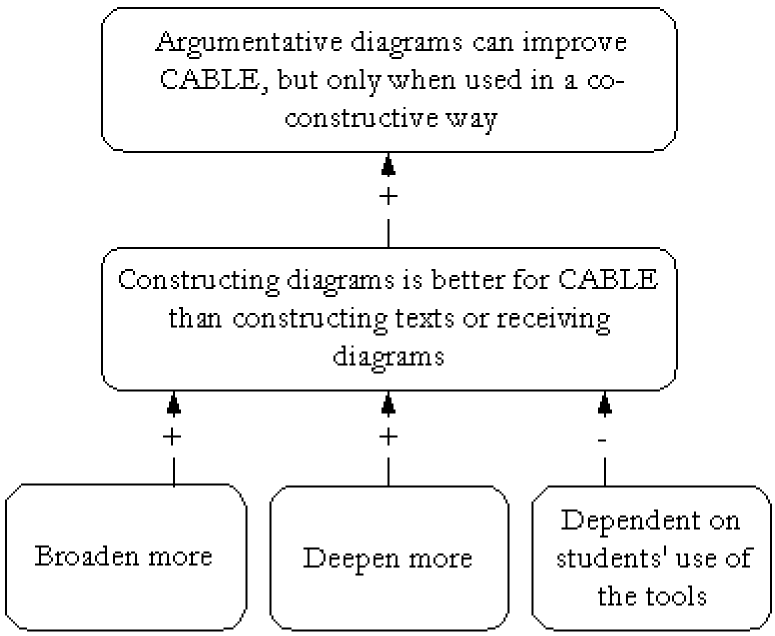
3

Representational tools in computer-supported collaborative argumentation-based learning:

How dyads work with constructed and inspected argumentative diagrams



Introduction

We probably all use argumentation in our daily lives. Whenever people have different opinions on an issue, they use argumentation to convince, to clear the air, or maybe just to have fun. This article focuses on collaborative argumentation with the goal of learning. Collaborative argumentation-based learning (CABLE) is increasingly used in education, because current practice - at least in the Netherlands - values peer collaboration and construction of knowledge. Through argumentative interaction, students exchange views and arguments, collaboratively constructing their knowledge of the "space of debate".

Using argumentation for collaborative learning can be difficult (e.g., Baker, 1996; 1999). Some researchers have suggested that representational tools can be used to facilitate argumentative learning (e.g., Suthers & Hundhausen, 2003). A tool that is often used is an argumentative diagram. Argumentative diagrams visualise the domain that is discussed. However, we do not know exactly how such diagrams contribute to learning or how learners should use them. For example, is it enough for learners to merely inspect argumentative diagrams, or should they actively construct such diagrams to facilitate learning? The present study aims to contribute to our understanding of the way in which representational tools can be used for learning by investigating argumentative diagrams in collaborative learning in secondary schools.

Collaborative Argumentation-Based Learning

People learn from argumentative interaction because it involves reasoning instead of merely retrieving information from memory (Andriessen, Baker & Suthers, 2003). In argumentation they have to make their thoughts explicit, which can aid learning through the self-explanation effect (Chi & Van Lehn, 1991). They also need to look at information from different sides, searching for causes and relations in the topic under discussion (Veerman & Treasure-Jones, 1999). Argumentative discussion may thus lead to a broader and deeper understanding of the "space of debate", which represents all possible positions and arguments regarding a certain topic. The space of debate can be finite, for example when students have to solve a problem in physics. In such cases, there are relatively stable ideas about what is correct and what not, which limits the exploration of the space of debate. In our view, there is more to be gained from argumentation when students work on open problems.

In argumentative problems without unique solutions, such as the desirability of genetically modified organisms, the space of debate is constructed by discussing different positions, ideas, and values. This space of debate is infinite. Learning is then defined as collaboratively broadening (i.e., using multiple viewpoints and subtopics) and deepening (i.e., using more elaborate arguments) the space of debate by constituting and transforming concepts and arguments. A broader and deeper understanding of the space of debate may lead to conceptual changes or attitude changes regarding the topic (Coirier, Andriessen & Chanquoy, 1999; Baker, 1996, 1999).

Several studies have found positive learning effects of collaborative argumentation. For example, Kuhn, Shaw and Felton (1997) found that dyadic discussion by adolescents on the topic of capital punishment improved reasoning about that topic. Erkens (1997) found that collaborative argument significantly improved 10-12 year old children's ability to solve problems. Finally, Reznitskaya et al. (2001) found that discussion by high school students promoted individual reasoning in writing persuasive essays. Students in the discussion-condition wrote essays that contained more arguments, counters, rebuttals, more formal argumentation, and more references to text information than students in the control condition who did not engage in collaborative discussion. These studies suggest that argumentation promotes learning, but such effects do not arise spontaneously.

Although argumentation is fairly often used in daily life, many people have difficulties constructing 'good' arguments. For example, Stein and Miller (1993) found that three-year old children are already capable of using arguments, but producing a well-substantiated argument is still very difficult for adults. Kuhn (1991) did an extensive study in which she asked adults to individually present and explain causal theories for societal issues. She found that people were apt at giving and explaining their opinions, but that only a third of them could come up with alternative views or counterarguments for their own view. Weighing different theories was seen even less often. When people discuss a topic together, they exchange views and arguments that can be used to construct knowledge. Unfortunately, people tend to ignore information or ideas that do not fit their own ideas (Wason, 1960; Chan, 2001). Additionally, social factors come into play in discussion. People may be afraid to lose face or to get into a fight. For effective argumentation, people have to be willing to argue, and need to have some common ground to make discussion possible. Taken together, these findings constrain the effectiveness of argumentation for learning.

Representational Tools

Representational tools have been suggested to support CABLE and alleviate many of the problems with argumentation. Representations can guide, constrain, or even determine cognitive behaviour (Zhang & Norman, 1994). Suthers and Hundhausen (2003) showed different effects on learners' discourse in the area of science. For example, matrix and diagram prompted students to discuss evidential relations more than a plain text.

Diagrams have been argued to have specific advantages as representational tools for argumentation-based learning (see Figure 1 in the method section for an example of an argumentative diagram). Argumentation can be visualised in diagrams by putting arguments in boxes and relations between them in arrows. Diagrams may benefit both construction and communication of arguments for many reasons, such as clarifying relations (Suthers, 2003), illustrating the structure of argumentation (Schwarz, Neuman, Gil & Ilya, 2000), giving overview (Larkin & Simon, 1987), helping to maintain focus (Veerman, 2000), and promoting reflection of alternative perspectives, solutions, and critiques (Kolodner & Guzdial, 1996). Thus, diagrams could be an important tool in supporting CABLE. However, until now there have been only a few studies showing that diagrams actually do support collaborative learning in the classroom (e.g., Toth, Suthers & Lesgold, 2003; Van Drie, Van Boxtel, Jaspers & Kanselaar, 2005).

In our view, there are at least three important questions to be answered about the conditions under which diagrams can be supportive for CABLE. The first question is: what are the specific advantages of diagrams over other representational tools? The most frequently used alternative for diagrams is a textual representation of arguments. A salient difference between text and diagram is linearity. A text is a linear representation, meaning that arguments are presented in a sequential fashion. In contrast, a diagram is non-linear because it displays arguments and argumentative relations in a two-dimensional space. It is exactly this two-dimensional space that has been argued to have specific advantages, because argumentation is in essence not linear (McCutchen, 1987; Coirier, Andriessen & Chanquoy, 1999). In addition, a diagram allows for multiple relations between arguments, by linking boxes with more than one arrow. Although text has many devices for expressing complex relations, such as advance organizers, and argumentative connectives that usually indicate a single relation between the previous and the following phrase (e.g., 'but' or 'because'), relations in a diagram are more salient (Suthers & Hundhausen, 2003). The text has to be processed sequentially when building a model of non-linear relations, while in a diagram this can be seen through parallel visual processes. Another advantage of diagrams over texts is that the limited space of boxes constrains the detail of the argument, allowing a clear overview; more topics can be represented with less detail. In spite of the theoretical advantages of diagrams for CABLE, texts may be easier to construct for people who are used to a narrative way of thinking (Chinn & Anderson, 1998). Although argumentation in itself is not linear, argumentative interaction happens through a linear dialogue. It may be difficult for students to represent their linear dialogue into a two-dimensional diagram. Thus, the first focus of the studies in this manuscript is on how the proposed differences between argumentative texts and diagrams contribute to broadening and deepening the space of debate in CABLE.

The second question regarding the use of diagrams in CABLE is about the situations in which a diagram would be conducive to learning processes (i.e., what processes and activities do diagrams foster, and when?). A clear distinction should be made between construction and inspection of diagrams (Cox, 1999). For example, the alleged benefits of showing the structure of argumentation may only arise if students actively construct diagrams themselves. On the other hand, giving an overview or helping to maintain focus may be promoted most when students inspect a diagram. The question of how to use a diagram is closely related to the question of when to use a diagram. For example, when students construct a diagram *during* discussion, it becomes a medium through which they discuss and build on each other's ideas (Suthers, Girardeau & Hundhausen, 2003; Van Drie, Van Boxtel, Jaspers & Kanselaar, 2005). Constructing a diagram *before* discussion activates prior knowledge, and helps students to structure and relate information about the topic. The advantage of individually constructed representations before discussion is also based on the hypothesis that people will argue more when they clearly see the things in which they differ (De Vries, Lund & Baker, 2002). *During* the discussion, diagrams form the basis for discussion if they present a clear and concise overview of the space of debate, or an individual point of view (Kanselaar et al., 2003). The differences between constructing and inspecting a diagram during CABLE have become more important in recent years, because new technologies make it possible to automatically present students with an argumentative diagram of the discussion they are engaged in. The linear, argumentative discussion is automatically put into a two-dimensional structure. This was also a goal for the SCALE process this study was part of. Thus, the second focus of our studies is the question whether it would help if someone else (e.g., an automatic system) represents argumentative texts as argumentative diagrams for later use during a discussion between students, compared to diagram construction by the learner. Because there was no automatic system available yet, the researchers changed texts into diagrams by hand (for more information, see method section).

The third question is how learners are actually using diagrams. While research has shown that different representations (e.g., diagrams and texts) provoke different learning activities (Zhang, 1997; Suthers & Hundhausen, 2003), representations do not determine learning activities. Perceived task goals, personal goals, and abilities may influence the realized benefits of argumentative diagrams as learning tools. For example, Postigo and Pozo (2000) described how a presented visual (mathematical) graph was only helpful for learning when interpreted globally. Local inspection led only to a focus on explicit elements, whereas global inspection required establishing conceptual relations based on an overall analysis of structure. Similarly, the benefits of the argumentative diagram in an open domain may only arise when students recognize its overall structure. It is in relating different arguments that co-construction and transformation of knowledge can take place. Therefore, it is also important to know whether students use the representations together or individually. While individual use of an argumentative diagram might lead to a simple accumulation of arguments, collaborative use can lead to conceptual change (see for example Roschelle, 1992, on convergent conceptual change). Our third question therefore aims at a qualitative exploration of what students actually *do* with the representations when exploring the space of debate. We inspect when students look at representations, what they do with it, and what they say about it to their partner.

The present study addressed all three aforementioned questions. We examined the quality of an open-ended discussion in terms of breadth and depth of the space of debate, and related this to the use of external representations. More specifically, we (i) compared a text and a diagram that students individually constructed before and after discussion, (ii) compared the collaborative use of a diagram students constructed themselves with a diagram that was made for them by the researcher based on students' own text, and (iii) investigated how students actually used these representations during collaboration.

The differences between texts and diagrams in terms of linearity, space, and ability to indicate relations could lead to different processes of broadening and deepening the space of debate. A diagram may be a good tool to broaden the topic under discussion, while a textual representation may lead to more deepening it in detail. The given diagram may lead to the broadest and deepest discussion, because students benefit from both linear detail they wrote in text, and structure and relations shown in the given diagram. However, if construction is more important than the modality of the representation in itself, students do not benefit from this given diagram.

Method

Participants

Students from seven classes in four upper secondary schools (pre-university) in the Netherlands participated (N=195). These schools volunteered to participate after a survey answered by 77 schools (Deliverable 8, SCALE-team, 2002). The students were 15 to 18 years old. A questionnaire given before intervention showed that collaboration in groups of two or more students was a fairly common practice, as was (individually) writing an argumentative text; students indicated having done this at least 1 to 5 times in the previous year. Seventy five percent of the students used computers to chat to other people, but not in school tasks. Half of the students indicated experience with collaborative work via the computer for school, but the other half said they had no experience with group work on the computer. Students were randomly divided into three groups according to condition (text *T*, diagram *D*, or given diagram made from text *TD*), and were put in dyads of students who differed in viewpoints and/ or arguments. Due to absenteeism, the number of dyads that were actually available for analysis (N= 58) is much smaller than the total number of dyads (N= 96). Dropout rates are about equally divided across conditions.

Design of the study

We used a collaborative writing task consisting of three phases (see Figure 1). The study included between-group differences in constructed representations before discussion (text T, diagram D), and in inspected representations during discussion (text T, diagram D, given diagram made from text TD). Representations were compared within groups (comparing individual representations from phase 1 with revise representations from phase 3). In phase 2, students discussed via chat and wrote an argumentative text together. The goal of writing gives direction and meaning to the discussion and may further broaden and deepen the space of debate.

Note that the given diagram made by the researcher (TD) was based on students' own ideas. It is a diagram made from the text students wrote themselves, not a diagram made on an expert's ideas to show the 'perfect' space of debate. In order to investigate differences in construction versus inspection of diagrams for learning, the diagrams should be comparable in terms of structure and content. Therefore, the given diagrams represent learners' own ideas instead of an expert's ideas.

Computer-based activities	*Condition D*	*Condition T*	*Condition TD*
Phase 1 *Individual preparation:* Students read sources and construct either text or diagram			
Between phase 1 and 2 Researcher constructs diagrams from texts in TD condition			
Phase 2 *Discussion and collaborative writing in dyads*: Dyads chat and write text collaboratively while their two individual diagrams (D), texts (T) or diagrams made from texts (TD) are available for inspection			
Phase 3 *Individual consolidation:* Students revise their own diagram or text from phase 1			

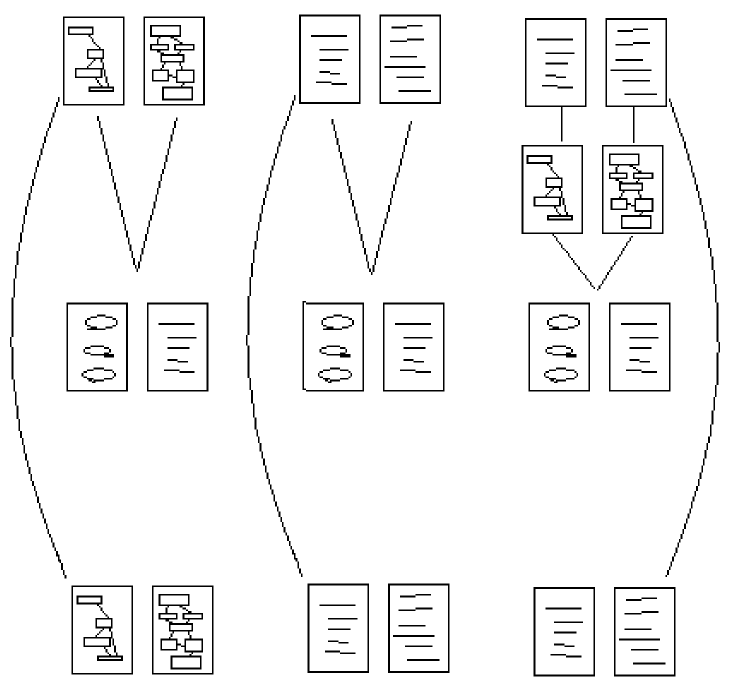


Figure 1. *Design of the study*

Computer environment

The computer environment that was used is *TC3*, developed at the Department of Educational Sciences in Utrecht to support collaborative argumentative writing in dyads (Jaspers & Erkens, 2000). For the present study an individual and a collaborative version of TC3 was developed.

When the individual version of TC3 was started, a user saw three windows: a chat window, (disabled), a window to write a text, and an information window.

Extra feature was a diagram (Figure 2), which popped up when the *diagram* button was clicked. The information window consisted of nine tabs containing the task assignment, a manual, criteria for assessment, and six information sources on the topic of genetically modified organisms (GMOs[6]). In the individual version students used either the text window or the diagram, depending on the condition they were in.

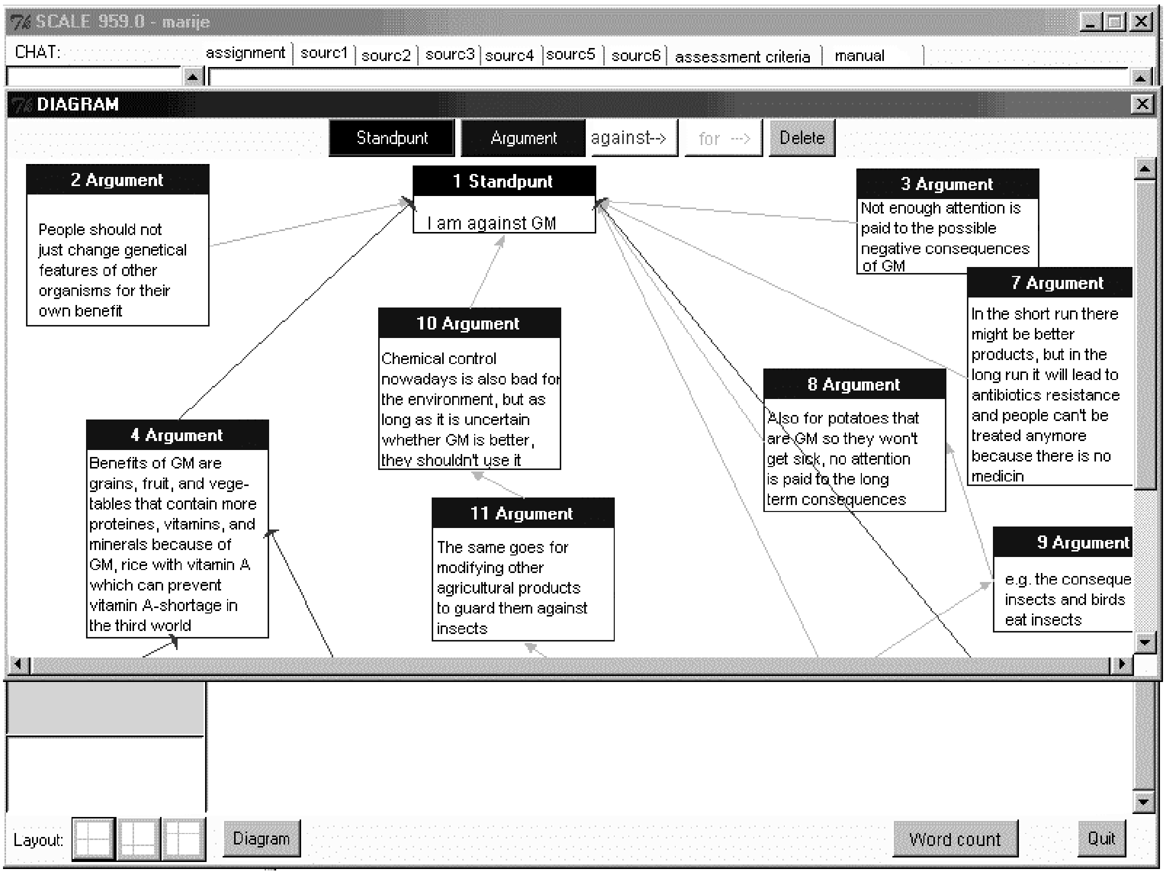


Figure 2. *An example of a diagram in the individual version of TC3 (translated from Dutch to English)*

The main windows in the collaborative version were the same as in the individual version of TC3 (chat, text, and information, see Figure 3). Students needed a number to log on, which enabled the program to link two computers. The chat window (1) could then be used to chat with the partner the student was linked to. It had three parts: below, a student could type his/her lines, in the middle part one could see what the other one was typing, and at the top the chat history was shown. The chat history was saved for all sessions. The text window

[6] We searched for information about GMOs on the Internet, in newspapers, and in magazines. Sources were chosen that reflect diverse standpoints and arguments. For more information on the selection procedure of the information, see SCALE report, Deliverable 1 & 2, 2002.

(2), used for collaborative text writing, could only be accessed in turns. Turn taking was done by using the traffic light (4). The information window (3) did not contain information sources on GMOs anymore, because we wanted students to make use of the information they put in their individual representations instead of the information in the official sources, to see what they thought was important themselves. The TC3 manual, information on the assignment, and assessment criteria were still available. Students had access to individually made texts or diagrams to look at their own standpoint and arguments or see their partners' representation. The representations were shown when clicking the buttons at the lower bar (5). All these windows could be accessed at all times, there was no specific order in which the task had to be carried out. Students could see on their own screen when their partner was writing chat or writing text, but not when their partner was looking at the representations or the task information. There was no shared pointer to refer to things; students could only refer in language. The boxes in the diagram are numbered for easy reference.

Procedure

The study consisted of six sessions of 50 minutes. Students were asked to discuss and write about the controversial topic of genetically modified organisms (GMOs). Since the topic of GMOs and argumentation are both part of the curriculum in upper secondary education, the experiments could be conducted during (six) regular hours in classes of Dutch language or Biology.

The six sessions were divided in three phases: (1) The individual preparation phase, about 80 minutes long; (2) the discussion and collaborative writing phase, about 150 minutes long; and (3) the individual consolidation phase, about 30 minutes long.

In the first phase, students gained initial knowledge on the issue of GMOs by reading information sources. They formed their opinion, supported with

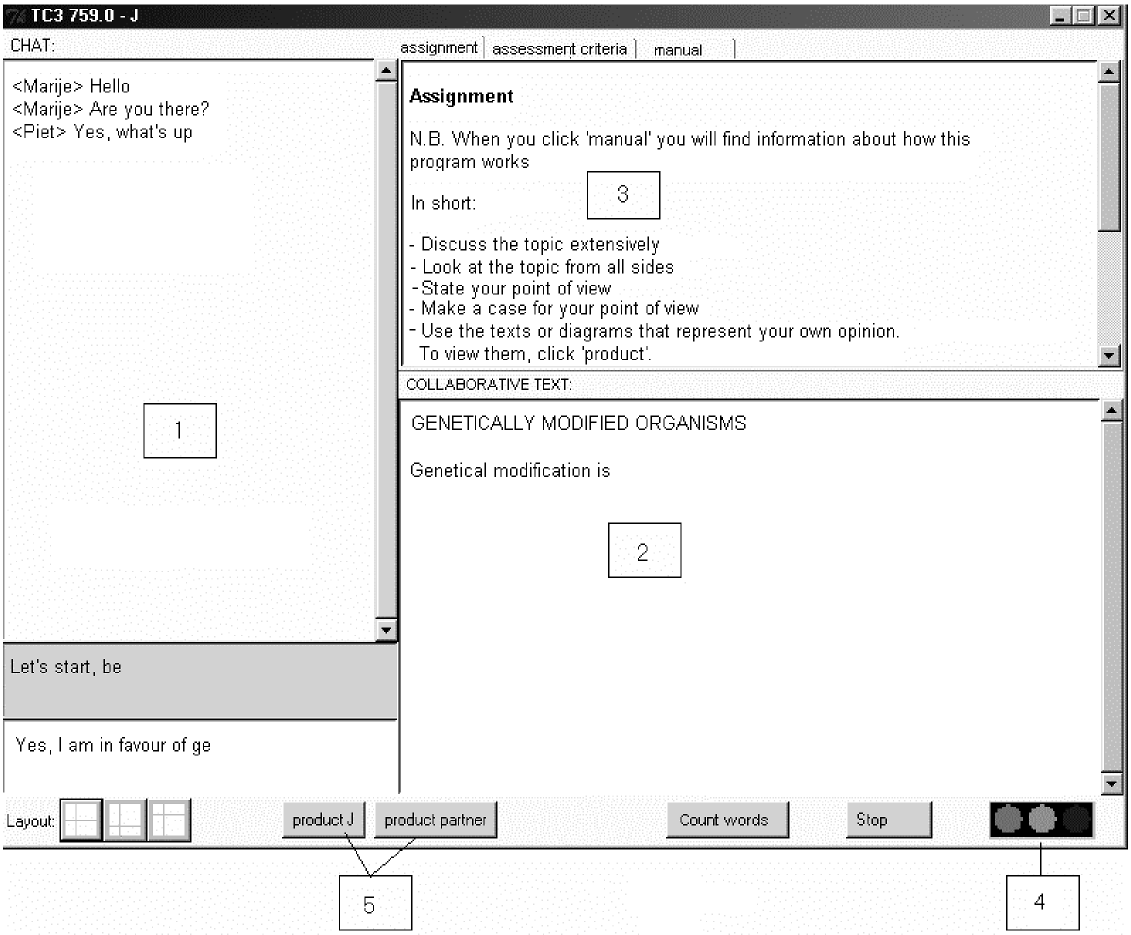


Figure 3. *Screen dump of the collaborative version of TC3 (translated from Dutch to English)*

arguments and counterarguments, and put that in a diagram or text.

Between the first and the second phase the researchers formed dyads of students based on the individual representation that the students had constructed in phase 1. Students with different opinions and/or different arguments were put together in order to provoke discussion between students with different ideas on GMOs. Not all dyads were comparable in terms of breadth and depth of the individual representations, but there were no differences in (un)equality of dyads across conditions.

For the given diagram condition TD, the researchers constructed diagrams of the texts students wrote. Diagrams contained sentences from the texts in boxes, and linguistic markers were used to construct arrows. Sentences in text were summarized in diagram to fit the boxes. Appendix A shows an example of a text that was represented in a diagram (translated from Dutch into English).

In the second phase, dyads chatted about GMOs and wrote an argumentative text together reflecting their opinion. Information sources were thus not available anymore, but two individual texts, two diagrams, or two diagrams made by the researcher of the texts were.

In the third phase, individual students went back to their individual diagram or text and revised these to represent what they thought and knew about the topic after debate.

Methods of analysis

The process of collaborative discussion and writing was analysed with MEPA (Multi Episode Protocol Analysis, Erkens, 2001). MEPA automatically divides the protocol in separate activities. The chat is separated by students pressing <Enter>; the text is separated by students taking turns. Every push on a button is separately logged, and the log file can be used to replay the whole process. The unit of analysis for chat followed MEPA's division of activities, except where students cut their sentences in chat or put different argumentative moves in one sentence. In phases 1 and 3, we did not analyse the process but only the product of the phase. The unit of analysis for the individually constructed texts and diagrams was a sentence, except when students put more argumentative moves in one sentence.

Rainbow. The Rainbow framework (Baker, Andriessen, Lund, Van Amelsvoort & Quignard, 2006), developed in the SCALE project, defines students' general collaborative activities in seven categories. These seven categories distinguish task-related activities (categories 4-7) from non-task related (categories 1-3) activities, and argumentative activities (categories 5-7) from non-argumentative activities (categories 1-4). A description of each of the categories can be found in Table 1. Together, they reflect the richness and diversity of a real-life discussion in a certain task environment. The categories can only be applied in their context, because activities derive their meaning within the collaborative process. The framework is descriptive, not normative, and provides information on both frequencies and sequences of activities. Because the categories are indicated with the seven colours of the rainbow, the sequence of activities can be easily viewed visually in an analysed protocol. Moreover, the Rainbow framework can be applied to all activities students carried out in the collaborative environment, revealing the mix of activities in chat, writing, representations and computer environment. Inter-rater agreement on ten protocols was .82 (Cohen's Kappa).

Table 1. *Rainbow categories*

Rainbow category	Explanation	Example
1 Outside activity	all interactions that do not have anything to do with the task	"How was the party yesterday?"
2 Social relation	all remarks about the social relation	"You are doing well!"
3 Interaction management	all remarks about communication, like checking presence, checking understanding	"Hello, are you there?"
4 Task management	all remarks and actions for managing the task	"It's your turn to write now"
5 Opinions	all statements about students' opinions	"I am in favour of GMOs"
6 Arguments	all arguments and counter-arguments students use to support or rebut a statement	"Because of Genetically modified food hunger in the third world will be banned"
7 Explore and deepen	all remarks that explore and deepen the (counter)arguments	"but hunger in the third world is not due to lack of food in the world, but to unequal division of food"

Breadth and depth of the space of debate. While Rainbow tells us about general activities in chat and writing, breadth and depth of the space of debate tell us about the extent to which the topic is explored. Breadth and depth of the space are based on categories five, six and seven of the Rainbow framework, because these categories comprise argumentative content. This is done firstly because we wanted to elaborate upon students' argumentative activities, the main concept in our studies. Secondly, the analysis of breadth and depth of the space of debate can be applied to both the individual representations and the collaborative discussion and writing, enabling us to view development of the space of debate in the three phases. The Rainbow framework, in contrast, was developed for the processes of collaborative activities only.

The *breadth* of the space of debate was defined as the amount of topics and subtopics mentioned. The debate on GMOs includes a variety of epistemological points of view (biological, agricultural, political, economical, ethical), and a variety of social actors in the debate (grain producers, researchers, citizens, farmers, politicians, non-governmental organizations). When students broaden the space of debate, they look for example at GMOs from the view of Greenpeace, but also from the view of the government, or they move from talking about consequences of GMOs for health to consequences for the environment. We distinguished five main topics in the GMOs issue, namely health, environment, affluence, worldview, and other. These topics were further divided into fourteen subtopics (such as affluence-hunger/food; affluence-costs/benefits). Inter-rater reliability of breadth between two judges was .75 (Cohen's Kappa).

When students talk in *depth* about the topic under discussion, they relate different concepts, and elaborate upon their arguments. For example, students do not only say that GMOs are bad for the environment, but also argue why this is the case, and give an example. For the analysis of the depth of the debate, a scoring system was developed in which all content-related utterances were analysed in their immediate context. The scheme consisted of four categories that received different weights: 1) stating an argument, 2) giving an example or explanation of an argument, 3) stating a support or rebuttal, and 4) explicit explanation of a relation between different arguments. We looked at episodes of content-related activities. For example, a support or rebuttal could only be scored if an argument is scored first. In the statistical analyses these categories were used separately. Hence, there were four scores that students could get for depth. However, because deepening the space of debate is shown by a line of argument, we also look at the total depth in which the scores on the separate categories are added. Table 2 gives an example of a scored protocol. The interrater reliability of the depth of the space of debate was .77 (Cohen's Kappa).

Table 2. *Example of scoring breadth and depth of the space of debate*

Content of argument	Who	From where	Breadth	What happens	Depth
I am pro, because it is good for the 3rd world, they can use extra vitamins	Maria	own diagram	Health-nutrients	argument and explanation	1+2
No, the 3rd world cannot afford GM, it is only meant for the rich West, and then nobody will buy products from the 3rd world anymore	Tom	new	Affluence-division	rebuttal, explicit relation and explanation	3+4+2
But the rich countries will help the poor countries with money and funding	Maria	new	Affluence-division	rebuttal and explanation	3+2
That happens already (funding), but with GMOs nobody will buy things from the 3rd world and they will become even more poor	Tom	new	Affluence-division	rebuttal and further explanation of argument #2	3+2
					22

Use of individual texts and diagrams during collaborative phase. The representations text, diagram and given diagram made from text are available for inspection during the second phase of discussion and collaborative writing. The analysis presented here addresses the third research question of how these representations are used by the students in collaboration. We as researchers and developers may have ideas on how representations should be used, but that may not be what students actually do. Students can use the representations in a local way (i.e., to copy elements), or in a more global way (i.e., as a starting point for a constructive discussion that leads to constituting or transforming knowledge). More specifically, we distinguished seven ways in which students used the representations during the collaborative phase: (1) to look at each other's representation and compare; (2) to remember own opinion or arguments; (3) to copy, to find arguments, as a source of information; (4) as a trigger to talk or write about; (5) as a starting point, followed by individual construction of new or transformed knowledge; (6) as a starting point, followed by collaborative construction of new or transformed knowledge; (7) unclear. These seven ways of using the individual representations were derived from the protocols, by looking at what students actually did during chatting and collaborative text writing. Like the other two analyses, this analysis can only be derived in its context. For example, the protocol shows that a student first clicks the button to view her own representation, then the button to view her partner's representation, and then chats: 'We are both in favour of GMOs'. This episode can then be categorized as (1).

Results

The description of the results follows the task phases. First we describe the individual breadth and depth of texts and diagrams in phase 1. Then we look at how dyads work in the collaborative chat and text writing (phase 2). In phase 2, we also take a more in-depth look in how students have used the representations in their exploration of the topic of genetically modified organisms. Finally, we look at the breadth and depth of texts and diagrams in phase 3, and their improvement over time.

Phase 1: Individual construction of Text or Diagram

Our first research question dealt with differences in breadth and depth of two types of constructed representations: texts and diagrams. Students individually constructed either text or diagram in phase 1. We conducted independent-samples t-tests on the breadth and depth scores between text and diagram conditions. There were no significant differences in *breadth* of the space of debate between texts and diagrams, $t(58) = 1.20$, $p > .05$. Results also showed no significant differences for any of the measures of *depth* (arguments, explanations/examples, supportives/rebuttals, explicit relations). In short, our expectation that diagrams would be broader, and texts would be deeper was not confirmed. This means that students in both conditions started their collaborative phase 2 with similar individual spaces of debate.

Phase 2: Collaborative discussion and text writing

The second research question was aimed at understanding differences in exploration of the space of debate when students had a text, a diagram, or a given diagram available for inspection. We first analysed the collaborative discussion and writing (i.e., phase 2) with the Rainbow framework to characterize the activities students performed in general. From these general activities, we isolated the argumentative content actions and analysed them on breadth and depth of the space of debate.

Activities in chat and writing in dyads: Rainbow. Results of Rainbow analyses are shown in Table 3. A one-way ANOVA was performed on the means of the Rainbow categories between the three conditions. We did not find significant differences for any of the Rainbow categories. Only Rainbow category 5 ‘Opinions’, $F(2,27) = 3.17$, $p = .06$, showed an almost significant difference; the percentage of opinions in the text-condition was higher than in the other two conditions. Therefore, we talk about the means of percentages over all three conditions (adding and dividing the three columns in Table 3) in the remainder of this section. Overall, less than 4% of students’ activities were outside activity, indicating that they were focused on the task. Most striking in the results was the fact that students invested a large amount of their activities (66%) in managing the task, specifically the writing task. For example, the students discussed who was to write, counted the words of their text, looked at their individual texts or diagrams, or worked on structure or spelling of the text. About 16% of all activity was spent on content interaction (categories 5, 6, and 7), chatting and writing about GMOs. This percentage is consistent with findings in other argumentative tasks (e.g., Veerman, 2000; Van Boxtel, 2000).

Table 3. *Mean frequencies (with standard deviations) and percentages of Rainbow categories*

Category		*Text*		*Diagram*		*Given diagram*	
		Mean Frequency (SD)	*Percentage*	*Mean Frequency (SD)*	*Percentage*	*Mean Frequency (SD)*	*Percentage*
1	Outside activity	6.40 (7.09)	1.2	23.10 (33.96)	4.6	30.40 (26.74)	5.4
2	Social relation	21.50 (18.86)	4.5	18.20 (10.68)	4.0	31.00 (31.50)	5.0
3	Interaction management	46.80 (29.07)	9.6	47.70 (19.36)	10.1	49.00 (30.94)	9.0
4	Task management	298.30 (102.12)	65.0	318.10 (88.72)	67.5	341.70 (102.32)	66.4
5	Opinions	27.50 (8.91)	6.5	18.00 (6.87)	3.9	24.00 (10.00)	4.7
6	Argumentation	19.20 (8.46)	4.5	20.00 (10.36)	4.2	17.50 (7.88)	3.7
7	Explore and deepen	35.40 (20.66)	8.8	25.10 (12.63)	5.7	29.50 (17.15)	5.7
	Total	455.10	100	470.20	100	523.10	100

Content related argumentative interaction in chat and writing. ANOVAs of breadth and depth in the collaborative phase showed a significant difference between dyads in the diagram and text condition for both *broadening*, $F(2,27) = 5.82$, $p < .05$, and *deepening*, $F(2,27) = 3.48$, $p < .05$, the space of debate. Post hoc tests indicated that dyads broaden more in the diagram condition than in the text condition and the given diagram condition, and also deepen more in the diagram-condition than in the given diagram condition. The text condition did not differ from the diagram and the given diagram conditions on deepening the space of debate. The significant difference in depth was due to a difference in examples and explanations, $F(2,27) = 7.96$, $p < .05$, with the diagram being deeper ($M = 18.4$; $SD = 5.9$) than the text ($M = 11.6$; $SD = 4.9$) and the given diagram ($M = 10.2$; $SD = 3.8$).

In short, students in the diagram-condition broadened and deepened the most. The given diagram condition did not benefit from constructing text and inspecting diagram. Students in the given diagram condition broadened and deepened the space of debate less than the dyads in the other conditions.

Broadening and deepening the space of debate in the collaborative phase 2 was done via chat and via collaboratively writing the argumentative text, both in different windows. We therefore split chat and writing (see Figures 4 and 5). Correlations are not significant between chat and text in either breadth ($r = -.25$, p

> .05) or depth (M_r = .13 on depth-measures separately, $p > .05$), indicating that there is no straightforward relation between dyads' exploration in chat and in collaborative writing.

In *chat*, a trend towards significance for *broadening* the space of debate in chat, $F(2,27) = 3.07$, $p = .06$, is seen, with dyads in the diagram condition broadening the most (M_{chat} = 8.8; SD = 2.4), and dyads in the given diagram condition broadening the least in chat (M_{chat} = 6.1; SD = 2.6). There is no significant effect in chat for *deepening* the space of debate between conditions, F (2,27) = 510.63, $p > .05$.

In *collaborative writing*, there is a trend towards significance for *broadening* the space of debate, $F(2,27) = 2.94$, $p = .07$, with dyads in the diagram-condition broadening the most ($M_{writing}$ = 9.5; SD = 1.4), and dyads in the text-condition broadening the least in writing together ($M_{writing}$ = 7.6; SD = 2.0). For *deepening*, a significant effect between conditions was found, $F(2,27) = 1022.43$, $p < .01$. Students in the diagram-condition (M = 10.9, SD = 2.2) use significantly more examples and explanations than students in the text-condition (M = 5.7, SD = 1.8) and the given diagram condition (M = 5.4, SD = 1.9), $F(2,27) = 24.04$, $p < .01$. The number of arguments used in writing shows a trend toward significance, with the diagram condition deepening with more arguments (M = 11.9, SD = 5.5) than the given diagram condition (M = 8.5, SD = 3.9) and the text condition (M = 8.0, SD = 2.4), $F(2,27) = 2.66$, $p = .09$. Figures 6 and 7 display the subcategories for depth and breadth separately in chat and in writing.

To summarize, dyads in the diagram-condition deepened their collaborative writing more than dyads in the other two conditions, mainly due to the amount of explanations and examples they use. Although the difference in broadening is not significant, the effect sizes are high, indicating that this difference might become significant with a larger sample size. Students in the given diagram did not benefit from both text-construction and diagram-inspection; they never score higher than the other two conditions on either breadth or depth of the space of debate.

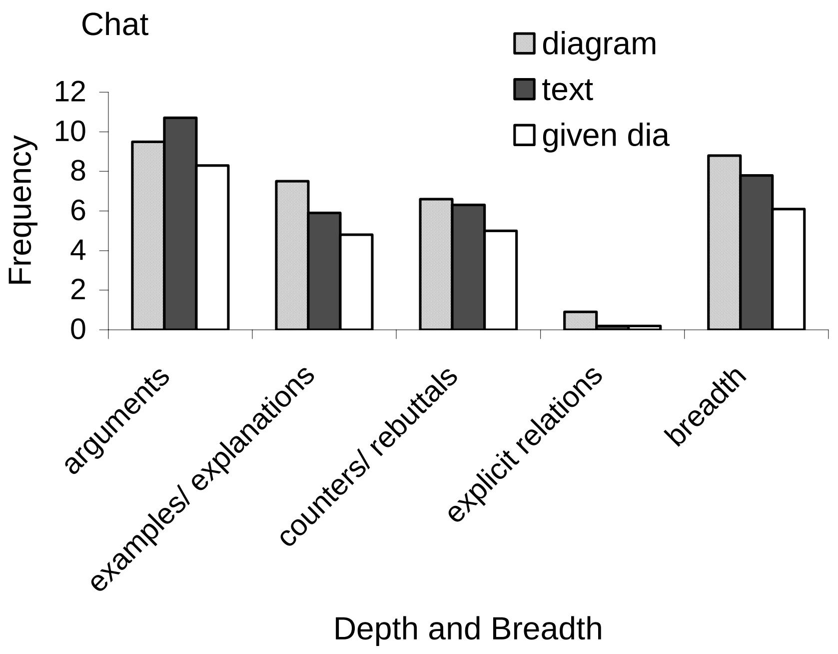


Figure 6. *Frequencies of depth and breadth in chat, split for the three conditions*

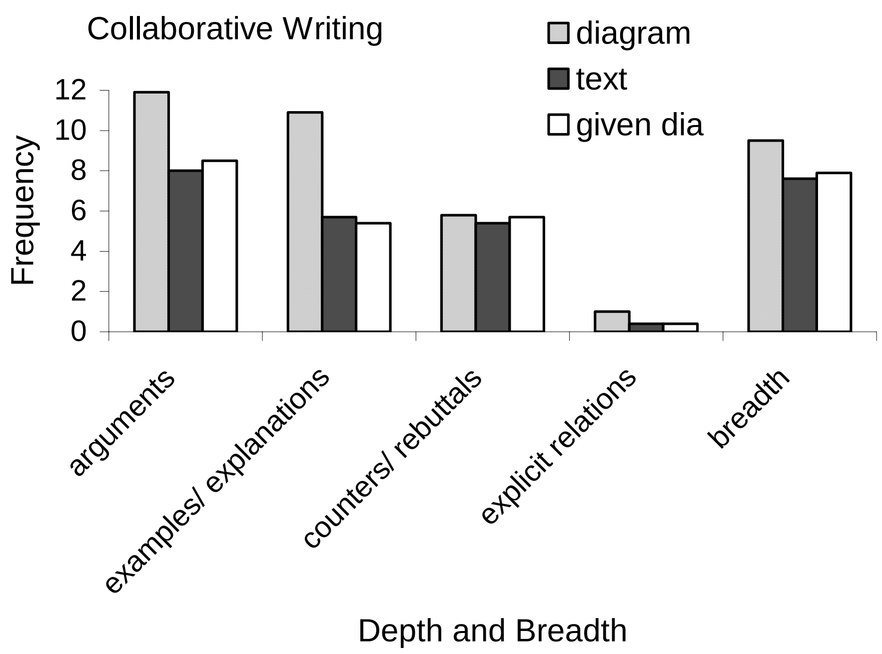


Figure 7. *Frequencies of depth and breadth in collaborative writing, split for the three conditions*

How students use the representations in exploring the space of debate. The third research question was concerned with variations between dyads in how they used the representations in their collaborative effort to broaden and deepen the space of debate. As can be seen from the results described above, standard deviations are rather high. This means that there are substantial differences between dyads in one condition. In this section a qualitative exploration of different types of dyads is presented in order to illustrate the sources of variation in the use of representational tools.

When we take a closer look at the differences in broadening and deepening for each dyad, three types of dyads can be distinguished. Some dyads broaden and deepen the space of debate very much in chat, compared to their individual representations, but not in writing. We call these the 'Mountains', because their broadening and deepening scores can be visually represented as a mountain 'Λ', namely relative low score on individual representations, high score in chat, and low score in collaborative writing. For other dyads the opposite is true (the 'Valleys', 'V'), and a third group of dyads shows a shallow individual representation, followed by a somewhat broader and deeper chat, and an even broader and deeper collaboratively written text (the 'Rising Slopes', '/'). Figure 8 displays scores of a Mountain, Valley and Rising Slope pair to visually show how they received their name. The scores of all dyads can be found in Appendix B. Two dyads could not be classified. From Appendix B it can be seen that the Rising Slopes are found more in the diagram condition (5 times) than in the other two conditions (2 and 3 times), the other types are about equally divided over conditions. Note that chat and collaborative writing can be done in parallel; students do not have to use chat first and text later.

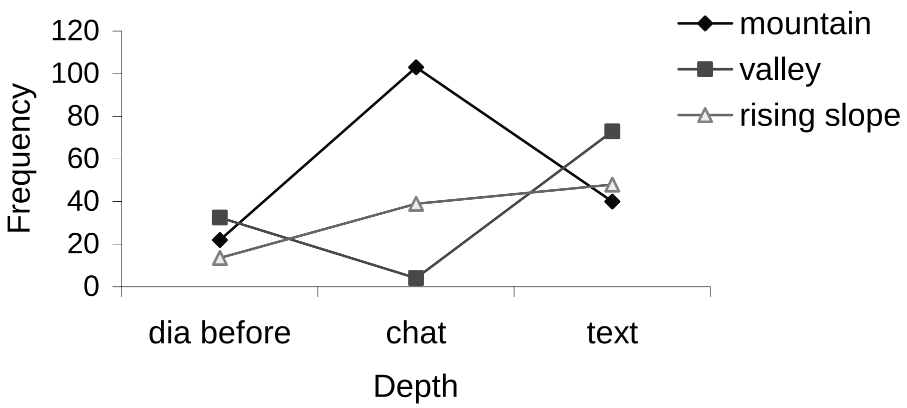


Figure 8. *Visual display of a mountain, a valley and a rising slope strategy*

Below, an example is given of a dyad that can be defined as 'Mountain' (text condition). After exchanging greetings, and some first exchanges of viewpoints, they chat:

Joleen	why are you in favour?
Mary	because I think it could solve a lot of problems in the world, e.g. famine *(telling from own individual text)*
Mary	I think it is not necessary in the West, we already have enough *(new)*
Joleen	And little farmers won't have any chance when gm-food is made by other companies *(telling from own individual text)*
Mary	Indeed, I also think we should still have natural food *(telling from own text)*
Joleen	Indeed, gm is ok, but there are boundaries *(telling from own text)*
Mary	It is important that it is safe for humans *(new)*
Mary	You shouldn't be dying after eating gm-food *(transforming)*

From this excerpt, all done in chat, it can be seen that students explore their space of debate in chat. Mary answers Joleen's question first with an opinion that she already put in her individual text, but then elaborates on her own opinion with a restriction that GMOs would not be necessary in the West. Joleen implicitly agrees with her on restrictions by giving another argument against GMOs. Giving arguments from the two individual texts combined with elaborations and transformations, these two students negotiate a joint standpoint with arguments in chat that they later put almost exactly in their collaborative text. The beginning of their collaborative text is displayed below:

> We are in favour of genetic modification, if we keep boundaries.
> We think that especially in the West genetic modification should not be used, because we already have more than enough here. Here in the West we often have food in abundance, so genetic modification wouldn't help us.
> In the third world, genetic modification can be a solution to famine. …

The next example is an example of a 'Valley' (given diagram condition). This excerpt displays the start of their collaboration. Note the immediate focus on writing the text:

Bill	Hi Colin, let's write!
Bill	You were against, weren't you?
Colin	I am against applications that haven't been researched yet. I am in favour of applications that have proved to be useful *(telling from own individual diagram)*
Bill	Me too
Colin	See Colin's product
Bill	*writes in text:* We are in favour of GMOs, only if the applications have been researched and proven *(telling from chat and partner's diagram)*

These two excerpts of chat and writing are typical of their categories. Valley-dyads are focused on writing; they see the collaborative text they have to write as the goal of the task. Chat is used mostly to manage the writing task. For example, the dyad from the excerpt above was aimed at managing the task (Rainbow category 4) in chat 86% of the time. Only 6 % of their chat was aimed at argumentative interaction (Rainbow categories 5 to 7). The Mountain-dyads are much more focused on exploring the space of debate before they start writing. However, their collaborative text is not as broad and deep as their chat. A Rising Slope-dyad shows characteristics of both Mountains and Valleys, some focus on discussion in chat, and some focus on broadening and deepening in text.

How are individual representations used in phase 2, the discussion phase? While students are discussing and writing together, they have access to the individual texts or diagrams they made. We logged frequency and timing of students looking at the individual representations. Unfortunately, due to an error in the logfiles, we cannot be certain which representation is a students' own, and which is their partner's. However, we can distinguish between the two representations. During collaboration, students looked about 16 times at one representation, M = 15.67, SD = 12.03), and 18 times at the other representation (M = 17.67, SD = 11.74), for about 20 seconds at a time. An ANOVA revealed no overall differences in how often, F (2, 27) = .39, $p > .05$, and how long, F (2, 27) = 1.13, $p > .05$, students looked at the individual representations between conditions. However, when distinguishing Mountains, Valleys and Rising Slopes, dyads defined as Mountains checked their individual representations significantly fewer times (M = 23.0; SD = 18.2) than Valleys (M = 44.5; SD = 16.1) and Rising Slopes (M = 46.2; SD = 24.3), F (2, 23) = 3.78, $p < .05$.

Dyads defined as Mountains, Valleys, or Rising Slopes do not only differ in frequency of using their individual representations, but also in the way in which they use these representations during the collaborative phase. In chat, Mountains as well as Rising Slopes used the individual representations more often than Valleys as a starting point for discussion, leading to a constructive collaborative discussion ($M_{mountains}$ = 1.7; $M_{risingslope}$ = 1.5; M_{valley} = 0). In writing together, Mountains showed less telling from individual representations than Valleys or Rising Slopes ($M_{mountains}$ = 0.7; $M_{risingslope}$ = 4.3; M_{valley} = 3.7). Mountains made less use of the individual representations during writing overall ($M_{mountains}$ = 1.7; $M_{risingslope}$ = 6.5; M_{valley} = 5.7).

Below, some examples are given of how students use the representations while they are discussing the topic of GMOs. The first excerpt (diagram condition, valley) is an example of how students use the individual representations to remember their viewpoints and arguments, and compare their representations. They first look at the diagram they made individually, and check their partner's representation. Then they start discussing each other's work. This often happens at the start of the collaborative phase, or at the start of a new lesson.

Vincent and Katie both start with looking at each other's diagram, then chat:	
Katie	Your point of view is neither for, neither against *(deduction from partner's diagram)*
Vincent	Yours is definitely in favour, wait, I'm going to read it carefully *(deduction from partner's diagram)*
< *Vincent opens Katie's diagram* >	
Katie	Ok
< *Katie opens Vincent's diagram* >	
Vincent	Your text* is good, how is mine?
Katie	Also good, but I think you are more in favour than against
Vincent	Yes, that's right
Vincent	Then let's write a text that is pro, but also has some aspects of against
Katie	A little in favour, a little against
Vincent	exactly
	…
Katie	Uhm…GMOs are good for society? (*transforming*)
< *Vincent opens own diagram, Katie opens Vincent's diagram* >	
Vincent	Genetic modification of food is good, as long as there are no harmful consequences? (*telling from own diagram*)

* Vincent refers to Katie's diagram as text.

Sometimes, the individual representations can give rise to a discussion that moves beyond what is written in the representations; it leads to an elaboration of ideas and arguments. In the discussion, this is mostly a collaborative activity, while during collaborative text writing, the elaboration is also often done by the individual who is writing. In the example below, Aisha and Odin exchange greetings first, and then spend some time looking at both individual representations. Then Aisha summarizes their two individual given diagrams by saying that she thinks they agree. Odin agrees with her finding, but then elaborates on his own ideas with a more subtle opinion that he didn't mention in his representation. This gives rise to new arguments, such as the one from Aisha at the end of the excerpt:

Example (given diagram, Mountain):

< *After exchanging greetings and looking at both individual representations:* >	
Aisha	I think we agree for a large part
Odin	indeed
Odin	Actually, I'm both in favour and against *(new)*
Aisha	How?
Odin	I think we should adjust our food production to the world population, or even more people will die from hunger *(telling from own diagram)*
Aisha	Exactly, but we cannot use it to make it easy for ourselves *(new)*

During text writing, individual representations are often used to find arguments and put them in the collaborative text, more or less literally. Students also occasionally look at the individual representations when their partner is writing, communicating found arguments to the other person by chat, or using them later when they write a part of the collaborative text.

Example (text, slightly Rising Slope):

After almost an hour of collaborating, Nelly is writing the introduction of the collaborative text, while Kim is looking at her individual text	
Kim	Shall I write something about diversity? *(topic from her individual text)*
Nelly	I'll do that, so you can look at my introduction and finish that

In summary, our qualitative analyses show that three strategies can be distinguished in collaborative argumentation-based learning, namely Mountains, Valleys and Rising Slopes. These strategies determine how dyads of students use representational tools. Students who display a deep discussion in chat (the Mountains and Rising Slopes) use their representations mostly to start a co-constructive discussion, while students who display a shallow discussion in chat (the Valleys) mostly copy information from their representations directly in their collaborative text.

Phase 3: Individual revision of Text or Diagram

In phase 3, after discussion, students revised their individual text or diagram from phase 1. We first compare texts and diagrams in the third phase. Then we compare the revised representations from the third phase with the constructed representations from the first phase, to get an indication of individual learning. Table 4 presents the means and standard deviations for breadth and depth of texts and diagrams in the first preparation and the third consolidation phase.

Table 4. *Means (with standard deviations) of breadth and depth in individual representations*

Phase	Type		Text		Diagram	
			M	*(SD)*	*M*	*(SD)*
Phase 1: Individual preparation	Breadth		5.9	(1.9)	6.4	(2.1)
	Depth	1	6.5	(2.8)	5.7	(1.8)
		2	2.0	(1.3)	3.6	(2.3)
		3	4.0	(2.3)	3.1	(2.4)
		4	0.3	(0.6)	0.0	(0.0)
Phase 3: Individual Consolidation	Breadth		6.8	(2.1)	7.6	(1.7)
	Depth	1	7.6	(3.1)	8.3	(2.0)
		2	3.4	(1.8)	4.2	(2.3)
		3	4.4	(2.6)	3.5	(2.2)
		4	0.3	(0.7)	0.1	(0.2)

Note 1. Depth: 1= arguments; 2= examples and explanations; 3= counters/rebuttals; 4= explicit relations

Note 2. No significant differences between Text and Diagram conditions

Texts versus Diagrams in phase 3. An independent-samples t-test revealed no differences in *breadth* of the space of debate between the diagram and the text condition after discussion, $t(57) = 1.55$, $p > .05$, nor in *depth* of the space of debate. This means that there was also no difference between texts and diagrams in breadth and depth of the space of debate after discussion.

Learning with texts or diagrams. Comparison of the revised representations after discussion with the constructed ones before discussion gives us an indication of what individual students have learned from discussion and collaborative writing, and whether the format of the representations has an effect. To examine the effects of the format of the representations on learning, an ANOVA was performed on the diagrams and texts made before and adapted after discussion. The design of the analysis was format (diagram, text) as a between-subjects factor, with repeated measure of time (before discussion, after discussion). Analysis revealed a significant effect of time, $F(1, 57) = 45.54$, $p < .001$, for *breadth*, indicating that all students had broadened their space of debate after discussion. However, there was no interaction with format, $F(1,57) = 0.61$, $p > .05$. There was also a significant effect of time, $F(1, 58) = 60.83$, $p < .001$, for *depth*, indicating that all students deepened their space of debate after discussion compared to before discussion (see Figures 9 and 10). A significant interaction with condition, $F(1, 58) = 4.07$, $p < .05$ was found. This was due to a significant difference in number of arguments. Before discussion, the number of arguments was lower in the diagram condition than in the text condition ($M_{textbefore} = 6.45$; $M_{diagrambefore} = 5.65$), but after discussion this relationship was reversed ($M_{textafter} = 7.63$; $M_{diagramafter} = 8.25$). To interpret the size of the difference, we used Cohen's *d* (1988) to calculate difference scores before and after discussion in both conditions. Cohen's *d* makes students' scores directly comparable, regardless of individual differences in breadth and depth. We found a mean effect size of 0.95 for depth, and 1.69 for breadth. These scores are considered to be large effects (Cohen, 1988), indicating that students have improved their space of debate considerably. Almost twenty percent indicated a change of opinion in their revised representation.

In summary, all students have learned significantly in terms of breadth and depth of the space of debate. Closer inspection of the representations revealed that they did this by adding boxes or sentences in their representation rather than changing existing texts, boxes or relations. There were no differences between texts and diagrams in the third phase, but students in the diagram condition deepened their knowledge more than students in the text-condition from phase 1 to phase 3.

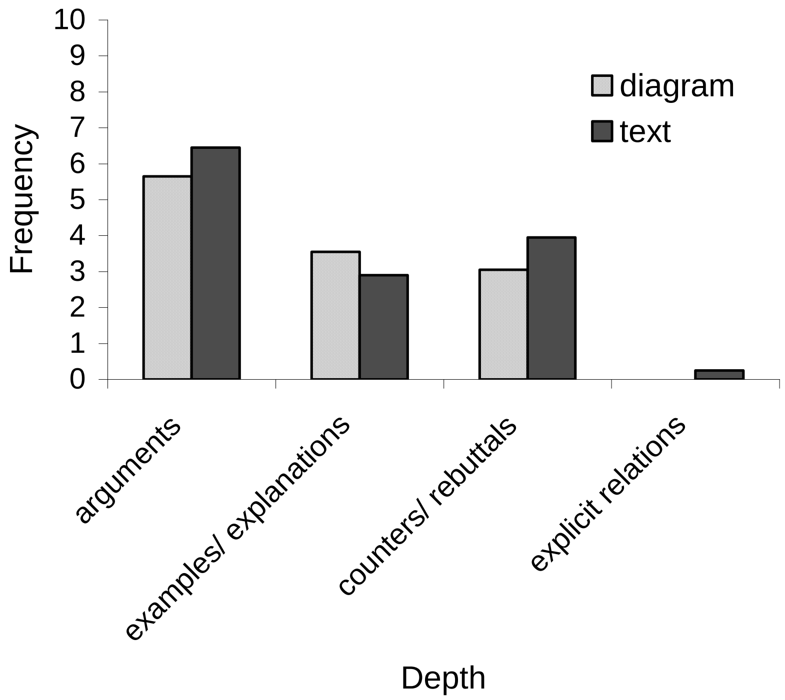


Figure 9. *Frequencies of depth of the space of debate before discussion, split for diagram and text condition.*

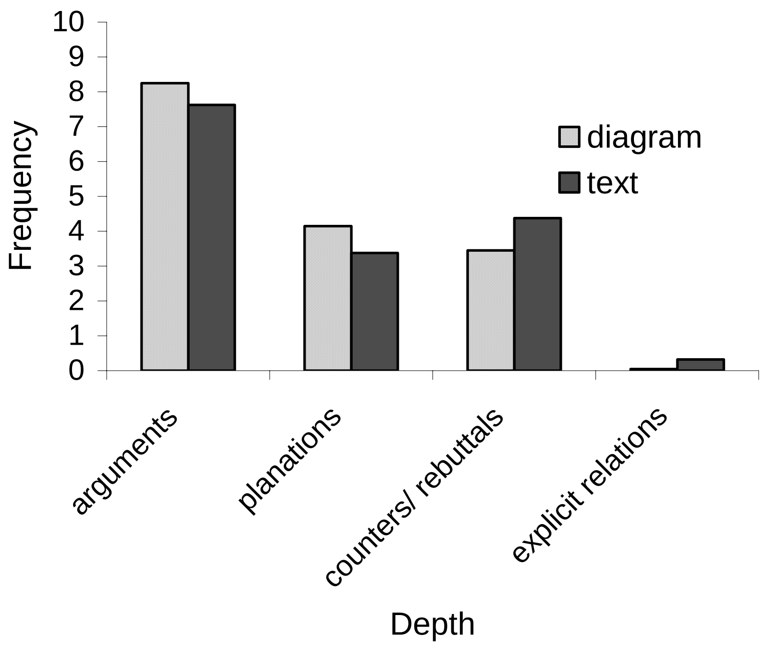


Figure 10. *Frequencies of depth of the space of debate after discussion, split for diagram and text condition.*

Although all students have individually made and revised either a text or a diagram, there was an extra condition of a given diagram made from a student's text by the researcher during the collaborative discussion and writing phase. Working with the given diagram might have affected the revised texts in a different way than working with the self-made texts. Therefore, we also performed an analysis in which the revised texts were split: revised texts from students who have inspected their own text, and revised texts from students who have inspected a given diagram from their text. T-tests showed no significant differences in *depth*, except for the number of arguments, $t(38) = 2.09$, $p < .05$, with the number of arguments in the revised text from the students who inspected their own text being higher ($M = 8.60$; $SD = 2.72$) than from students who inspected the given diagram ($M = 6.65$; $SD = 3.16$). This difference was not present in the preparation phase. The breadth was also significantly different for revised texts, $t(37) = 2.60$, $p < .05$ between students who inspected their own text ($M = 7.55$; $SD = 1.87$) and students who inspected the given diagrams ($M = 5.89$; $SD = 2.11$). Unfortunately, this difference was already present in the preparation phase, indicating an unwanted difference in students' placements into conditions.

In line with our expectations all students learned from the discussion and collaborative writing. We found that the self-made diagrams were deepened more than text after discussion compared to before discussion. This was mainly due to the increase in arguments. Students in the diagram-condition learned the most, and students in the given diagram condition the least ($d_{autodia} = 1.4$; $d_{dia} = 2.0$).

Discussion

This study began with the premise that supporting collaborative argumentation-based learning (CABLE) with diagrams would be beneficial for learning. Research on diagrammatic support for CABLE left us with questions about exactly when and how to use diagrams. Below, several important factors are uncovered to answer these questions.

Diagrams and texts

The benefits of diagrams for CABLE as described in literature have been found in our study. Diagrams have functioned as input for the discussion phase, and gave rise to a broader and deeper discussion. In this sense, they both stimulated and guided CABLE (Suthers, 2003). The diagram students made before discussion helped to maintain focus during discussion (Veerman, 2000). For example, when students were finished talking about one topic, or when they did not know what to write anymore, they looked at their diagram to find something new to discuss or write about. Students also sometimes reflected on alternatives when discussing the representations (Kolodner, 1996). However, all these findings were not only seen with diagrams, but also with text, indicating that the benefits of a diagram for CABLE are generalisable to texts as well.

No differences between diagrams and texts were found in the way students individually explored the topic of GMOs in breadth and depth before they engaged in discussion. The non-linearity of boxes and arrows in a diagram did not lead to a broader diagram, nor did the narrative of text result in a deeper text. The fact that there were no individual differences in breadth and depth between the conditions does make them more comparable for the collaborative phase.

We have to be careful in interpreting the above-mentioned findings. Argumentative diagrams are, unlike most diagrams, verbally oriented, display a lot of information, and are not very spatial (see Lohse, Biolsi, Walker & Reuter, 1994, on classification of visual representations). Although the very essence of diagrams is supposed to ensure a non-temporal, non-causal way of thinking in premises and conclusions, we suggest that students designed the diagram in a narrative way. They focused on the content of the boxes, less on structure of the diagram, or on relations between the boxes. This enhances the similarities between texts and diagrams. Whether it is the nature of argumentative diagrams or students' lack of experience to work with them that causes these similarities needs to be further explored.

The similarity of texts and diagrams in terms of breadth and depth of the space of debate implies that students could write an argumentative text equally well as construct an argumentative diagram when they are asked to individually represent their own space of debate. However, the students who constructed and revised a diagram before and after discussion broadened and deepened their space of debate more than the students who wrote and revised a text. This result implies that students in the diagram-condition have learned more during discussion and writing collaboratively. The discussion about diagrams may have stimulated students to gather more arguments. It may be easier to collaboratively broaden and deepen the topic of GMOs when inspecting diagrams than when inspecting texts, because the diagrams give a quick overview and are easier to compare and refer to. Another possibility is that students in the diagram-condition learned from the translation they had to make from the collaborative text to revision of the individual diagram (Ainsworth, Bibby & Wood, 2002). The cognitive effort that is involved in translation from a collaborative linear text to an individual two-dimensional diagram might have been what triggered learning to take place. The function of the individual constructed representations as inspectable representations during collaboration will be discussed in the next section.

Construction and inspection of diagrams

Construction and inspection of diagrams clearly served different functions in our study. Students constructed diagrams individually in phase 1 to indicate the extent of their space of debate. Although students' knowledge of the space of debate need not be complete in their representation, the activity of constructing an argumentative representation can support and shape students' reasoning (Bell, 1997; Stern, Aprea & Ebner, 2003). The inspection of the diagrams during collaboration in phase 2 supports this finding. Dyads inspected their representations to reason from it. The construction of the diagram thus helps explication and activation of knowledge, while the inspection of the diagram is good to support the discussion.

In this study, students inspected a text or diagram they made themselves, or a given diagram that was made by the researcher from a text they wrote themselves. Students inspected these representations in all conditions as frequently, but the inspection of the self-made diagram led to the most broadening and deepening of the space of debate, especially during collaborative writing (as opposed to chatting together). Sometimes (e.g., in the revision phase, and in chat) the constructed diagrams provide better results than the texts, indicating an effect of kind of representation that is constructed, while other times (e.g., in the collaborative writing task) the constructed diagrams provide better results than the diagrams made for the students, indicating an effect of constructed versus presented representation. Active construction of a diagram is important; the given diagram made from text did not lead to a broader and deeper space of debate. The text students made themselves and the diagrams they received seemed to be two unrelated things for them. Only once in our data did we see a student saying to his partner: "look, they made nice diagrams from our texts!" All others did not talk about the texts being transformed into a diagram. Moreover, the diagrams were sometimes referred to as text ("your text is good, how is mine?"). This indicated a focus on the sentences (content) in the boxes, and not on the structure of the diagram. We will return to this finding below.

Construction of diagrams cannot be easily compared to inspection of (given) diagrams, because the construction was an individual activity, while the inspection was a collaborative activity. In another article, we have compared the two kinds of diagrams in the same task, i.e., construct a diagram together versus inspect a diagram together (Munneke, van Amelsvoort & Andriessen, 2003). We found that the collaboratively constructed diagram led to more task management, aimed at constructing this diagram together. The inspected diagram led to more discussion on opinions, probably because the differences in opinion are clearly visible in the diagrams (Baker, 2003). Students also broadened the space of debate more when inspecting diagrams. However, it is hard to say how construction of a diagram *supports* the discussion, because the discussion is not an activity solely done in chat anymore, but partly shifts to the diagram (Suthers, Girardeau & Hundhausen, 2003; Van Drie et al., 2005).

How students make use of argumentative diagrams

The three representations in our study (diagram, text, and given diagram made from text) did indeed lead to differences in broadening and deepening the space of debate, but different approaches to the task were at least equally important. We identified three kinds of strategies dyads used in their exploration of the space of debate. Mountains had a broad and deep discussion in chat, but not in writing; Valleys had a shallow discussion in chat, but a broad and deep in writing; and Rising Slopes had a medium discussion and a somewhat broader and deeper writing. Difference in strategy proved to be related to difference in the use of the individual representations. Dyads that used their representation as information source for their collaborative text showed a shallow debate. Dyads that used their representations as a starting point for discussion showed a deep discussion, and went beyond what was written in their individual representations to collaboratively construct or transform that knowledge. Apparently, students' strategies interact with the affordances and constraints of different representational tools in determining the extent to which they collaboratively explore the space of debate, irrespective of the format of the representation (although we saw more Rising Slopes in the diagram condition than in the other two conditions).

In general, students' collaborative writing did not differ much in terms of breadth and depth, but the chats were very different. Some dyads discussed GMOs at length, while others used chat only to manage the writing task. Chat seems to be important for the amount and complexity of knowledge construction, but it does not ensure a broader and deeper argumentative text. With chat the topic can more easily drift, not only in analogy, but also because the topics drift off the screen (into the conversation history). With collaborative writing, students constantly work together on the same artefact, which may encourage convergence more. Additionally, in the development of our task we started from the assumption that students first discuss the topic of GMOs in chat, and then write a text about it. We expected that the extent to which the topic was discussed in chat would determine the breadth and depth in text (while also taking into account that writing the text might broaden and deepen the space of debate even more). This has probably been a false assumption. We did not find any correlations between chat and writing. A deeper space of debate does not necessarily produce a deeper text. It is very well possible that students see chat and text as two unrelated tasks, or that they choose to communicate via either chat or text and don't see the need to use both. Research on multiple representations shows that every representation has its own affordances and constraints, and will be used for different (sub)tasks (e.g., Ainsworth, Bibby & Wood, 2002; Grawemeyer & Cox, 2005). For future research, we would like to have students discuss and use representations without having to write a collaborative text. This way they might be focused on co-construction during chat instead of being focused on finishing the writing-task. Every task activity has to be meaningful to students. It appears that a task that is supposed to support another task is not meaningful in itself for students.

Collaborative argumentation-based learning with diagrams

Argumentative effort in chat and writing leads to broadening and deepening the space of debate. External representations such as a diagram support exploration of the space of debate by providing a basis to talk about, a way to focus on differences between partners, or an information source to tell from.

It is not the case that a diagram supports broadening, while a text supports deepening of the space of debate, and a given diagram support both. Rather, the diagram showed the best results in both broadening and deepening the space of debate. Broadening and deepening are not separate activities (at least not for the students); they both contribute to elaboration of the space of debate.

Our study suggests that a given diagram is not useful for learning. The translation from the textual representation to the diagrammatic representation is made for the students, while our results imply that the *construction* of the diagram is important to support learning. This is much more in line with constructivist theories, in which people learn by actively constructing knowledge instead of passively acquiring it (e.g., Bruner, 1990, Von Glaserfeld, 1989). Another explanation is that the given diagram distracts students. This is not very likely, however, because students did not mention the fact that they had received such a given diagram. We suggest that a given diagram (either generated by a person or computer program) may only be helpful for learning when students have to actively engage with them, and can claim their authorship.

Close inspection of how students used the diagram when exploring the space of debate showed that students do not go beyond the local level of diagram-inspection much. Students explore only to a certain extent. The advantage of a diagram over a text or a given diagram was mostly found in the amount of arguments and the amount of examples and explanations given. This means that the diagram deepened only in a 'shallow' way, with arguments and examples (the first and second category of depth), but not with counterarguments, rebuttals, or explicit relations (the third and fourth layer of depth). It appears that almost all students fail to look at the diagram in a global way, that is, they do not benefit from its overall structure. There are relations in a diagram of course, as there are in text, but we do not see arrows that specifically relate different subtopics or arguments to weigh these. Those explicit relations are necessary to get a grip on the space of debate and to reach conclusions. Instead, students discussed or wrote about each part of the diagram (or text) separately. More research is needed into students' ability and motivation to pay attention to structure and relations in diagrams. They need to be made aware of the possibilities a diagram has. The positive effects we found for the diagram suggest that it can be a better tool for CABLE than the text or the given diagram, but the differences we found in strategies suggest that these effects could be much higher.

Our results imply that representational guidance (whether constructed or inspected) is not a matter of 'plug and play'. Using a representation in CABLE does not automatically lead to broadening and deepening the space of debate, because dyads work very differently with the representations. Representations are used to put information directly from representations to collaborative text, to compare viewpoints and arguments, or to transform the information from the representations into collaborative knowledge during discussion.

For research this means that frequencies don't tell us everything about when and how to use diagrams for CABLE. Much more subtle processes, such as dyads' usage of diagrams, contribute to the results students can obtain. Research into tool support should therefore address not only effects, but also processes of tool use.

For school practice, our results give rise to the question of how to get students to change strategies. Is it possible to help students make optimal use of the tools? One possible solution is to restrict tool and task in such a way that only certain strategies are supported. However, a tool and task that can be employed to support different strategies, topics, and levels of expertise best supports argumentation-based learning in different situations and for different students. This greatly improves usability in schools. Another idea is to put more emphasis on reflection. After using the tool, students can be asked to look back at their collaboration and tool use, and comment on it. This might enhance students' insight in the affordances of the tool. The teacher then plays a very important role in guiding student's collaborative argumentation-based learning with tools.

At the moment, dyads explore the topic only to a certain extent. They sometimes seem to use the available tools as if they try to paste pieces together with a hammer. When they start using their tools in a more effective way, they could make a beautiful knowledge-artefact together.

Appendix A: Diagram made by the researcher from student's text

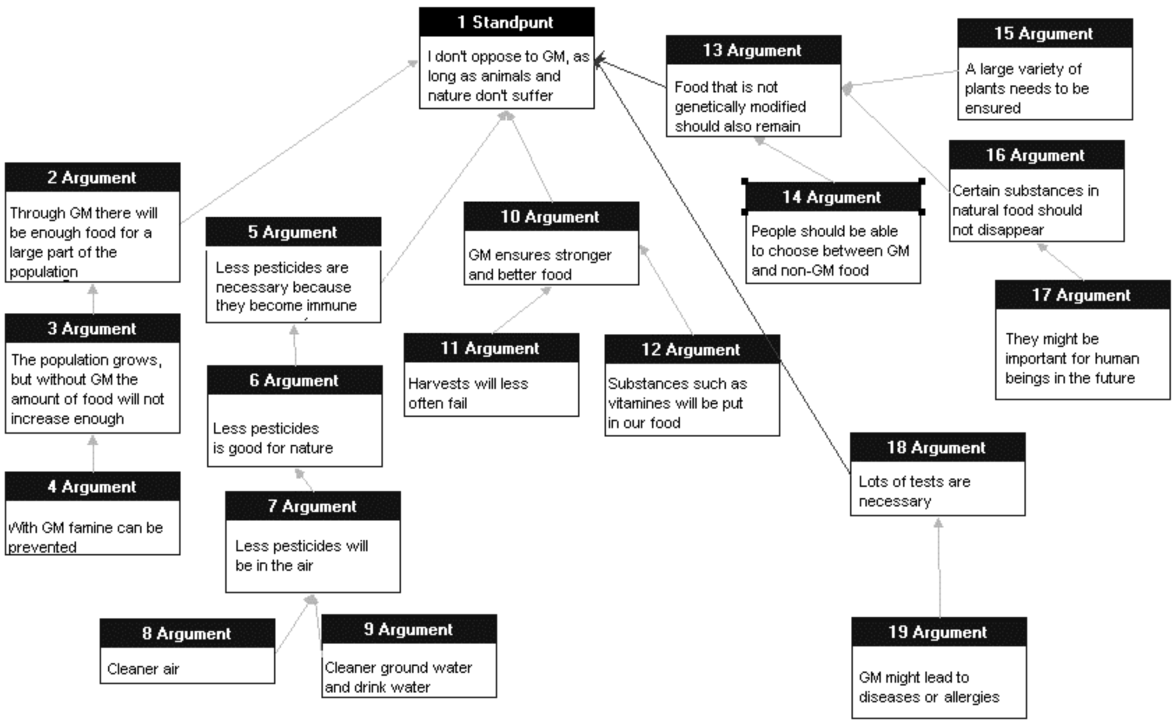
I don't oppose to genetic modification of food, as long as animals do not suffer from it, and nature won't be damaged.

By genetically modifying food there can be enough food for a large part of the population. The population continues to grow but without genetic modification the amount of food will not increase enough. Thus, there won't be enough food in the long run and there will be famine.
Genetic modification will also ensure stronger and better food. Harvests will less often fail and there will sometimes also be other substances (mostly with a positive influence) in the food, such as more or other vitamins.

However, food that is not genetically modified should also remain to exist, in order for people to be able to choose what kind of food they want, and to ensure that certain substances that are not present in genetically modified food, but are present in natural food, will not disappear. Because they might be important for human beings in the future.
It is also important that certain kinds of plants will not disappear at all, because a large variety of plants need to be ensured.

If it is true that fewer pesticides are necessary because of genetic manipulation, because they have become immune, then this is better for nature and that is a large plus. Less chemical substances will be in the air, and this ensures a cleaner air, and thus also cleaner ground water and drink water.

I do think that good tests with the food need to be done first to see whether it is harmful, whether it cannot lead to allergies or diseases.



Appendix B: Broadening and Deepening scores

Condition	Dyad	Type	Representation before		Chat		Text		Representation after	
			depth	*breadth*	*depth*	*breadth*	*depth*	*breadth*	*depth*	*breadth*
Diagram	1	1	18.5	5.0	73	9	43	9	21.0	5.5
	2	1	22.0	4.5	103	13	40	7	23.5	5.0
	3	3	13.5	4.5	39	7	48	11	23.0	6.5
	4	2	32.5	9.0	4	7	73	11	33.0	9.0
	5	3	32.0	8.5	29	5	60	11	33.5	9.5
	6	3	30.5	8.0	36	8	57	8	35.0	8.0
	7	1	26.0	7.5	96	12	65	11	31.5	8.5
	8	3	23.0	6.5	37	8	62	9	31.0	7.5
	9	5	15.5	7.0	14	10	52	9	22.0	8.0
	10	3	5.5	3.5	48	9	51	9	17.5	8.5
Text	11	4	44.5	7.0	32	9	27	6	46.0	7.5
	12	2	24.5	6.5	5	5	25	6	32.5	7.5
	13	2	26.0	5.5	4	3	38	9	32.5	7.5
	14	1	29.0	8.0	81	9	56	5	32.0	10.0
	15	1	32.5	7.5	51	11	35	7	42.5	8.5
	16	3	36.5	8.5	39	8	41	10	39.5	8.5
	17	3	29.5	7.5	59	8	51	11	37.0	8.5
	18	1	16.5	6.0	75	9	30	7	17.0	6.0
	19	3	22.5	6.0	28	10	31	9	23.0	6.5
	20	5	23.5	4.5	48	6	38	6	25.5	5.0
Given diagram	21	2	24.5	5.5	14	5	47	6	29.0	6.0
	22	5	22.5	4.5	49	7	45	8	23.5	5.5
	23	2	28.0	6.0	0	1	46	11	28.5	7.0
	24	3	21.0	6.5	22	6	63	8	30.5	8.5
	25	2	9.5	2.0	3	3	48	11	9.5	2.0
	26	1	25.5	5.0	45	9	12	7	32.0	6.0
	27	1	27.0	5.5	48	8	30	6	29.5	7.0
	28	3	18.0	5.0	31	6	38	10	21.5	6.0
	29	5	20.0	4.0	71	7	33	7	22.5	4.5
	30	1	21.0	4.0	54	9	18	5	22.0	4.5

Note. Type: 1 = mountains; 2 = valleys; 3 = rising slopes; 4 and 5 = other types not taken into account for analyses.