Instructional Science **30:** 155–186, 2002.

Collaborative argumentation in academic education

ARJA VEERMAN[1]*, JERRY ANDRIESSEN[2] & GELLOF KANSELAAR[2]
[1]*TNO – Human Factors, Kampweg 5, 3705 LA Soesterberg, The Netherlands;*
[2]*Utrecht University, Dep. of Educational Sciences, Heidelberglaan 2, 3584 CS Utrecht;*
**e-mail: veerman@tm.tno.nl, website: http://www.tno.nl*

Received: 7 December 2000; in final form: 23 May, accepted: 14 September 2001

Abstract. The general purpose of this research is to discover principles for the design of educational tasks that provoke collaborative argumentation. The specific research question concentrates on the relationship between question asking and argumentation and is examined in three different collaborative learning tasks involving advanced university students. These studies aim at providing criteria for organising educational situations that elicit argumentation during which opinions change and new knowledge is being created, within constraints (course duration, exam criteria, student expectations) set by current higher education. We discuss some factors influencing argumentation (the role of the student, peer, tutor, task, instruction and medium) and specific attention is paid to question asking. Then we report three studies conducted at our educational department. These studies involve comparable students, a similar domain, but differ in many other respects: the mode of communication (oral, typewritten), the presence of the tutor, instruction on argumentation and/or question asking, assigned task goals (competition, consensus), and the type of required outcome. Each study reveals prominence of different types of questions and question generation mechanisms. In addition, the relations found between question asking and argumentation change between studies. In comparing and interpreting these studies, we discuss results in the light of provoking collaborative argumentation in regular academic learning situations.

Keywords: academic education, argumentation, collaborative learning, computer-supported collaborative learning, instructional design, question asking

Introduction

In this article we report three related studies about university students that argue in different tasks that are all part of the same curriculum on learning with computers. We assume that there is an important role for argumentation in academic learning. This assumption rests on two reasons. First, under certain circumstances, argumentation can be a tool for fostering reflection and deep thinking (see e.g., Andriessen & Coirier, 1999). Second, academic knowledge cannot be understood as something fixed that can be transmitted from experts to students. To participate fruitfully in academic discourse, it is crucial for students to understand the nature of scientific knowledge as a process of permanent negotiation (Petraglia, 1997). Hence, we see argumen-

tation as a tool to promote critical thinking as well as an essential quality of the discourse-to-be-acquired in academic education.

Inspired by such thinking and by the book "Rethinking University Teaching" (Laurillard, 1993), tutors engaged in the curriculum of Educational Sciences at Utrecht University started to rethink their academic teaching. Under the motto "teach what you preach", traditional learning activities such as studying from textbooks for exams and listening to lectures were substituted by more student-centred activities. Tutoring sessions, collaborative learning projects and electronic discussions in which students could actively ask, share, refine, explain and discuss knowledge, were organised as central learning activities in the curriculum. In this article, we explore the effect of some of these activities on students that have limited experience in argumentation and collaborative learning. In their formal educational careers, many students have developed a pragmatic attitude when it comes to learning. To pass exams is often more important than reaching deep understanding or gaining new insights (Veerman, Andriessen & Kanselaar, 2000). It is important to know how regular students behave in argumentative tasks as these students represent the default case in many educational systems.

We report the results of three instructional interventions in which students are asked to perform argumentative tasks in different settings; the third situation is an electronic one. These three studies form a temporal sequence. Although they involve the same students, no systematic comparisons at that level are intended, as experimental design between the three studies was not strived for. The purpose is to discover a number of principles for the design of educational tasks that provoke collaborative argumentation in actual practice. In this context, the relationship between question asking and argumentation is specified and closely examined.

The studies we report on can be regarded as part of our research about the role of collaborative argumentation in computer-supported learning (Veerman, 2000). Not enough is known about how argumentation can be used in 'normal' academic settings, what the effects are on students' learning, what characteristics of the environment affect this learning and how students should be coached, given practical constraints of available student and tutor time.

The remainder of the article is organised in the following way. First, we present our perspective concerning the role of collaborative argumentation in academic learning and the role of question asking to provoke argumentation. Then, we describe important contextual factors that affect collaborative argumentation in education: the role of the tutor, (peer) students, task design, instruction and the medium. We will then describe the three studies in sequence, and propose a common system for analysis of the

resulting data. After we discuss the results, we will tentatively draw conclusions on principles for the use of collaborative argumentation in academic learning.

Factors affecting collaborative argumentation in education

From a rhetorical perspective on academic learning, education can be framed as an ongoing argumentative process (Petraglia, 1997). It is the process of discovering and generating acceptable arguments and lines of reasoning underlying scientific assumptions and bodies of knowledge. From a socio-cultural perspective on education, students should acquire practice and expertise in this activity, through sustained and, to some extent, guided efforts in meaning negotiation. However, little is known about what it takes to make such things succeed.

Question asking

As a first step, we thought about what methods would trigger argumentation in collaborative learning situations. One main topic that came up was question asking, since this can be viewed as a fundamental component in collaborative learning situations (Graesser, 1993). Question asking can trigger students to check each other's information and to provide explanations and justifications (Webb, 1995). King (1990) found that especially higher-order questions elicit explanations, in which justifications could be enclosed. Such questions include for instance open-answer questions, deep-reasoning questions aimed at causes, consequences and/or expectations and goal-oriented questions.

Graesser, Person & Huber (1993) proposed to not only analyse different *types* of questions but also the so-called "question generation mechanisms". In other words: what is the reason that a question is asked? For instance: a student can ask a verification question to infer knowledge but also to monitor common ground. Question generation mechanisms can give a better insight in causes and effects of question types since the context in which they are stated is taken into account.

According to a number of models in cognitive science, question generation is a fundamental component in cognitive processes that operate at deep conceptual levels, such as comprehension of text and social action, the learning of complex material, problem solving and creativity. However, to fruitfully ask good questions students may need explicit training (King, 1990). "It is well documented that student questions in learning are very infrequent and unsophisticated (...). That is, they are normally shallow, short-answer questions that address the content and interpretation of explicit

material. They are rarely high-level questions that involve inferences, multi-step reasoning, and the application of an idea to a new domain of knowledge, the synthesis of a new idea from multiple information resources or the critical evaluation of a claim" (Graesser & Person, 1994, pp.105–106). Students have difficulties to identify their own knowledge gaps, they are afraid of asking "bad" questions, and do not like to confront each other or a tutor.

A descriptive framework

We hoped question asking could provoke students to engage in meaningful argumentation. In that case, we would have data to support the need for instruction about question asking, especially high-level questions. For question asking to serve that role, we needed more information concerning the relations between question asking and argumentation in our own collaborative learning situations. We wanted to know what types of questions are asked and which are more likely to trigger argumentation, and to what extent this relationship is sensitive to the situation and the type of instruction. To describe more systematically what we considered as important situational variables, we use a general descriptive framework that incorporates the main factors affecting a collaborative learning situation (Figure 1). These factors (and many others) interact in complex ways, and at this stage we do not intend to present an integrated account of all related factors. We used this framework to describe the variations in task design of the three studies reported in this paper. The factors we address are described in the sections below.

The student

We wanted our students to ask each other high-level questions in order to engage in meaningful argumentation. Research suggests that students argue when doubt or disbelief arises, with respect to their personal attitude strength (Krosnick & Petty, 1995). In argumentation, attitude strength can bias the evaluation of scientific information: evidence supporting one's attitude is seen as more compelling than evidence that disagrees with it (Lord, Ross & Leppler, 1979). The tendency towards favouring a personal position and attacking the opposition, thereby neglecting the possible plausibility of the opposition or flaws in the personal stance has been found in many studies (e.g., Hightower & Sayeed, 1995; Stein, Calicchia & Bernas, 1996). This type of biased behaviour limits the negotiation space and, therefore, hinders knowledge transformation. In the second and third study, we instruct our students to be aware of multi-sided argumentation and to ask each other critical questions.

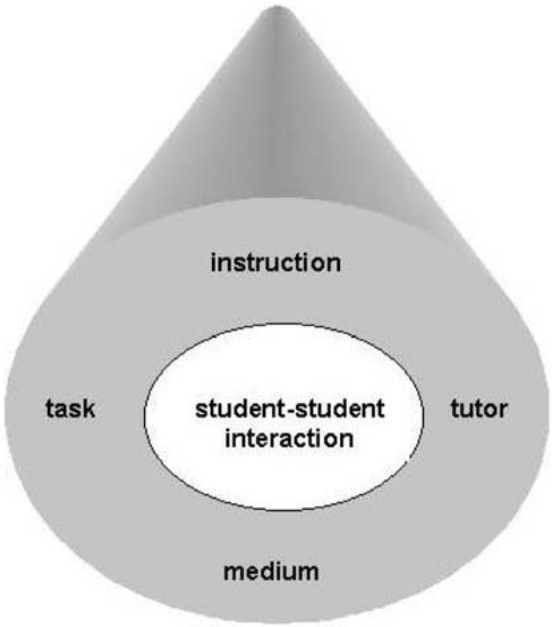

Figure 1. Factors affecting student interaction in a collaborative learning situation.

The peer

Interacting with fellow students can make learning realistic and relevant, both of which are seen as critical conditions for a learner to seriously consider alternatives (Petraglia, 1997). However, especially in authentic learning situations politeness strategies may inhibit students to engage in critical discussion. To prevent students from being "too nice" to each other, one of the possible interventions is to give them competitive roles, for example to let them defend (predefined) conflicting positions (Stein & Miller, 1991; Stein, Calicchia & Bernas, 1996). Such an intervention may distance students from routine conversational rules and trigger argumentation more easily. In the second and third study we considered this aspect in our design.

The tutor

In education the role of the tutor can be crucial. Tutors may signal misconceptions, model students' behaviour by breaking down a task, making abstract situations concrete or vary the context in which problems occur. The tutor can do all of this, but the main point is that the students should do this themselves. If students rely on their tutors too much, they will acquire dependencies instead of responsibilities. It has been found more than once that students do not tend to ask many questions with a tutor present (e.g., Graesser & Person,

1994). In study 1, the tutor acts as an expert trying to discuss the students' conceptual problems. In the other two studies we examine argumentation without a tutor present.

The task
Elaborate argumentation is more likely to occur under specific task constraints (Coirier, Andriessen & Chanquoy, 1999; Quignard & Baker, 1999). To arrive at a shared conclusion or solution, collaborative argumentation should be a necessary activity. Tasks fostering meaningful argumentation should be open-ended, so students can share and learn as much as possible from each other's differences in prior knowledge, experiences, beliefs and values. In all three studies the learning task is open-ended. In study 3 students also have to construct a shared solution.

Instruction
One of our main problems was what kind of instruction is needed to trigger students' engagement in argumentation without explicit training or scripting of the situation. The problem was that training into question asking, collaboration, or argumentation is not a part of our regular curriculum. Since we decided to study argumentation in tasks that were part of the curriculum, the choices for research variables were partly based on what was possible in the particular course. In the first study we exploratively examined question asking and argumentation in regular tutoring sessions. In the second study we concentrate more closely on the concept of "asking questions". In the third study we examine if competitive roles for students prevent them from being "too nice" to each other and trigger collaborative argumentation (Stein & Miller, 1991; Stein, Calicchia & Bernas, 1996).

The medium
We use oral as well as electronic communication, obviously very different media for serving similar goals (Collins, Neville & Bielaczyc, 2000). Features of electronic tools such as a text-based dialogue history or graphical representations may be used for reflection and structuring text-based interaction (Veerman & Treasure-Jones, 1999). Compared to oral discussion, computer-mediated communication (CMC) is a "slow" mode of communication. Time delays allow participants to re-read and reflect on information and to more easily share multiple perspectives and attitudes relative to a particular topic or issue. In the third study students work with Belvedere, an electronic tool for text-based communication and graphical argumentation.

In this study, we aim at finding some principles for provoking collaborative argumentation in real academic learning situations, considering the factors mentioned above. Our approach is to first analyse argumentative

Table 1. Main factors in the design of the studied learning situations

		Tutor	Medium	Students' role	Instruction
Study 1:	Tutoring	Yes	F2F	–	–
Study 2:	Collaborative discussion	No	F2F	Competitive behaviour	Question asking
Study 3:	Electronic discussion	No	CMC	Competitive versus Consensual behaviour	Basic instruction on argumentation

discussions on questions asked and arguments given by the participants, in different educational situations within the context of the curriculum of our department. We use the same category system for analysing the discussions in all three studies so that comparison between the studies is possible. Our specific research question can be formulated as: *What is the relationship between question asking and argumentation in collaborative learning tasks in academic education?*

Three studies

In this section, we describe three studies that each involved small groups of undergraduate students who collaboratively worked on short, "real-time" tasks as part of an eight-week course in Educational Technology and Computer-based learning (CBL). In each task, students had to work in pairs or triplets on open-ended problems. With or without a tutor present, they were asked to externalise (incomplete) knowledge, beliefs and values and to use each other as a source of knowledge and reflection in order to reach a (shared) solution. In sequence of admission, the tasks were aimed at reaching the following goals:

(1) Evaluating the learning goals of a student-designed CBL program

(2) Developing insight into a theoretical framework

(3) Designing didactics for a student-designed CBL program

The tasks were respectively conducted as face-to-face (F2F) tutoring, F2F discussion, and electronic discussion. In Table 1, an overview is given of the main differences considering the factors that affect collaborative argumentation in the three studies (see also Figure 1). Before we present the studies in sequence of the data collection, we give details and an example of our system for analysing data on collaborative argumentation in each of the three studies.

Table 2. Category system for analysing argumentative fragments on question asking and argumentation

Categorisation system for separated utterances	Examples
(A1) Question types	
1. Goal-oriented question	*"Why have you chosen this subject?"*
2. Cause-consequence question	*"Why shouldn't the teacher's reactions lead to learning?"*
3. Evaluative question	*"Are you satisfied with the assignment?"*
4. Other open questions	*"Why?", "Can you give me a definition?"*
5. Verification question	*"Is that true?"*
6. Other closed questions	*"Is the student wrong or the tutor?"*
(A2) Question generation mechanisms	
1. Inferring knowledge	*"O yeh, why do you think that?"*
2. Correcting knowledge	*"What operationalisation do you use?"*
3. Monitoring common ground	*"If I understand you correctly, you say that principles are facts?"*
4. Other . . .	*"Any more suggestions?"*
(B) Argumentation	
1. Neutral argument	*"I don't know if the tutor's feedback is helpful or not"*
2. Positive argument	*"So, I am sure this is true because . . ."*
3. Negative argument	*"No, that is not correct. When that happens it means that . . ."*

Data analysis

All F2F sessions were recorded on audiotape (study 1 and study 2). The electronic sessions were logged automatically on the computer (study 3). All sessions were screened on the presence of **argumentative fragments**, defined as oral or written exchanges in which at least some doubt is expressed (in the form of a question or counter-argument) and at least one argument is given (Van Eemeren, Grootendorst & Snoeck Henkemans, 1995). Subsequently, all argumentative fragments were analysed in terms of individual verbal utterances defined as single messages exchanged between participants. Based upon the Verbal Observation System (Erkens, 1997), work on analysing argumentative text production (Andriessen, Erkens, Overeem & Jaspers, 1996) and the Question Categorisation System (Graesser, Person & Huber, 1993), each utterance in an argumentative fragment was coded as a type of question or argument (see Table 2).

For the analysis of question asking, we used the Question Categorisation System (Greasser, McMahen & Johnson, 1994). Based upon this system, we could code all questions on two dimensions: (A1) as question types (e.g., "open" or "closed" question) and (A2) as question generation mechanisms (underlying goals of questions). Thus, each question should be coded with one code from set A1, and one code from set A2.

Grounded on earlier, exploratory research we reduced the original system from 13 to 6 types of questions and from 20 to 4 question generation mechanisms (Veerman, 1996). Thus, each question could be coded on two dimensions, for example: a verification question ("Is this true?") could be aimed at monitoring common ground ("Is that true?" interpreted as: "Do I understand you correctly?") but also at inferring knowledge ("Is that true?" interpreted as: "Tell me why. I don't believe you"). We expected higher-order questions (e.g., open-answer questions, questions about goals, causes and consequences), aimed at reasoning processes (inferring knowledge) to be most effective for triggering argumentation (King, 1990).

Arguments (B) were coded as neutrally, positively or negatively related to the main claim or problem statement. Coding was based upon both content and linguistic features ("but", "and", "thus", "however", etc.). Positive arguments contained justifications, positive examples, specifications, explanations and refutations of contradicting information. Negative arguments were against the claim or in favour of an opposite of (a part of) the claim or problem statement. Neutral arguments contained continuations ("and"), disjunctives ("maybe or maybe not") and conceptual statements.

The inter-judge reliability score of the Question Categorisation System was 0.94 (based upon the amount of agreement), on both question types and question generation mechanisms (Graesser, Person & Huber, 1993, p.162). For practical reasons, we decided to only measure the inter-judge reliability score for categorising arguments ourselves (neutrally, positively and negatively oriented) by defining Cohen's kappa (Cohen, 1968). After two days of training, two independent judges scored each five new protocols from the first study. As a result, a Cohen's kappa of 0.77 could be established. For qualitative data that depends on the interpretation of categories, this score can be regarded as "substantial" (Heuvelmans & Sanders, 1993).

In the next section, we will show an example of an argumentative fragment analysed on question asking, argumentation, multiple perspective taking and balancing the argument.

Example

To give an example, we present an argumentative fragment collected in the first study on tutoring sessions (see Figure 2). In this example, three students

(S1, S2, S3) and a tutor (T) discuss the issue of offering help in a CBL program. The main claim (Claim 1) under discussion is: "help is offered when the student makes a mistake". The students describe what they think the help system should do, the tutor disagrees. The students finally convince the tutor of their opinion (which can be considered as non-prototypical!). Text and explanations of the argumentative fragment are presented and the analysed fragment is shown (Figure 2).

Nr.	Text argumentative fragment	Explanations
	CLAIM: "help is offered when the student makes a mistake"	
1	T: (reads from the student's paper) "Help is offered when the student makes a mistake". No, I don't think that is true.	Negative statement in relation to the claim
2	S2: Yes, it considers the ideas of Burton, who	The ideas of Burton can be considered as a justification for the claim
3	T: Yes? A help system is . . . , or isn't that true? The help in Word Perfect, that is pressing a help button, isn't it?	The tutor tries to find out what the students mean
4	S1: No, that is uhh . . . , a help system . . . is something placed in the menu bar.	Explanation within the argumentative fragment; can be considered as a neutral argument in relation to the claim
5	S2: I made some nice translations (from Burton's arguments), but maybe they are not correct.	Idem
6	S1: Do you really see the help as part of the menu bar?	Checking the other student's utterance
7	T: Yes, yes, that's what I thought. Well . . . , I don't think you have to copy this literally.	Argument against the claim: Burton does not have to be copied
8	S3: But the help system can have different types of explanations in different ways, can't it?	Evaluating the help of features of the system, aimed at correcting the idea that there may be only one type of help provided by the system
9	T: I mean, in Word Perfect is it obvious that uhh . . . . The other thing is, when a student makes a mistake, help must be triggered. Uh. . .	Argumentsing, not aimed at choosing a specific point of view
10	S1: (reads Burton aloud) "In the system help is available on request or during errors". So, during errors he says uhh . . .	Using Burton as a justification for the claim
11	T: "Upon request"? That is what I read first, . . . yes. Yes, so that's possible. Yes, uhh . . . ok. Help is offered when a student makes a mistake.	Agrees, thinks what Burton mentions is possible

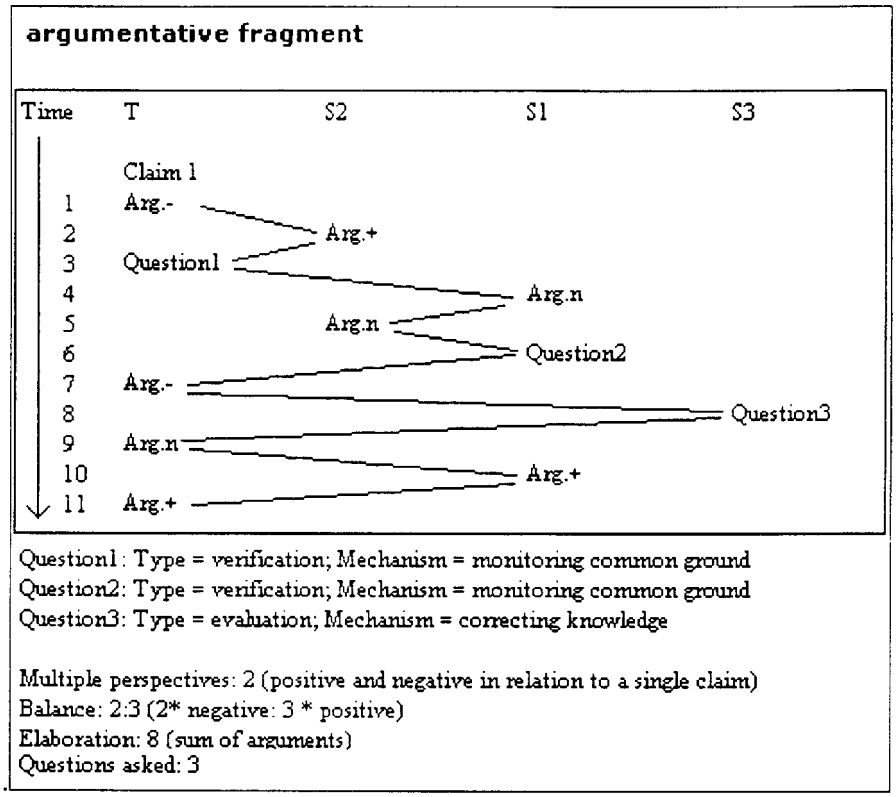

Figure 2. Example of an analysed argumentative fragment.

Study 1: tutoring sessions

Purpose

The study was aimed at the specific research question: *What is the relationship between question asking and argumentation in collaborative learning tasks in academic education?* To find any answers, we first needed to get more insight about the ways our own students (and tutors) argued in educational practice. Therefore, a pilot study was organised around natural, face-to-face (F2F) tutoring sessions. In tutoring sessions generally single students or small groups (dyads, triples) discuss educational issues with a tutor. While, in general, we did not expect much argumentation to occur, our specific interest concerned the nature of the questions and the question generation mechanisms that triggered argumentation.

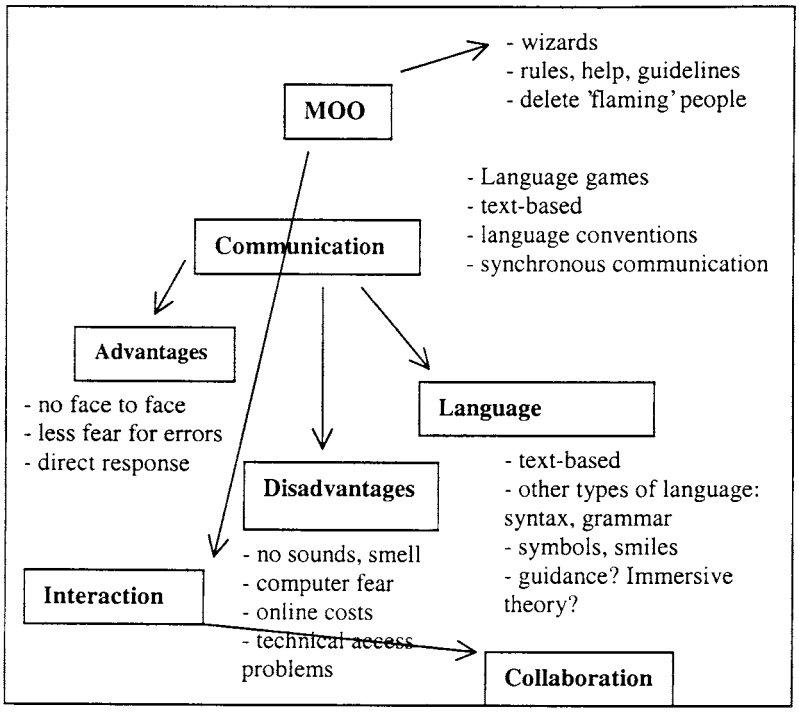

Figure 3. Example of a concept map, subject to one of the tutoring sessions.

Method

The study on tutoring sessions was implemented as an open-ended problem-solving task as part of an 8-week course on developing computer-based learning (CBL) programs. After an introduction to the course and a presentation of some principles for instructional design, 23 students organised themselves in 11 units of one, two or three students each (2*1 student; 6*2 students; 3*3 students). All students had reached a comparable level in Educational Sciences. The units organised were the same for the duration of the course.

The first assignment was to construct learning goals for a CBL program the student units had to design. To this end students were instructed to use concept map techniques (Novak, 1991) and to construct a plan on paper (size A1/ A2), in which the learning goals of the CBL program were described, interrelated, organised and justified (see Figure 3). In the second week of the course, this plan was subject to a one-hour tutoring session, in which the course tutor evaluated the feasibility of the learning goals by discussing

the students' assumptions within each student unit. Neither the students nor the tutor received specific instruction how to act or communicate. However, the students were informed that the goal of the session was to discuss their ideas, plans and problem solutions with the tutor. The tutor was expected to raise most questions and to trigger discussion. His interest was in coaching students to improve their design.

Results

The data analysis revealed that the tutoring sessions hardly contained any argumentative discussion. From the 11 one-hour tutoring sessions analysed, only 25 argumentative fragments (parts of a protocol of a few minutes each) could be gathered in which in sum 72 questions were asked (mean ≈ 3 questions per fragment) and 767 argumentative utterances were produced (mean ≈ 31 arguments per fragment). Since single fragments were too small and varied too much in terms of the number of contributions to use for separate statistical analyses, we decided to collapse all 25 fragments, to look at the questions and arguments produced by the tutor and students.

With respect to the question asked, Table 3 shows that the tutor asked most questions (72%), of all types, thereby mainly aiming at correcting and inferring knowledge from the students. The students, however, mainly asked verification questions and other types of closed answer questions. They were particularly focused at reaching common ground. Considering argumentation the tutor produced more negative arguments than positive ones, while the students produced the same amounts of both types of arguments.

Across all 25 argumentative fragments, we measured the relations between question asking and argumentation. Since all categories had a skewness and/or kurtosis > 1 (including both students' and the tutor's expressions), we used Spearman's nonparametric measure of correlation (Baarda & De Goede, 1989). In Table 4 it is shown that question asking in general and some specific question types and generation mechanisms were significantly related to the incidence of arguments, particularly to negatively oriented arguments. In addition to open questions, verification and cause-consequence types of questions, questions aimed at correcting knowledge and at monitoring common ground showed significant relationships with argumentation.

Discussion of study 1

First, as was to be expected, students and tutor acted according to different roles, which are reflected in the nature of the questions and the question generation mechanisms by each party. The tutor asked most questions, and these questions aimed at inferring or correcting ideas, rather than at moni-

Table 3. Frequencies of question types, generation mechanisms and argument types in 25 argumentative fragments

Categorisation system for separated utterances	Tutor	Students	Total
(A1) Question types			
1. Goal-oriented question	5	–	5
2. Cause-consequence question	2	–	2
3. Evaluative question	9	–	9
4. Other open questions	13	–	13
5. Verification question	12	11	23
6. Other closed questions	11	9	20
Total number of questions	52	20	72
(A2) Question generation mechanisms			
1. Inferring knowledge	20	–	20
2. Correcting knowledge	20	–	10
3. Monitoring common ground	7	16	37
4. Other . . .	5	4	5
Total number of questions	52	20	72
(B) Argumentation			
1. Neutral argument	20	18	38
2. Positive argument	161	183	344
3. Negative argument	203	182	385
Total number of arguments	384	383	767

toring common ground. The students only asked closed questions, serving to monitor common ground. Second, with respect to the relation between question asking and argumentation, we found that almost any question type correlated with some form of argumentation, except evaluative questions and other closed questions. The generation mechanisms of correcting knowledge (by the tutor) and monitoring common ground (often by the students) correlated significantly with argumentation as well.

Questions (by the tutor) aiming at inferring knowledge did not correlate with argumentation. This is unfortunate, because one of the main reasons for our study was to see to what extent students would transform their knowledge as a result of collaboration in tutoring sessions. Argumentation did not serve this goal here. It looks like that the best a tutor can do is aiming for correcting knowledge and having some argumentation from that. Students

Table 4. Significance of correlations between question asking and argumentation in study 1

Question asking/ Argumentation	Sum of arguments	Negative arguments	Positive arguments
Sum of questions	*	*	–
(A1) Question types			
1. Goal-oriented question	–	*	–
2. Cause-consequence questions	*	–	–
3. Evaluative question	–	–	–
4. Other open questions	*	*	–
5. Verification question	*	–	–
6. Other closed questions	–	–	–
(A2) Generation mechanisms			
1. Inferring knowledge	–	–	–
2. Correcting knowledge	*	*	–
3. Monitoring common ground	*	*	–
4. Other . . .	–	–	–

*: significant relation: $p < 0.05$; –: $p > 0.05$.

uniquely focus on the tutor's ideas and try to find out what he means. The fact that they did not ask any open question can be taken as indicating poor skill or motivation to really find this out. It could be that students did not prepare sufficiently or that they lacked the skills to critically question and effectively oppose the tutor. Small group tutoring is an expensive form of education, in terms of tutor time, and we found it not very effective here.

To overcome some of the problems just mentioned, we decided to organise a more egalitarian collaborative learning situation (F2F). First, we wanted the students to show a stronger attitude in defending positions. Therefore, we decided not to bring the tutor back on stage and, moreover, we provided the students with predefined conflicting claims and a competitive role (to win the discussion!). Second, students were provided with instruction on critical question asking (particularly asking open questions and verification questions aimed at correcting and inferring knowledge). We expected the students to be encouraged to resist critique, to discuss their positions and to engage in multiple-sided argumentation.

Study 2: collaborative learning sessions

Purpose

In line with the first study, we developed a second study that specifically aimed at the following research question: *What is the relationship between question asking and argumentation by university students in a competitive collaborative learning task and with instruction about asking questions?* We based our design on the results of the first study and organised a collaborative learning situation (F2F), without a tutor present. Our main goal was to provoke students to engage in argumentation by competitive instruction and information about critical question asking.

Method

We developed two short (open-ended) argumentative discussion tasks as part of an 8-week introductory course on Educational Technology for undergraduate students. For several years, this course has focused on the book "Rethinking University Teaching" (Laurillard, 1993). The book is used as a theoretical framework to discuss affordances of media applications in higher education (hypertext, simulations, CMC, etc.). The book centres on a discussible "conversational framework" (Bostock, 1996) that describes crucial activities necessary to complete the learning process in teaching-learning dialogues.

In order to stimulate insight into the implications of this framework, we developed two 10-minute argumentative discussion tasks in the third week of the course. Students with a comparable background in Educational Sciences were required to prepare two chapters of "Rethinking University Teaching" (Laurillard, 1993) at home about the role of (A) feedback and (B) tutoring strategies in tutoring sessions. Before the meeting, they had to take an individual knowledge test, in which they were tested on their descriptive knowledge of concepts and activities mentioned in the conversational framework. During the meeting, they were randomly paired and instructed to competitively discuss a protocol of a student-tutor dialogue. In two 10-minute sessions they were asked to defend controversial claims in relation to the protocol and to win the discussion. The claims were:

> (Ad A) "The tutor provides feedback that improves / does not encourage the student's learning process"
> (Ad B) "The tutor's strategy improves / does not encourage the student's learning process"

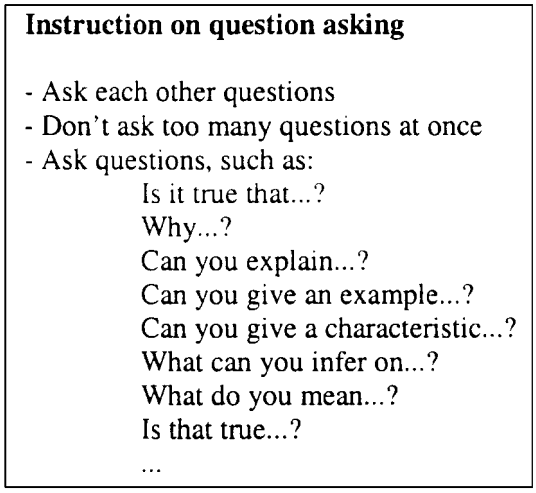

Figure 4. Instruction on question asking.

The use of controversial claims together with the instruction on competitive behaviour (win the discussion!) was expected to provoke students to engage in argumentative discussion. In addition, we explored the effects of providing student groups with a short instruction on critical question asking as shown in Figure 4. We expected such groups to be triggered to ask more critical questions than groups that did not receive instruction.

In our practical setting 24 students were available to participate in the study. In order to gather enough data for analyses – specifically with respect to the question asking intervention – we organised the study as is presented in Table 5. The 24 students were randomly assigned to twelve pairs. Subsequently, six pairs engaged in discussion task A (group 1), six pairs engaged in task B (group 2). Then, the students randomly switched partners within a group, and formed new pairs. All pairs then received instruction on question asking and the groups switched tasks.

At the individual level of analysis, the same students were randomised, thus, some order effects could not be ruled out. However, at the group level of analysis the second student pairs formed could be treated as being new and independent units from the first pairs organised.

Results

The knowledge tests showed that students did not prepare very well as only 36% of the students passed the test. We analysed the data in the same way as in the formerly described study on the F2F tutoring sessions. First of all, we

Table 5. Organisation of study 2

Discussion task	*Set of 24 individual students*	
Phase 1	No additional instruction	
	– task A	6 student pairs (group 1)
	– task B	6 student pairs (group 2)
Phase 2	Instruction on question asking, switch partners within groups	
	– task A	6 student pairs (group 2)
	– task B	6 student pairs (group 1)

gathered all 24 collaborative learning sessions, which could all be considered as complete argumentative fragments. A fragment in the current study also comprises a complete session. One of the student pairs accidentally did not receive instruction in the second session. We had 13 sessions without instruction on question asking and 11 sessions with instruction. Secondly, within these argumentative fragments we analysed separated utterances on question types, question generation mechanisms and argument types. In addition, we measured argumentative fragments on multiple perspective taking, balanced argumentation, the total number of arguments and the sum of questions asked by the use of T-tests (see Table 6).

The overall results showed that the 24 argumentative fragments included in sum 1422 statements (mean = 58 per fragment) and 156 questions (mean = 6.5 per fragment). Compared to the study on tutoring sessions, twice as many questions were asked in 1/6 of the time available (10 minutes versus 1 hour). Students asked each other many verification questions, but also many open questions. Questions were mainly aimed at inferring knowledge and at monitoring common ground.

Student groups without instruction on question asking (phase 1) produced a similar number of questions and arguments of all types as the instructed student groups. The main difference between the two groups was that instructed students asked significantly more cause-consequence questions and questions aimed at inferring knowledge.

We continued our analyses by measuring correlations between question asking and argumentation. Categories with a high skewness and/or kurtosis were included in Spearman's nonparametric correlation measurement. The results are shown in Table 7.

Compared to Table 4 the number of significant correlations is much lower. In this study, goal-oriented questions, verification questions and questions aimed at monitoring common ground were significantly correlated

Table 6. Frequencies, means per session, standard deviations (s.d.) and p-values considering question types, question generation mechanisms and argumentation across not instructed groups (13 sessions) and groups with instruction (11 sessions)

	Phase 1 (n=13)			Phase 2 (n=11)			
	freq.	mean	s.d.	freq.	mean	s.d.	p-value
(A1) Question types							
1. Goal-oriented question*	10	0.8	1.2	12	1.1	1.0	0.28
2. Cause-consequence question*	7	0.5	1.4	13	1.2	1.0	0.04
3. Evaluative question*	0	0.0	0.0	1	0.1	0.3	
4. Other open questions	18	1.4	1.3	23	2.1	1.8	0.39
5. Verification question	36	2.8	2.3	25	2.3	1.9	0.61
6. Other closed questions*	4	0.3	0.5	7	0.6	1.0	
Total number of questions	75	5.8	4.6	81	7.4	3.5	0.41
(A2) Question generation mechanisms							
1. Inferring knowledge*	32	2.5	3.0	44	4.0	2.1	0.04
2. Correcting knowledge*	14	1.1	1.5	8	0.7	1.0	0.87
3. Monitoring common ground	29	2.2	2.0	28	2.6	2.3	0.82
4. Other . . .*	0	0.0	0.0	1	0.1	0.3	
Total number of questions	75	5.8	4.6	81	7.4	3.5	0.41
(B) Argumentative fragments							
1. Neutral argument*	41	3.2	3.0	58	5.3	6.5	0.46
2. Positive argument	366	28.2	12.0	306	27.8	14.7	0.43
3. Negative argument	343	26.4	13.9	308	28.0	15.7	0.69
Total number of arguments	750	57.7	25.7	672	61.1	30.8	0.36

* skewness > 1 and/or kurtosis > 1; categories are included in a nonparametric test (Mann-Whitney).

with different types of argumentation. Remarkably, no significant relationships were found between the most frequent question generation mechanisms (inferring knowledge and correcting knowledge) and (types of) argumentation. However, monitoring common ground was significantly correlated with argumentation.

Discussion of study 2

We have interpreted the results as follows. Students established a strong motivation to engage in discussion, given the large number of questions and arguments produced in the 10-minute sessions. Students proved to be able to ask every type of question without instruction on question asking, which suggests they suppressed that ability in the first study. Argumentation was not related to the relatively frequent question generation mechanisms of

Table 7. Significant relationships between questions asked, the number and types of arguments across the 24 collaborative learning sessions

Question asking/ Argumentation	Sum of arguments	Negative arguments	Positive arguments
Sum of questions	*	*	–
(A1) Question types			
1. Goal-oriented question	–	–	*
2. Cause-consequence questions	–	–	–
3. Evaluative question	–	–	–
4. Other open questions	–	–	–
5. Verification question	*	*	–
6. Other closed questions	–	–	–
(A2) Generation mechanisms			
1. Inferring knowledge	–	–	–
2. Correcting knowledge	–	–	–
3. Monitoring common ground	*	*	–
4. Other . . .	–	–	–

* significant relation: $p < 0.05$; – no significant relation $p > 0.05$.

correcting and inferring knowledge. This could mean students tried to elicit argumentation but with no effect. Recall the poor result of the knowledge test, which may explain the lack of arguments related to these attempts. It seemed that questions were asked to infer and/or correct knowledge, but arguments were given just to keep the discussion going, without arriving at new insights or conclusions according to the protocols. Students focused on winning the argument that started from the predefined conflicting claims, mainly by producing reasons supporting their own claims. To continue this line of reasoning, the problem may not have been the questions asked, but the arguments given in reply. One important reason for this could be that students had to defend assigned stances, about which their knowledge appeared to be limited, and their discussion did not serve any significant goal apart from the discussion itself.

In order to find more critical argumentation, we thought it would be a good idea to increase commitment to the collaborative learning task by allowing students to defend their self-generated claims. Recall, this was also the case in study 1, with the tutor. In addition, students were asked to submit a joint product to the tutor as a result of their discussion. Furthermore, we decided to use electronic communication instead of face-to-face contact. While CMC

communication is much slower and less questions and arguments might be the result, the slow pace of communication and the availability of a discourse history could also offer the participants more opportunity for reflection on the discussion. In addition, the environment we used affords the construction of graphical representations, which could be used as a permanently available product under construction.

Study 3: electronic collaborative learning sessions

Purpose

We based our design at the results of the first and second study and organised a third study as electronic collaborative learning sessions. To provoke students to engage in deep, critical and thorough forms of argumentative discussion, they had to predefine their own claims for discussion, to work together in an electronic workspace – which included a tool for graphical argumentation – and to aim at reaching a shared product. We varied instruction on competitive versus consensual behaviour in order to assess the effects on students' (graphical) argumentation. The research question for this third study can be formulated as: *What is the relationship between question asking and argumentation in a collaborative learning task with self-defined claims in a CMC environment which affords synchronous chat and joint diagram construction?*

Method

The third study was integrated in an 8-week course on developing CBL programs. The course was similar to the course we described in Study 1, however, this study was conducted a year later with a different cohort of students. Another main difference was the use of Belvedere, a synchronous network-based tool developed by the Learning Research and Development Center at the University of Pittsburgh (Suthers, Toth & Weiner, 1997).

Belvedere can be used for constructing argumentative diagrams online with individuals or groups of students of any size. The working screen of the program displays private and shared windows. To communicate with a partner the student has a text-based chat box in which multi-line messages can be created and sent. Messages will then be displayed, linked with the writer's name, in the shared chat-history. Adding data into the diagram is constrained; students must use the predefined set of boxes (hypothesis, data, unspecified) and links (for, against, and). These are shown in the menu bar in Figure 5.

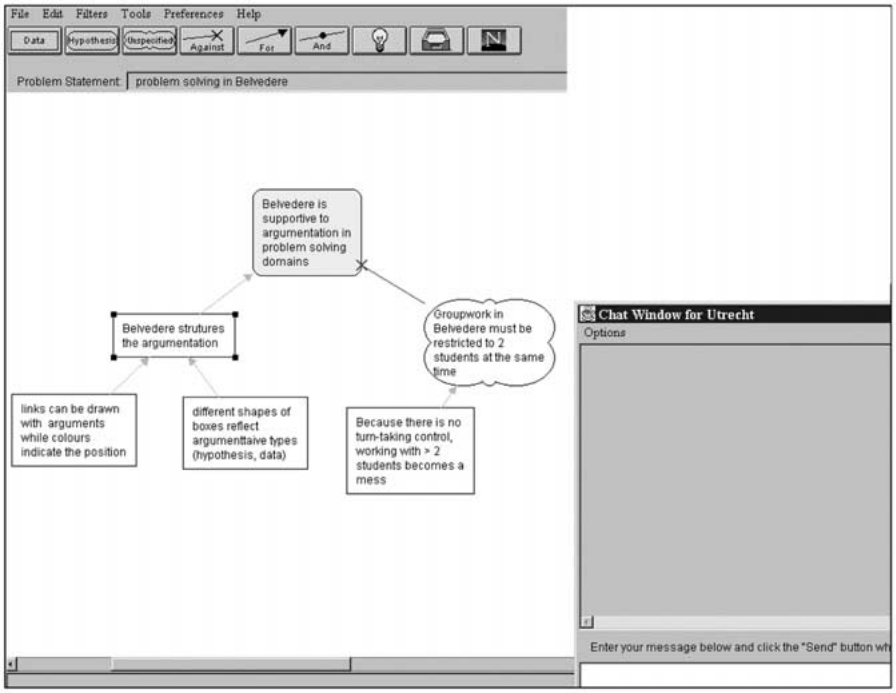

Figure 5. Interface of Belvedere.

In the first week of the course, 14 students received an introduction into principles for instructional design. Then they were asked to group themselves in 7 pairs and to start with the construction of learning goals for an educational software program they had to design within their own group. After defining these goals, they were asked to produce conflicting claims on three aspects that their design should deal with:

- what pedagogical strategies to use
- how to sequence learning activities
- what programming tool to use

In the second week of the course, all student pairs had to discuss their self-defined conflicting claims one by one, by use of the Belvedere system (about an hour of discussion per claim). Besides basic instruction on the technical use of Belvedere, they all received guidelines on how to engage in argumentative discussions (see Figure 6). We figured that perceiving the discussible nature of the task would support students' mental preparation.

The use of self-defined, conflicting claims together with the guidelines to engage in collaborative argumentation were expected to provoke students to have deep argumentative discussions. In addition, we explored the effects of

Guidelines to engage in collaborative argumentation

- Be critical in argumentation
- Use task-related arguments
- Detect feigned arguments (e.g. based on misinterpretations)
- Be co-operative (e.g. do not use pressure, aggression)
- Use multiple perspectives in argumentation
- Argumentse on arguments
- Ask 'open' and verification question

Figure 6. Basic guidelines to engage in collaborative argumentation.

providing student pairs with additional instruction on *competitive* or *consensual* behaviour. In the *competitive* condition student pairs were instructed to behave as competitively as possible, even when already convinced by the other party (considering the conflicting claims). They were told to end the argument only when getting stuck or not being able to think of any more (pro or contra) arguments. Then, a (sub) conclusion could be drawn and a next discussion could be provoked. In the consensual condition student pairs were instructed to behave competitively while discussing the conflicting claims but to concede when being convinced by their peer student. We expected the students in the competitive mode of instruction to argue more and to take more perspectives into account than students in the consensual mode of instruction. Due to politeness strategies, the latter group was expected to agree too soon.

The study was organised as presented in Table 8. Seven student pairs had to discuss three claims each. With respect to the first claim, some of the pairs received instruction to behave competitively; the other pairs were instructed to be aimed at reaching consensus. Considering the second claim, all student pairs switched their roles from being competitive to consensual and vice versa, as was the same case with respect to the third claim. In interpreting the results, we have to bear in mind that in this practical reseach design no adjustments could be made with respect to possible order effects.

Results

We analysed the data in the same way as the formerly described studies on F2F tutoring sessions and collaborative learning sessions. In analysing the chat discussions and diagrams, it appeared that the third pedagogical claim for discussion ("What programming tool to use") was not appropriate since

Table 8. Organisation of study 3

Discussion task	*7 self-organised student pairs*	
	Competitive instruction	Instruction on consensus
– discussion claim 1	Pair 1,2,3	Pair 4,5,6,7
– discussion claim 2	Pair 4,5,6,7	Pair 1,2,3
– discussion claim 3	Pair 1,2,3	Pair 4,5,6,7

students had too little experience in programming. They had trouble to define a clear claim to start with and to engage in further chat discussion. As a consequence, no chats or diagrams were produced that could be used in our analyses.

Chat discussions

Each of the 14 text-based chat discussions that we analyzed could be recognised as an argumentative fragment (7 student pairs * Belvedere sessions on claim 1 + claim 2). Again, like in study 2, a fragment comprised a complete session. Recall, each session took about one hour per claim. However, compared to the face-to-face discussions these electronic chat discussions were about six times less "thick", due to time delays that are common for this mode of communication. Within these fragments, we analysed separated utterances on question types, question generation mechanisms and different types of arguments. In addition, the Belvedere diagrams were analysed on the number of neutrally, positively and negatively oriented arguments in relationship towards the self-defined claims.

The results showed, first of all, that across all Belvedere chat discussions 80 questions were asked and 410 arguments were produced. Students mostly asked verification questions and open answer questions. They aimed at inferring knowledge, correcting knowledge and monitoring common ground. The students in the consensus group produced more positively than negatively oriented arguments (see Table 9). Between conditions, no significant differences were found with respect to question asking and argumentation. The only (weak) effect was that consensually instructed student pairs were more often inclined towards correcting knowledge than competitively instructed student pairs.

We continued our analyses by measuring correlations between question asking and argumentation across all chat discussions. We found an overall and significant relationship between inferring knowledge and the number of negatively and positively oriented arguments triggered ($r = 0.58$; $p < 0.05$

Table 9. Frequencies, means (mn), standard deviations (s.d.) and p-values (t-test) considering question types, generation mechanisms and argumentation within the Belvedere chat conditions of competition versus consensus

	Belvedere chat discussions Competition (n = 7)			Consensus (n = 7)			p-value
(A1) Question types	freq.	mn	s.d.	freq.	mn	s.d.	
1. Goal-oriented question*	4	0.6	0.8	10	1.4	1.4	
2. Cause-consequence question*	2	0.3	0.5	6	0.9	1.9	
3. Evaluative question*	2	0.3	0.8	3	0.4	0.5	
4. Other open questions*	10	1.4	1.5	12	1.7	1.5	0.71
5. Verification question*	13	1.9	2.0	13	1.9	2.3	0.81
6. Other closed questions*	1	0.1	0.4	4	0.6	1.1	
Total number of questions	32	4.6	2.8	48	6.9	4.9	0.30
(A2) Question generation mechanisms							
1. Inferring knowledge*	14	2.0	1.8	13	1.9	1.7	1.00
2. Correcting knowledge*	8	1.1	2.2	13	1.9	0.9	0.07
3. Monitoring common ground*	6	0.9	1.6	15	2.1	3.0	0.39
4. Other . . .*	4	0.6	1.1	7	1.0	1.4	
Total number of questions	32	4.6	2.8	48	6.9	4.9	0.30
(B) Argumentative fragments							
1. Neutral argument*	11	1.6	3.3	30	4.3	2.9	0.54
2. Positive argument	117	16.7	14.1	102	14.6	6.9	0.72
3. Negative argument*	102	14.6	15.4	48	6.7	5.0	0.23
Total number of arguments	230	32.9	28.4	180	25.7	9.5	0.54

* skewness > 1 and/or kurtosis > 1; categories are included in a nonparametric test (Mann-Whitney).

resp. $r = 0.64$; $p < 0.05$). Other relationships found were between open question types and the total of arguments stated ($r = 0.63$; $p < 0.05$) and between verification questions and the number of neutral arguments ($r = 0.60$; $p < 0.05$).

Diagrams

During each chat discussion a student pair produced an argumentative diagram. In sum, fourteen diagrams were gathered and analysed, which related to the chat discussions of claim1 and claim 2. These diagrams contained 181 arguments in sum (see Table 10). Furthermore, competitively instructed student pairs included 30% more arguments in their diagrams than student pairs that were aimed at reaching consensus (107 against 74

Table 10. Argumentation in the Belvedere diagrams

(B) Argumentation	Belvedere diagrams Competition (n=7)			Consensus (n=7)			p-value
	freq.	mean	s.d.	freq.	mean	s.d.	
1. Neutral argument	–	–	–	–	–	–	–
2. Positive argument	57	8.1	2.5	49	7.0	3.3	0.48
3. Negative argument	50	7.1	2.7	25	3.6	1.8	0.01
Total number of arguments	107	15.3	4.5	74	10.6	4.5	0.08

arguments). This was due to the high production of negatively oriented arguments.

Discussion of study 3

The study showed that students engaged in argumentation only when they were sufficiently prepared for the assignment. Their lack of knowledge about programming inhibited them to discuss the third pedagogical aspect. The argumentative task design appeared to be successful, encouraging students to engage in argumentation based upon their self-defined conflicting claims and the requirement to submit a jointly constructed product (the diagram). Questions of all types were found, as well as all different question generation mechanisms. Students provided with competitive instruction produced better-balanced argumentative chat discussions than students aimed at reaching consensus: they elaborated on both positive and negative sides of the argument. Shown as a trend in the chat discussions, these findings were confirmed within the diagrams: the constructed diagrams in the competitive condition showed an even more balanced argumentation than the chat discussions. Finally, it seemed that the assignment affected the nature of question generation mechanisms. Consensus instruction groups showed significantly more correction-aimed questions.

Discussion and conclusions

The studies described in this article explored the question of how argumentation in collaborative learning tasks could be provoked, considering contextual factors such as the role of the tutor, (peer) student, task, instruction and medium. We specifically concentrated on the role of question asking in relation to argumentation. We conducted three studies in sequence. Study 1

was a session in which two or three students discussed face to face with a tutor the learning goals for the CBL program they were designing. In study 2 pairs of students orally discussed two opposing claims about the effectiveness of a learning dialogue, first without instruction, the second time with instruction about question asking. In study 3 groups of students discussed the pedagogical goals and the learning strategies for the CBL program they were designing, by using Belvedere.

Although the studies differed on many aspects, in all studies we collected argumentative fragments that were analysed similarly with the same category system on question asking and argumentation with the same type of users. Therefore, we would like to propose some tenative conclusions. To support discussion, the summary of results is displayed in Table 11. After discussion of the results we present some perspectives for future research and for educational practice.

First, we would like to compare the studies with respect to questions asked, question generation mechanisms, and the obtained relationships with argumentation. Study 1 confirmed the well-known result that students do not ask the tutor any high-level questions. They asked short questions, aiming at monitoring common ground, that is, in order to find out whether there ideas were correct. They argued a lot, however, presumably as a reaction to the tutor's high-level questions, especially those aiming at correcting knowledge. Tutor questioning directed at inferring new knowledge did not have any success in the sense of eliciting arguments.

Study 2 revealed that our students were capable of asking high-level (open) questions, even without additional instruction, when there was no tutor and when they were asked to competitively discuss an assigned claim. However, this study also showed that their open questions did not significantly correlate with argumentation. Argumentation was significantly related to verification questions and to monitoring common ground. While instruction about question asking in phase 2 of this study led to more open questions (cause-consequence) and more questions aiming at inferring knowledge, these questions did not lead to more argumentation.

Study 3, where students could discuss there own claims in an electronic environment, established a significant correlation between open questions and argumentation, as well as between questions aiming at inferring knowledge and argumentation. Instruction to arrive at consensus, compared to competitive instruction, significantly increased the frequency of questions aiming at correcting knowledge, and reduced the number of negative arguments. In our introduction we discussed research supporting the claim that high-level questions and argumentation are important for learning and that the two might be related, in the sense that high-level questions could provoke argumentation.

Table 11. Summary of results

	Study 1		Study 2		Study 3	
	Tutor	Student	Phase 1	Phase 2	Compete	Consent
Number of groups		N = 11	N = 13	N = 11	N = 7	N = 7
(A1) Question types						
1. Goal-oriented questions	5	–	10	12	4	10
2. Cause-consequence questions	2	–	7	13**	2	6
3. Evaluative questions	9	–	–	1	2	3
4. Other open questions	13	–	18	23	10	12
5. Verification question	12	11	36	25	13	13
6. Other closed questions	11	9	4	7	1	4
Total number of open questions	29	0	35	49	28	31
Total number of closed questions	23	20	40	32	14	17
Significant Correlations with number of arguments (any type)	Goal-oriented, cause-conseq., evaluative, other open, verification		Verification		Other open, verification	
(A2) Question generation mechanisms						
1. Inferring knowledge	20	–	32	44**	14	13
2. Correcting knowledge	20	–	14	8	8	13*
3. Monitoring common ground	7	16	29	28	6	15
4. Other . . .	5	4	–	1	4	7
Significant Correlations with number of arguments (any type)	Correcting Monitoring		Monitoring		Inferring	
(B) Argumentative fragments						
1. Neutral argument	20	18	41	58	11	30
2. Positive argument	161	183	366	306	117	102
3. Negative argument	203	182	343	308	102	48
Total number of arguments	384	383	750	672	230	180

**: $p < 0.05$; *: $p < 0.1$, with respect to result in previous column.

The three studies showed different relations between question asking and argumentation. There are relations between the two, and some types of questions seem to trigger more arguments than others, but this heavily depends on the situation. The relationship is very complex, as situations can give rise to argumentation without many high-level questions being asked, while in other cases good questions (open, high level questions) do not always lead to argumentation.

It seems that question asking does affect argumentation, but the way in which this is realised, depends on the task, instruction, the medium, and the roles of the participants (tutor and students) in the learning situation.

The studies reported do not allow specifying in precisely what way, but we can propose some tentative suggestions for task design. If the goal of the educational arrangement is to promote (1) participants asking high-level questions, and (2) to engage in high-level argumentation, then the following recommendations might apply:

- The student: it seems that students are able to ask any type of question and are able to argue, but only in situations that provide them with appropriate responsibility and serve some goal that is useful, such as a common product. This obviously does not mean that they always ask the right questions and give the right arguments at the right moment. This probably requires more instruction and information than we could provide in our studies.
- The peer-student: our students seemed to be able to deal with conflicting claims and competitive instruction. This type of instruction worked well, considering the amount of produced arguments. On the other hand, instruction to agree increased the quality of the question asked somewhat. More insight is needed into the nature of the collaboration process to be able to decide on the choice between competitive an consensus instruction in specific cases. In addition, we propose that electronic tools such as Belvedere could be a great help in supporting many aspects of collaboration.
- The tutor: although the tutor may be able to ask good questions, he often ends up supplying most of the "correct" information during the session (compare Graesser, A.C., C.L. McMahen & B.K. Johnson, 1994). The process of answering a question in a tutoring session is therefore a collaborative process in which participants construct an answer together. Therefore, individuals need to be trained in how to ask the right questions at the right time in a manner that is sensitive to the content of the world knowledge.
- The task: all three tasks we used in this study were open-ended, but the degree of argumentation differed between them. In order to evaluate successful argumentation, in the sense of students learning from each other, the content of the arguments should have been examined. It seems that production and negotiation of good arguments requires better preparation than we have been able to observe with our students.
- Instruction: our efforts to manipulate instruction had some effects, but these effects seemed to be only indirectly related to better argumentation. Instruction in asking questions was beneficial in the sense that more open questions were generated, but probably no better arguments. It can be supposed that in order for open questions to be effective, participants need to feel they are able and willing at providing some

answer to them. The instruction to compete may increase the balance of the discussion, in the sense of equal numbers of positive and negative arguments.

- The medium: how much we would have liked to conclude that Belvedere was successful in supporting the generation of questions aiming at inferring (new) knowledge, of course our study does not allow that conclusion. The slow pace of communication reduced the number of contributions, but not their quality, in terms of the relative amounts of different questions and arguments. The content diagram seemed to summarize the chat discussion, in terms of the relative proportions of positive and negative arguments. These are only suggestions that have to be researched more deeply.

Finally, we like to recommend more research about the role of verification questions in particular and the importance of seeking common ground (e.g., King, 1990; Hakkarainen, 1995; Erkens, 1997). Verification questions correlated significantly with argumentation in all three studies. While it is possible that verification questions serve to monitor common ground, and argumentation has a role in that (Andriessen, Erkens, van de Laak, Peters & Coirier, in press), there have also been suggestions that verification questions serve as important triggers to correct knowledge, by eliciting argumentation (Veerman, 2000).

As a conclusion, to support collaborative argumentation for learning purposes, we first of all have to know more about the specific needs students have in different educational situations (both F2F and CMC) considering collaborative argumentation. It seems that triggering argumentation is not the main issue, it is much more important to create environments in which it is rewarding to have a meaningful argument. Secondly, we have to think about how to adapt the tutor's role or other forms of (electronic) support to the role of the other factors already provided by the task environment (task features, instruction, electronic characteristics of CMC systems, etc.). The need for research results specifying how to provide effective support increases, especially for collaborative argumentation in electronic environments. Since educators discovered the Internet, CMC systems have increasingly been implemented for educational purposes. We think these systems include promising characteristics to support learning through collaborative argumentation. However, as a research community we have to resist the current technology push by the development of a serious pedagogical, empirical research pull that this issue deserves. In this respect, research and development considering such systems is only starting.

Acknowledgment

We would like to greatly thank Dan Jones, Dan Suthers and Jos Jaspers for their extensive support on the Belvédère computer system during the years of our experiments.

References

Andriessen, J.E.B. & Coirier, P. (1999). *Foundations of argumentative text processing*. Amsterdam: Amsterdam University Press.

Andriessen, J.E.B., Erkens. G., Overeem, E. & Jaspers, J. (1996, September). *Using complex information in argumentation for collaborative text production*. Poitiers, France: Paper presented at the UCIS '96 conference.

Andriessen, J.E.B., Erkens, G., van de Laak, C., Peters, N. & Coirier, P. (in press). Argumentation as negotiation in collaborative writing. In J. Andriessen, M. Baker & D. Suthers, eds, *Arguing to learn: confronting cognitions in computer mediated communication*. Dordrecht: Kluwer.

Baarda, D. & De Goede, M.P.M. (1989). *Basisboek methoden en technieken* [Handbook on methods and technics]. Culemborg: Stenfert Kroese/ Educatieve Partners.

Bostock, A.J. (1996). A critical review of Laurillard's classification of educational media. *Instructional Science* 24: 71–88.

Cohen, J. (1968). Weighted kappa: nominal scale agreement with provisions for scales disagreement of partial credit. *Psychological Bulletin* 70: 213–220.

Coirier, P., Andriessen, J.E.B. & Chanquoy, L. (1999). From planning to translating: The specificity of argumentative writing. In J.E.B. Andriessen & P. Coirier, eds, *Foundations of argumentative text processing*, pp. 1–28. Amsterdam: Amsterdam University Press.

Collins, A., Neville, P. & Bielaczyc, K. (2000). The role of different media in designing learning environments. *International Journal of Artificial Intelligence in Education* 11: 144–162.

Erkens, G. (1997). *Coöperatief probleemoplossen met computers in het onderwijs: Het modelleren van coöperatieve dialogen voor de ontwikkeling van intelligente onderwijssystemen* [Co-operative problem solving with computers in education: Modelling of co-operative dialogues for the design of intelligent educational systems]. Utrecht, The Netherlands: Brouwer Uithof.

Graesser, A.C., McMahen, C.L. & Johnson, B.K. (1994). Question asking and answering. In M. Grensbacher, ed., *Handbook of Psycholinguistics*, pp. 517–538. San Diego, CA: Academic Press.

Graesser, A.C. & Person, N.K. (1994). Question asking during tutoring. *American Educational Research Journal* 31: 104–137.

Graesser, A.C., Person, N.K. & Huber, J. (1993). Question asking during tutoring and in the design of educational software. In M. Rabinowitz, ed., *Cognitive foundations of instruction*, pp. 149–172. Hillsdale, NJ: Lawrence Erlbaum.

Hakkarainen, K. (1995, August). *Collaborative inquiry in the computer-supported intentional learning environments (CSILE)*. Nijmegen, The Netherlands: Paper presented at the 6th European conference for Research on Learning and Instruction.

Heuvelmans, A.P.J.M. & Sanders, P.F. (1993). Beoordelaarsoverstemming [Inter-judgement reliability measurement]. In T.J.H.M. Eggen & P.F. Sanders, eds, *Psychometrie in de praktijk*, pp. 443–469. Arnhem, The Netherlands: CITO.

Hightower, R. & Sayeed, L. (1995). The impact of computer-mediated communication systems on biased group discussion. *Computers in Human Behaviour* 11: 33–44.

King, A. (1990). Enhancing peer interaction and learning through guided student-generated questioning. *Educational psychologist* 27: 111–126.

Krosnick, J.A. & Petty, R.E. (1995). *Attitude strength: Antecedents and consequences*. New Jersey: Lawrence Erlbaum.

Laurillard, D. (1993). *Rethinking university teaching: A framework for the effective use of educational technology.* London: Routledge.

Lord, C.G., Ross, L. & Lepper, M.R. (1979). Biased assimilation and attitude polari-zation. The models: Some practical issues. *Psychological Bulletin* 114: 533–541.

Novak, J.D. (1991). Clarify with concept maps: A tool for students and teachers alike. *The Science Teacher* 58(7): 45–49.

Petraglia, J. (1997). *The rhetoric and technology of authenticity in education.* Mahwah, NJ: Lawrence Erlbaum.

Quignard, M. & Baker, M. (1999). Favouring modellable computer-mediated argumentative dialogue in collaborative problem-solving situations. In S.P. Lajoie & M. Vivet, eds, *Artificial Intelligence in Education*, pp. 129–136. Amsterdam: IOS Press.

Stein, N.L. & Miller, C.A. (1991). I win-you lose: The development of argumentative thinking. In J.F. Voss, D.N. Perkins & J.W. Segal, eds, *Informal Reasoning and Education*. Hillsdale, New Jersey: Lawrence Erlbaum.

Stein, N.L., Calicchia, D.J. & Bernas, R.S. (1996). Conflict talk. Understanding and resolving arguments. In T. Givon, ed., *Typological studies in language: Conversational analysis*, pp. 233–267. Amsterdam: John Benjamins.

Suthers, D., Toth, E. & Weiner, A. (1997). An integrated approach to implementing collaborative inquiry in the classroom. *Proceedings of the Conference on Computer Supported Collaborative Learning: CSCL'97*, pp. 272–279.

Van Eemeren, F.H., Grootendorst, R. & Snoeck Henkemans, A.F. (1995). *Argumentatie* [Argumentation]. Groningen, The Netherlands: Woltersgroep.

Veerman, A.L., Andriessen, J.E.B. & Kanselaar, G. (2000). Learning through synchronous electronic discussion. *Computers & Education* 34(2–3): 1–22.

Veerman, A.L. (2000). Computer-supported collaborative learning through argumentation. Print Partners Ipskamp: Enschede (thesis). http://www.library.uu.nl/digiarchief/dip/diss/1908992/inhoud.htm.

Veerman, A.L. & Treasure-Jones, T. (1999). Software for problem solving through collaborative argumentation. In J.E.B. Andriessen & P. Coirier, eds, *Foundations of argumentative text processing*, pp. 203–230. Amsterdam: Amsterdam University Press.

Webb, N.M. (1995). Group collabroation in assessment: multiple objectives, processes and outcomes. *Educational Evaluation and Policy Analysis* 17(2): 239–261.