Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computers in Human Behavior xxx (2004) xxx–xxx
Computers in Human Behavior
www.elsevier.com/locate/comphumbeh

Effects of representational guidance on domain specific reasoning in CSCL

Jannet van Drie *, Carla van Boxtel, Jos Jaspers, Gellof Kanselaar

Department of Educational Sciences, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

Abstract

Computer-supported collaborative learning (CSCL) aims at enhancing and supporting peer interaction and the joint construction of products through technology. This study investigated the effects of the joint construction of external representations on the collaborative process and the learning outcomes. By providing representational guidance, the study aimed at promoting co-elaborated and domain-specific reasoning. Since it is assumed that the representational format may be of influence on the collaborative process and outcomes, three representational formats, namely an argumentative diagram, an argument list and a matrix, were compared with a control group. Sixty-five student pairs from pre-university education collaborated on a historical writing task in a CSCL environment. The analyses included analyses of interaction processes in the chat, the quality of the co-constructed representation, the quality of the essay and the scores on the individual posttest. The results indicated that each representational format has its own affordances and constraints. For example, Matrix users talked more about historical changes, whereas Diagram users were more focused on the balance in their argumentation. However, this did not result in differences in the quality of historical reasoning in the essay, nor in outcomes on the posttest.

Keywords: Computer-supported collaborative learning; External representations; Historical reasoning; Interaction analyses

* Corresponding author. Tel.: +31 30 253 4926; fax: +31 30 253 2352.
E-mail address: j.vandrie@fss.uu.nl (J. van Drie).

0747-5632/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2004.10.024

1. Introduction

Current trends in the field of learning and instruction stress the importance of active knowledge construction and collaborative learning. Technology can play a major role in implementing these new trends in education (Kanselaar, De Jong, Andriessen, & Goodyear, 2000). Technology can support the construction of knowledge by representing learners' ideas and understandings and it can function as a social medium to support learning by dialogue. Computer-supported collaborative learning (CSCL) aims at enhancing and supporting peer interaction and the joint construction of products by the use of technology (Lipponen, 2002). The key factor that determines the success of CSCL can be found in the quality of the interaction processes students engage in. After all, meaningful learning in a collaborative environment is related to the quality of the interaction processes (Van der Linden, Erkens, Schmidt, & Renshaw, 2000). Although research in the field of CSCL has resulted in positive learning outcomes (Lethinen, Hakkarainen, Lipponen, Rahikainen, & Muukonen, 2001), the use of CSCL is no guarantee for a productive dialogue (Kirschner, 2002; Stahl, 2002; Veldhuis-Diermanse, 2002). More research is needed to reveal under which conditions CSCL can lead to the intended knowledge construction. Important questions in this respect are: what *kind* of interaction processes promote collaborative knowledge construction and *how* can such interaction be provoked and supported?

Studies on collaborative learning processes are conducted from different perspectives. A distinction can be made between a domain-specific, an elaboration, and a co-construction perspective (Van Boxtel, 2004). The *domain-specific perspective* focuses on the propositional content and quality of the discourse. From this perspective an important question is whether the students make progress from their everyday reasoning towards a deeper understanding and more scientific ways of reasoning on the topic at hand. Types of talk that are of interest from a domain-specific perspective are the explication of one's own conceptions, the comparison of these conceptions with new information and interpretations of others, and the search for meaningful relations.

The *elaboration perspective* focuses on the types and quality of the cognitive processes during group work. Elaborative activities – such as the verbalization of prior knowledge, questioning, and the creation of meaningful relations by giving examples, using analogies, reformulating or referring to previous experiences – are considered important ingredients of a productive student interaction. From this perspective, it is important in a collaborative learning situation to promote elaborative talk. Elaborative talk is often constituted by the asking and answering of questions and through the elaboration of controversy by providing justification and argumentation.

The *co-construction perspective* puts the contingencies of the actions of both partners and the mediational role of tools in the centre. From this perspective, an important question is whether knowledge is really shared and co-constructed. In many groups participants do not equally contribute. Sometimes one of the participants does almost all of the talking and work, while the others passively watch

and wait. Kumpulainen and Mutanen (1999) distinguished different modes of social processing. The *individualistic* mode implies that students work individually in the group and do not share ideas or try to co-construct meanings, the *dominative* mode reflects unequal participation, and the *collaborative* mode reflects joint meaning making. Co-construction of knowledge implies that meanings are extended, deepened or transformed because participants build on each other's contributions (Van Boxtel & Van Drie, 2003). Joint meaning-making and co-construction of knowledge requires a shared focus and coordination on the task content level, the meta-cognitive level and the socio-communicative level (Erkens et al., 2003). In many studies one of the perspectives prevails, whereas a multi-perspective approach may have advantages in order to make progress in the design of collaborative learning environments. In the study reported here, a multi-perspective approach is adopted to investigate the role of representational tools for supporting historical reasoning in a computer-based collaborative inquiry and writing environment.

2. The potential of external representations

As has been stated before, using a CSCL environment does not automatically result in knowledge construction. This was confirmed in a previous study (Van Drie, Van Boxtel, & Van der Linden, in press), where a CSCL environment that enables students to collaboratively engage in a historical inquiry task and the collaborative writing of an essay was used. Students used a shared text processor, a private notepad, and had access to information sources. All communication between the collaborating students took place in an integrated chat facility. A writing task was used for several reasons. Previous research has shown that a writing task can deepen students' knowledge and understanding (Klein, 1999; Tynjälä, Mason, & Lonka, 2001) and may result in deeper historical understanding (Boscolo & Mason, 2001; Voss & Wiley, 1997). Moreover, collaborative writing can trigger critical reflection, externalization of thinking and immediate feedback (Gere & Stevens, 1989). Especially writing an argumentative text may result in a productive discussion, for learners may have different views or use different confronting arguments (Giroud, 1999; Veerman, 2000). Finally, small group inquiry task in which students jointly write an essay are more often used in current Dutch history education. The results of this study indicated that although students learned from the task and were engaged in historical reasoning in their chat discussions, the reasoning episodes were often very short and of poor quality. Furthermore, the collaboratively written essays did not show the quality that was expected.

Based on these results, different ways to promote and raise the level of historical reasoning both in chat and essay were considered. A possible way to support and improve collaborative learning in a computer-supported learning environment is by using representational tools. In the following sections, the potential of the collaborative construction of external representations in a historical writing task in CSCL is considered. The focus is on the way the construction of an external representation

may support historical reasoning, elaboration, and co-construction in student interaction.

2.1. Supporting historical reasoning

From a domain-specific perspective it is important to know whether tools in a CSCL environment can promote thinking and reasoning within the domain at hand. In this study the focus is on the domain of history. Historical reasoning can be considered as a key aspect of building historical knowledge. Historical reasoning is always constructed in relation to a historical question or hypothesis. It implies that the learner situates historical phenomena in time, uses historical concepts, and organizes information to describe processes of change and continuity, to explain a historical phenomenon or to compare historical phenomena. Moreover, it implies supporting claims with arguments, making use of historical sources, and taking into consideration the trustworthiness, representativeness and usefulness of the sources (Van Boxtel & Van Drie, 2004). Historical events, processes, and structures need to be organized to build an interpretative historical case (Leinhardt, Stainton, Virji, & Odoroff, 1994).

This study investigated how the construction of an external representation influences the process of historical reasoning. The format used to display information is an important dimension of external representations (De Jong et al., 1998). Suthers and Hundhausen (2003) argue that the cognitive and social affordances of a representation depend on the representational notation. Different representational formats may support particular components of historical reasoning. For example, the construction of a causal diagram may provide guidance when learners are asked to explain a historical phenomenon, whereas a matrix can be a useful format to organize aspects of change and continuity. Furthermore, the representational artifact constructed in a representational tool can function as a writing aid. Experimental studies of Suthers et al. showed that representational notations can have significant effects on learners' discourse during the collaborative construction of external representations in the area of science. They compared the construction of three types of external representations: text document, matrix and diagram. In their study, students worked together behind one computer. The matrix group represented significantly more evidential relations; the empty cells in a matrix seemed to have prompted users to fill in all available evidential relations.

2.2. Supporting elaboration

From an elaboration perspective it is important that students are stimulated to engage in elaborate activities. Zhang and Norman (1994) state that external representations guide, constrain or determine cognitive behavior. Much research on the use of external representation focuses on the (individual) use of *presented* external representations. However, in a CSCL environment students are supposed to actively engage in the construction of their own knowledge, which implies that students

themselves should (co-)construct representations. Cox (1999) claims that the self-construction of external representations may help to translate information from one type of representation to another, thus supporting deeper understanding of the underlying concepts and situations. Moreover, the *collaborative* construction of an external representation can promote verbalization of own conceptions, the (re-)ordering of information, and can provide perceptual assistance. A graphical representation, for example, can make information explicit and can direct attention to central problems and relations and help to distinguish core issues from more peripheral ones (Suthers & Hundhausen, 2001).

2.3. Supporting co-construction

An important condition for the co-construction of knowledge is that group members participate and contribute more or less equally and that they coordinate their activities. In face-to-face collaboration coordination is partly constituted by gesturing and using facial expressions (Schegloff, 1991), whereas the lack of these impose certain constraints on the coordination processes in electronic communication. In an electronic discourse via a chat facility it is important to coordinate and maintain focus on the main issues (Veerman, 2000). From a co-construction perspective the question is whether external representations can contribute to the construction and maintenance of a shared understanding and a joint problem space between co-learners (Crook, 1998; Veerman & Treasure-Jones, 1999). According to Suthers and Hundhausen (2003), an external representation can increase the conceptual complexity that can be handled in group interactions and facilitate elaboration on previously represented information. From this perspective, the representation can facilitate the co-construction of meanings through building on each other's contributions.

While reasoning within a domain, elaboration and co-construction are intertwined. Collaboration can stimulate the articulation of task-related knowledge and information. This verbalization makes it possible for ideas to be questioned, criticized and elaborated, and thus generates explanations, justifications and a search for new relations, which are important aspects of elaboration, historical reasoning and co-construction of knowledge.

3. Aim of the study

This study aims at supporting historical reasoning in a writing task in CSCL by providing representational guidance. The collaborative construction of external representations may support collaborative knowledge construction in a CSCL environment through the facilitation of (domain-specific) cognitive and communicative processes. Moreover, external representations may be helpful for organizing available information in the preparation of the co-authoring of an essay. The main question of this study is whether and how the co-construction of an external

representation influences the collaborative process of knowledge construction and have an effect on the learning outcomes with respect to reasoning and learning in the domain of history. To examine the influence of the representational format, three different representational formats will be compared. The focus is especially on the appearance and quality of domain-specific reasoning, elaboration, and co-construction, for these processes are believed to constitute positive learning outcomes of collaborative learning.

4. Method

4.1. Design

The study consists of a pretest–posttest design with four conditions. In the experimental groups, students were asked to co-construct an argumentative diagram with arguments pro and contra (Diagram condition), a list of arguments pro and contra (List condition) or a matrix in which changes can be described and characterized (Matrix condition). In the control condition, no representational tool was available and students did not receive instruction to co-construct an external representation. The experiment was conducted in two phases: in the first year the experiment was conducted for the Diagram and List condition, in the next year for the Matrix and Control condition.

4.2. Participants

Participants in this study were 157 students from six history classes in secondary (pre-university) education, aged 16–17. Three classes from two schools participated over a period of two years (same schools, teachers and level). The experiment took place at school, during the history lessons and lasted for six lessons in two weeks time. The students worked in pairs, each behind a computer and the pairs were divided over two computer labs. Pairs in which one of the students missed more than one lesson were excluded. The analyses included 130 students (65 pairs). Within their class, the students were randomly assigned to pairs and to one of the two conditions. In the first year of the experiment, 16 student pairs participated in the Diagram condition and 14 in the List condition. In the second year 18 student pairs participated in the Matrix condition and 17 in the Control condition. So, in each year the conditions were randomized over classes and schools.

4.3. Task and learning environment

Students performed a historical inquiry task that involved studying historical sources (such as texts from textbooks, different interpretations of historians, photos, tables, and interviews) and writing an essay of approximately 1000 words. The task was about the question of whether the changes in the behavior of the Dutch youth in

the nineteen sixties were revolutionary or not. Students worked for six lessons (of 50 min) on the task and did not receive instruction on the subject in advance. Students in the experimental conditions were instructed to collaboratively construct an external representation, for which they could use the historical sources. After finishing the construction of the representation, they could start co-writing the text, for which they could use the constructed representation and the sources. Students in the control group performed the same task, without the instruction to construct an external representation.

Students worked in a CSCL environment called Virtual Collaborative Research Institute (see http://edugate.fss.uu.nl/vcri; Jaspers & Erkens, 2002). VCRI is a groupware program that enables students to work collaboratively on an inquiry task and essay writing. Each student works at one computer, physically separated from the partner. Communication takes place by means of chat. Fig. 1 shows the main screen of VCRI in the diagram condition. Information about the task and relevant historical sources can be found in the database menu. The upper left window contains a chat facility and the chat history. The lower left window contains a shared text processor that can be used by taking turns. The upper right window contains a private notepad. In the lower right window, the representational tool is shown (in this figure the diagram). The representational tools in the experimental conditions are all

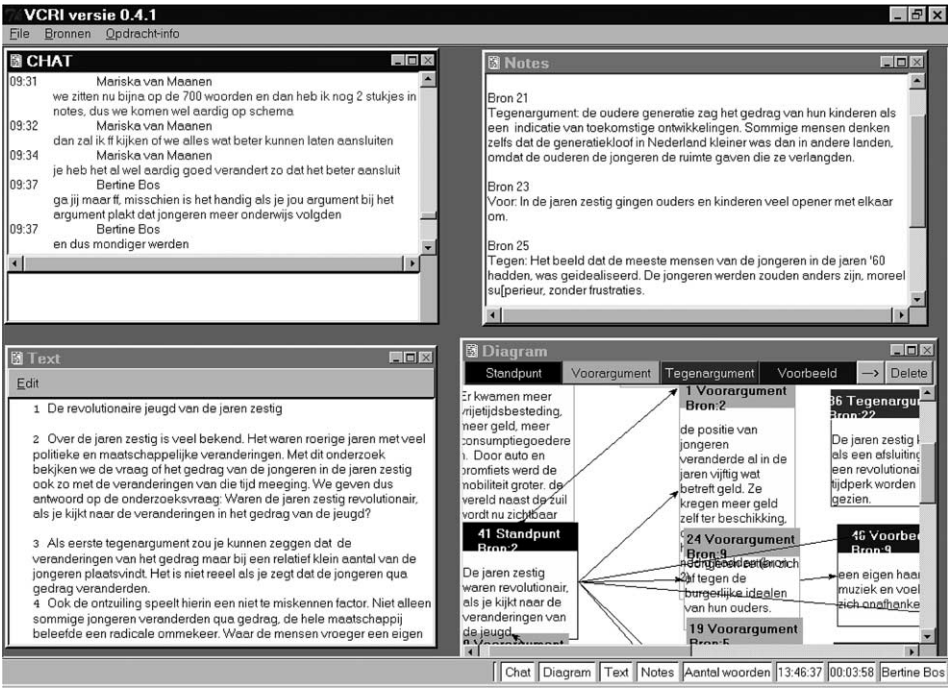


Fig. 1. The main screen of VCRI for the diagram condition.

shared tools. In the control group, no such shared tool was available. Below the representational tools that were used in the experimental conditions will be described.

4.3.1. Argumentative diagram

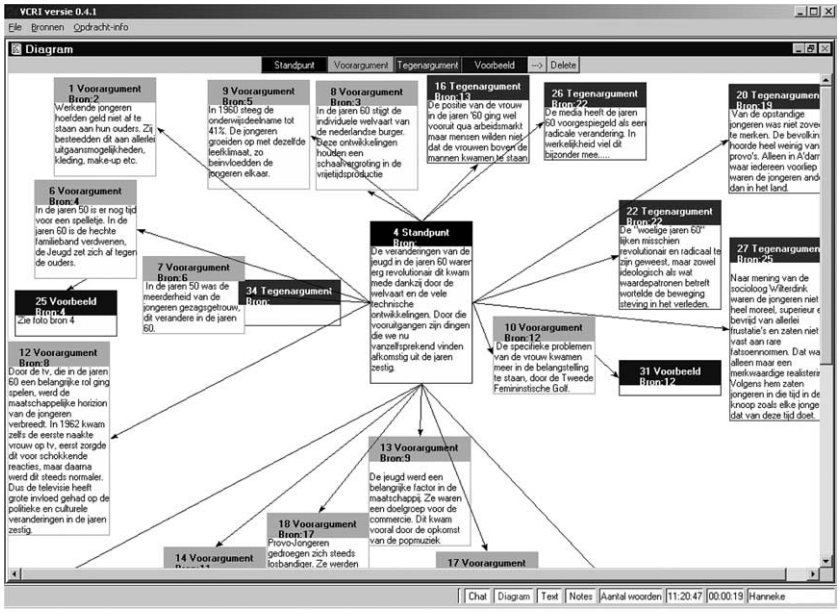
The task requires students to take a point of view on a historical issue and support it with arguments. In an argumentative diagram, a point of view and arguments pro and contra can be graphically represented. Fig. 2 shows the diagram tool that was used in the study. Standpoint, arguments pro, arguments contra and examples can be represented in text-boxes, each with their own color. All text-boxes can be linked to each other by arrows. Furthermore, students can refer to the source from which the argument or the example derives in each box.

4.3.2. List

Whereas an argumentative diagram organizes and links arguments in a two-dimensional graphical way, the list organizes arguments in a linear way. In the List condition a list tool as is shown in Fig. 3 was used. In this window, students can put together arguments pro and arguments contra.

4.3.3. Matrix

The argumentative diagram and list focus on the process of argumentation, and do not pay explicit attention to more domain-specific aspects of the argumentation process required by the task. The task is about historical change. Historical change



Legend: Standpunt = standpoint; Voorargument = argument pro;
Tegenargument = argument contra; Voorbeeld = example

Fig. 2. Example of a diagram constructed by one of the dyads (in Dutch).

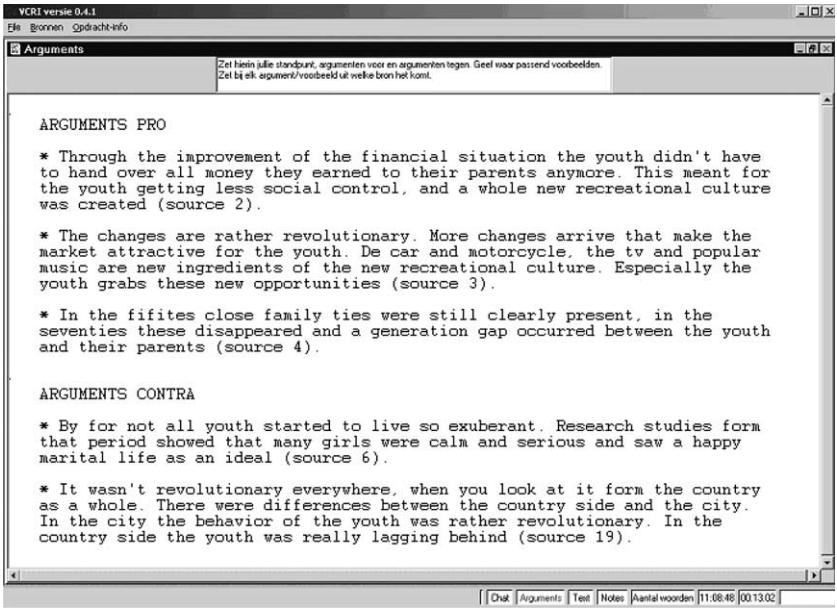


Fig. 3. Example of a list constructed by one of the dyads (translated from Dutch).

can occur in very different areas, for instance beliefs, economic situation or family life. In describing and evaluating processes of change, historians often distinguish between political, economic, social, and cultural changes. Furthermore, they characterize changes according to their tempo and impact. A matrix is a format in which historical changes can be characterized using several columns. Fig. 4 shows the matrix tool used in the study. The matrix consists of a table format that can be filled in by the students. In the second column the number of the source can be added, in the third column students can describe the historical changes (or aspects of continuity) and in the fourth whether they think the change can be defined as revolutionary or not (yes or no). The last column contains a ''sort'' function. Students can categorize the changes in the way they choose. When they push on the sort button, all the changes are sorted. For example, the students who made the matrix in Fig. 4 categorized the changes as cultural, political and economic. After sorting, all changes with the same label are listed.

4.4. Instruments and analyses

The analyses focused on the process of collaboration as well as on the products as outcomes of the collaboration. All actions taken during working on the task were logged. The analyses included the interaction processes in the chat protocols, the jointly constructed representations, the collaboratively written essays, and the results of the individual pretest and posttest. The analyses of the processes focused on

id	Source	Description Changes	Revolutionary change?	Sort
5	00	Old ways of living, traditions and cultures were partly lost, secularization and decline of the youth movement	YES	Culture
6	00	Interference by the government decreased during the 1950s but did not disappear	NO	Politics
7	00	strong demand for fundamental changes during the 1960s	YES	Politics
9	02	Youngsters earned more money, so they did not need to give up their complete earnings to their parents	YES	Economy
8	03	New types of leisure activities due to increasing wealth	No	Economy
10	04	Parents had more time to spend time with their families	NO	Culture
11	05	Because the number of people pursuing an education increased rapidly, social differences between people diminished.	YES	Education
13	07	Nozems were not accepted by society. They were considered hopeless cases.	NO	Politics
12	08	The influence of television grew strongly during the 1960s, a structural depillarization took place, although depillarization was still visible in broadcasting	YES	Culture
17	09	The rise of new music was widely noticeable. Young people in particular were attracted by it.	Yes	Culture
14	10	Youth culture changed significantly between the 1950s and the 1960s; as became visible through the rise of popular idols, and a tendency to revolt against grownups.	YES	Culture

Fig. 4. Example of a matrix constructed by one of the dyads (translated from Dutch).

historical reasoning, elaboration, and co-construction. Below the different analyses will be described in more detail.

4.4.1. Chat protocols

The interaction processes in the chat protocols were coded by using MEPA, a computer program for Multiple Episode Protocol Analysis (Erkens, 2002; see http://edugate.fss.uu.nl/mepa). The interaction processes in the chat were coded on the level of utterances and were first analyzed on the dimension of Task Acts. Five main categories were distinguished: utterances related to the content of the task at hand (Task), to procedures to perform the task (Procedures), talk about the technical functioning of the computer-program (Program), social talk (Social), and greetings at the start or ending of a working period (Greetings). The categories Task and Procedures were divided in subcategories. Task utterances in which students say something about the past, give an interpretation of the past, or of the merit of sources, were coded as Historical reasoning.

Utterances coded as Historical reasoning were further analyzed on a episodic level. An episode is defined as several subsequent historical reasoning utterances which belong to the same type of historical reasoning. These analyses are related to the three perspectives described above: domain-specific, elaboration and co-construction, and can be found in Table 1. With regard to the domain-specific perspective, six types of historical reasoning episodes were distinguished, namely episodes in which (1) historical phenomena are situated in time, (2) the past is

Table 1
Coding definitions of historical reasoning episodes

Category	Description
Type	Content of the historical reasoning episode is related to:
Time	Historical time
Description	Describing the past
Change	Describing changes
Explanation	Giving explanations
Source	Interpretation and evaluation of sources
Standpoint	Providing a point of view and arguments
Elaboration	Extended historical reasoning episode, which starts with a:
question	Question
conflict	Negation, counter-argument or critical question
reasoning	Statement
No elaboration	Historical reasoning episode that does not contain extended reasoning
Degree of co-construction	Degree to which the episode is co-constructed:
Co-constructed	Both students contribute equally to the reasoning
Dominated	One student dominates the reasoning
Individual	only one student contributes to the reasoning
Co-elaborated historical reasoning	*Extended historical reasoning episode to which both students equally contribute*

described, (3) changes (or continuity) are described, (4) the past is explained, (5) the sources are discussed, and (6) a point of view is taken and supported with arguments. The historical reasoning episodes were next analyzed on the appearance and type of elaboration. Three types of elaboration were distinguished: question, conflict and reasoning (Van Boxtel, 2000). Episodes that did not contain an elaboration, were coded as No Elaboration. Furthermore, the degree of co-construction in the historical reasoning episodes was analyzed. Inspired by the modes of social processing of Kumpulainen and Mutanen (1999) the following distinction was made: individual reasoning, reasoning dominated by one of the students, and co-construction.

Historical reasoning episodes were coded as *co-elaborated historical reasoning* whenever the reasoning reflected both elaboration and co-construction. Table 2 shows an example of co-elaborated historical reasoning (Van Drie et al., in press). In this example Paula and Wendy discuss their point of view on the question of whether the sixties were revolutionary or not. Paula and Wendy co-construct their meaning on this subject. First, they talk about which point of view they are taking, and whether they both agree on this. They ask questions which elicit elaboration, such as "What is our opinion?", "Why?", and "Which arguments pro are we going to use?". They both, in turn, add arguments to support their meaning, and they elaborate upon the reasoning of the partner, as is shown in lines 16 and 17.

Table 2
Example of co-elaborated historical reasoning (fragment of chat protocol, translated from Dutch)

1	Paula	*But what is our opinion?*
2	Wendy	*To start with, what is your opinion?*
3	Paula	*The sixties were revolutionary.*
4	Wendy	*Why?*
5	Paula	*Because the consequences are still noticeable now.*
6	Wendy	*Okay, that is true.*
7	Paula	*But it already started in '50.*
8	Wendy	*That's true too, but that's got little to do with the revolutionariousness, so the sixties were indeed revolutionary!?*
9	Paula	*If you think so too.*
10	Wendy	*Yes, sure I do too.*
11	Paula	*Okay.*
12	Wendy	*Which arguments pro are we going to use?*
13	Paula	*That young people became a group, and that they had their own opinion.*
14	Wendy	*The depillarization*[a] *went on strongly.*
15	Paula	*Yes, and the consumptive society arose.*
16	Wendy	*People started to think more flexible about sex, which meant greater freedom for young people.*
17	Paula	*Yes, young people had more freedom anyway, because of their being financially independent they were able to leave their parents more early and they were independent of their parents.*
18	Wendy	*Exactly.*
19	Paula	*Okay.*

[a] Depillarization refers to the breakdown of the strict division in socio-religious groups or pillars (Protestants, Catholics, Socialists, and Liberalists) that existed in Dutch society since the beginning of the twentieth century.

For all categories mentioned in this section (with the exception of asymmetry for that was calculated on the basis of information that was logged by the computer) the inter-rater reliability was measured by two coders (the first and second author) over four randomly chosen chat protocols from a pilot study. The agreement varied between 83% and 98%, and Cohen's κ varied between 0.69 and 0.95.

4.4.2. External representations

The external representations the students produced in the experimental conditions were scored on the number of arguments pro and contra. The inter-rater reliability (Cohen's κ) over 12 randomly chosen representations was 0.89 for the arguments pro and 0.78 for the arguments contra. The number of sources referred to in the representation, the total number of arguments used (arguments pro plus arguments contra), and the balance of arguments pro and contra were also included in the analyses. The balance refers to the difference between the number of arguments pro and the arguments contra. A higher score on this measure means less balance. It should be noted that based upon the sources that were provided, more arguments pro could be identified compared to arguments contra. In sum, sixteen arguments pro could be identified from the sources and ten arguments contra.

4.4.3. Essays

Essays were scored on six aspects of historical reasoning: time references, changes and continuity, explanations, use of sources, argumentation, and the use of historical concepts. The scoring took into account both amount and quality. For example, the number of explanations given, and the quality of the explanations given. Furthermore, a score was given for the structure of the complete essay. The maximum score on the essay was 60 points. The inter-rater reliability between two coders on ten essays, turned out to be 0.59 (Cohen's κ; agreement 72%). This outcome is considered sufficient, since assessing texts is highly interpretative. However, it was decided to score all essays independently by two coders, compare the results, and discuss differences until agreement was reached.

4.4.4. Pretest and posttest

The pretest and posttest focused on subject knowledge about the sixties, since the aim of the task was to improve subject-matter knowledge. The test contained seven open answer questions and one multiple-choice question. The items were constructed in line with the different aspects of historical reasoning (see Table 3).

The pretest and the posttest consisted of the same questions, only for some items different historical sources (for example a different picture or text) were used. The maximum score on both tests was 79. The inter-rater reliability of the scoring, on ten randomly chosen tests, varied between 0.70 and 1.00 (Cohen's κ). After excluding item 1a, in which the students had to give associations on the fifties, the item homogeneity (Cronbach's α) turned out to be acceptable (pretest 0.72; posttest 0.64).

Table 3
Description of the items of the pre- and posttest and maximum scores

Item	Description	Components of historical reasoning	Max score
1a	Associations fifties[a]	–	7
1b	Associations sixties [a]		7
2	Situating historical phenomena in time[b]	Time	10
3	Giving a definition of four concepts	Concepts	8
4	Giving examples of four concepts	Concepts	8
5a	Giving characteristics of the youth in the fifties	Change	4
5b	Describing changes in the behavior of the youth in the sixties	Change	4
6a	Giving causes for the changes in the behavior of the youth in the sixties	Explanation	10
6b	Indicating the most important cause	Explanation	2
7a	Giving arguments pro the given statement	Standpoint	6
7b	Giving arguments contra the given statement	Standpoint	6
8a	Interpretation of a source	Source	2
8b	Interpretation of a source	Source	2
8c	Evaluating the trustworthiness of both sources	Source	3

[a] Association item: item in which students are asked to give associations on the fifties and the sixties in a mind map.

[b] Multiple-choice items. The answers were correct or false.

4.4.5. Questionnaire

After finishing the assignment the students were asked to fill out a questionnaire that contained evaluative questions about the task and the computer environment.

4.5. Hypotheses

In this study the influence of the co-construction of an external representation in a writing task in CSCL is investigated. Three different representational format (a diagram, list and matrix) are compared to a control group, which included the same task without the collaborative construction of an external representation. It is expected that compared to the control group, the experimental conditions (Diagram, List and Matrix) would show more co-elaborated historical reasoning in the chat, would produce better essays and would have higher scores on the posttest. In order to verify the assumption that the amount of co-elaborated historical reasoning, the quality of the constructed representation, the quality of the essay and the scores on the posttest are positively related, the correlations between these variables will be calculated.

Based upon the characteristics of the different representational formats used, differences between the three experimental conditions in historical reasoning, the constructed representations and the written essays are expected. Historical change and argumentation are important elements of the task used in this study. It is hypothesized that the chat protocols in the Matrix condition, compared to the Diagram and the List, would show more talk about historical changes, for the columns in the matrix direct attention to describing historical changes and deciding whether each of these changes can be considered revolutionary or not. In addition, for the Matrix condition higher scores on the aspect of historical change in the essay and on the posttest are expected. The List and Diagram both focus on the process of argumentation and it is therefore expected that students in these conditions would show more discussion about their standpoint in the chat, and have higher scores for argumentation in the essay and on the posttest. Moreover, it is expected that the graphical format of the diagram would have more potential to promote balance in the representation of arguments pro and contra than a linear format. The balance between arguments pro and contra is more salient in a diagram, since one can easily see whether it is even or uneven.

5. Results

In this section, the results of the analyses of the interaction processes, the collaboratively constructed products and the individual learning outcomes are presented, as well as the outcomes of correlational analyses. To test differences between the conditions, both univariate and multivariate analyses of variance were carried out. First, the conditions were compared on students' subject knowledge about the nineteen sixties, measured by the pretest. A one-way ANOVA revealed that the scores were

significantly different for the four conditions on the individual level ($F(3, 126) = 16.17$; $p \leqslant 0.00$). Also at the level of pairs (for which the average score of both students was taken) a one-way ANOVA yielded a significant difference between the conditions ($F(3, 61) = 14.03$; $p \leqslant 0.00$). Post hoc tests (Dunnett's C) indicated that the student-pairs in both the Diagram ($M = 27.8$; SD = 2.7) and List condition ($M = 27.9$; SD = 4.6) scored significantly higher on the pretest ($p \leqslant 0.05$), than the student-pairs in the Matrix ($M = 21.8$; SD = 5.2) and Control condition ($M = 18.7$; SD = 6.1). Therefore, the average pretest-score of the dyads was used as a covariate in the univariate and multivariate analyses. To indicate differences between the conditions simple contrast analyses were carried out. Differences are considered significant when $p \leqslant 0.05$. When different analyses were conducted, this will be explicitly mentioned.

5.1. Co-elaborated historical reasoning in the chat protocols

The mean length of protocols was 361.4 utterances (SD = 171.5). ANCOVA revealed that the length of the protocols (see Table 4) was significantly different

Table 4
Mean frequencies and standard deviations of task acts in chat protocols ($N = 65$)

Task acts	Diagram ($N = 16$)		List ($N = 14$)		Matrix ($N = 18$)		Control ($N = 17$)	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Task*	77.3	(30.6)	102.1	(52.5)	128.2	(60.7)	153.2	(72.6)
Historical reasoning*	16.4^{-}	(14.3)	17.6^{-}	(11.0)	34.3^{+}	(23.4)	35.2^{+}	(35.3)
Representation	19.8^{+}	(14.2)	8.6^{-}	(6.7)	13.9	(10.5)	–[a]	
Text construction*	11.5^{-o}	(6.9)	22.0^{-}	(13.1)	22.6^{-x}	(14.4)	39.2^{+}	(22.3)
Revision*	4.3^{-o}	(3.8)	12.1^{-}	(15.6)	9.1^{-x}	(12.1)	17.7^{+}	(16.9)
Goal	8.2	(5.7)	13.4	(14.2)	17.3	(13.0)	13.9	(10.9)
Resources	3.0	(3.8)	7.1	(9.1)	7.6	(8.8)	6.5	(7.2)
Evaluation task*	8.5^{-}	(4.8)	10.1^{-}	(8.3)	14.8^{-}	(11.1)	22.8^{+}	(13.0)
Word count*	5.4^{-}	(3.6)	10.1^{-}	(9.1)	11.4^{-}	(8.5)	18.4^{+}	(13.9)
Procedures*	121.4	(60.6)	132.4	(80.0)	187.3	(65.7)	192.2	(81.7)
Coordination	75.6	(44.9)	90.5	(53.0)	91.6	(34.9)	106.0	(58.4)
Task approach*	18.6^{-}	(10.6)	20.7^{-}	(8.6)	51.4^{+}	(30.4)	53.1^{+}	(25.2)
Planning	7.3	(4.6)	7.6	(7.3)	13.3	(9.0)	15.3	(14.7)
Turn taking*	17.7^{-}	(15.5)	16.1^{-}	(13.8)	28.5^{+}	(15.8)	15.0^{-}	(12.6)
Evaluation	2.3	(2.7)	2.6	(3.9)	2.5	(3.6)	2.9	(3.8)
Program	17.6	(15.3)	28.0	(28.6)	13.3	(13.1)	12.4	(13.8)
Social	35.9	(43.0)	29.3	(25.2)	35.1	(56.4)	41.1	(43.3)
Greeting	19.5	(8.5)	23.1	(8.8)	19.1	(7.8)	27.5	(8.0)
No Code	0.4	(1.8)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total	271.9	(108.5)	320.7	(170.2)	384.0	(137.9)	426.5	(180.1)

Note. Means are unadjusted means.
Post hoc analysis: $^{+}$ sign. > $^{-}$; ^{x} sign. > ^{o}.
[a] In the control group no representational tool was used.
* $p \leqslant 0.05$.

for the conditions ($F(3,60) = 4.47$; $p \leqslant 0.01$). Simple contrast analyses indicated that the Control condition produced more chat-utterances than Diagram and Control, and that the Matrix produced more utterances than the Diagram. Thus, although all students had worked for 6 h on the task, the students in the Control and Matrix condition produced more utterances in the chat.

All utterances in the chat protocols were first coded on the level of Task Acts and are presented in Table 4. In general, the analyses of the Task Acts showed that most of the Task Acts were directly related to the assignment: 46% of the utterances were related to Procedures and 33% to Task. About 21% of the utterances was about the technical functioning of the program (Program), social talk (Social) and Greetings at the beginning or the end of a session. Furthermore, it worth noting that a relatively high percentage of utterances was coded Coordination (26%).

A MANCOVA, with the average pretest scores of the pairs as covariate, was used to test whether there were differences between the conditions for the subcategories of the variable Task and the variable Procedures. First, a MANCOVA for the variable Task was conducted. The subcategory Representation was left out, since the Control condition had no representation and therefore no Task Acts related to Representation. An overall effect was found ($F(21,158) = 2.68$; $p \leqslant 0.00$), with significant differences for the subcategories: Historical Reasoning ($p \leqslant 0.01$), Text Construction ($p \leqslant 0.00$), Revision ($p \leqslant 0.00$), Task Evaluation ($p \leqslant 0.01$), and Word Count ($p \leqslant 0.05$). It was expected that the Experimental conditions would show more historical reasoning compared to the Control condition, however, the results indicated differently. Both the Control and Matrix condition scored higher compared to Diagram and List on Historical Reasoning. In addition, the Control condition scored significantly higher compared to the other tree conditions on Text Construction, Revision, Task Evaluation, and Word Count. The Matrix condition scored higher than the Diagram on Text Construction and Revision. In other words, the Control condition scored higher on those variables that were directly related to the process of text-writing. An ANCOVA was carried out on the subcategory Representation, which yielded a significant difference between the three experimental conditions ($F(2,44) = 3.97$; $p \leqslant 0.05$). Significantly more utterances were related to Representation in the Diagram, compared to the List condition.

Next, a MANCOVA was conducted on the subcategories of the variable Procedures. An overall effect was found ($F(15,164) = 2.30$, $p \leqslant 0.01$), with significant differences for the subcategories: Approach ($p \leqslant 0.01$), and Turn Taking ($p \leqslant 0.05$). Both in the Control and Matrix condition the students talked significantly more about the approach of the task, compared to the students in the List and Diagram condition. In the Matrix condition significantly more utterances were related to Turn Taking compared to the other three conditions.

As shown above, both the Control and Matrix condition scored higher compared to Diagram and List on the amount of historical reasoning. Historical reasoning episodes were also analyzed on the type of historical reasoning, the amount of elaboration and the degree of co-construction. The results of these analyses are shown in Tables 5 and 6. Notice, that instead of the number of episodes, the number of utter-

Table 5
Mean frequencies and standard deviations of historical reasoning in chat protocols and the results of analyses of variance ($N = 65$)

Historical reasoning	Diagram ($N = 16$)	List ($N = 14$)	Matrix ($N = 18$)	Control ($N = 17$)	F	p
Time	0.4^{-} (1.1)	1.3^{-} (2.8)	1.2^{-} (3.0)	5.1^{+} (6.5)	3.77	0.02^{*}
Description	3.4^{-} (4.9)	2.3^{-o} (3.2)	5.2^{x} (5.9)	6.8^{+} (11.1)	2.87	0.04^{*}
Change	3.4^{-} (5.0)	2.9^{-} (4.2)	14.1^{+} (15.4)	6.1 (11.8)	4.57	0.01^{**}
Explanation	0.4 (1.3)	0.9 (2.7)	1.2 (3.5)	0.8 (2.9)	0.37	0.78
Source	0.7 (1.8)	0.1 (0.5)	1.4 (5.2)	0.9 (3.0)	0.55	0.65
Standpoint	8.4 (6.6)	10.1 (7.1)	11.2 (11.5)	15.5 (17.1)	2.61	0.06

Note. Means are unadjusted means. Post hoc analysis: $^{+}$ sign. > $^{-}$; ^{x} sign. > ^{o}.
* $p \leqslant 0.05$.
** $p \leqslant 0.01$.

Table 6
Mean frequencies and standard deviations of elaboration, co-construction and co-elaboration in historical reasoning episodes in chat protocols and the results of analyses of variance ($N = 65$)

	Diagram ($N = 16$)	List ($N = 14$)	Matrix ($N = 18$)	Control ($N = 17$)	F	p
Elaboration	12.2^{-} (12.7)	11.8^{-} (10.4)	20.9^{+} (18.2)	27.1^{+} (32.4)	5.31	0.00^{**}
– Question	6.1^{-} (9.1)	7.2^{-} (8.4)	10.5^{+} (9.0)	13.2^{+} (12.2)	4.15	0.01^{**}
– Conflict	0.9 (3.0)	1.4 (4.0)	2.0 (6.1)	0.9 (2.7)	0.80	0.50
– Reasoning	5.1^{-} (5.3)	3.1^{-} (4.3)	8.4 (11.7)	13.0^{+} (21.7)	3.31	0.03^{*}
No elaboration	4.6 (4.8)	5.9 (6.1)	13.4 (16.8)	8.1 (5.5)	2.04	0.12
Degree of co-construction						
– Co-construction	7.6^{-} (10.5)	10.3^{-} (9.8)	22.8^{+} (20.0)	28.4^{+} (31.1)	5.36	0.00^{**}
– Domination	3.4 (5.1)	2.2 (4.1)	4.2 (7.1)	3.3 (4.2)	0.83	0.48
– Individual	5.8 (4.9)	5.0 (5.9)	7.3 (7.4)	3.5 (4.1)	1.39	0.26
Co-elaboration	*6.7^{-o} (10.0)*	*9.6^{-} (10.3)*	*13.4^{-x} (13.7)*	*$23.9\ (30.4)^{+}$*	*5.47*	*0.00^{**}*

Note. Means are unadjusted means. Post hoc analysis: $^{+}$ sign. > $^{-}$; ^{x} sign. > ^{o}.
* $p \leqslant 0.05$.
** $p \leqslant 0.01$.

ances that are part of the different types of historical reasoning episodes are mentioned in this table. In this way, the length of the episodes can be taken into account.

With respect to the different components of historical reasoning, the results indicated that most historical reasoning was about historical changes in the sixties and the point of view students take (see Table 5). Both of these aspects were central to the task at hand. It was expected that the Matrix condition would show more talk about historical changes, and that the Diagram condition would show more utterances reflecting students' point of view regarding the historical issue at hand. A MANCOVA on the components of historical reasoning revealed an overall effect ($F(18, 161) = 2.06$; $p \leqslant 0.01$). Significant differences between the conditions were

found for the categories Time ($p \leqslant 0.05$), Description ($p \leqslant 0.05$), and Change ($p \leqslant 0.01$). Simple contrast analyses revealed that students in the Matrix condition talked significantly more about historical changes compared to students in the Diagram and List condition, which is in line with the expectations. Furthermore, students in the Control condition made more time-references compared to the students in the other three conditions. Both the Control and Matrix condition scored significantly higher on Description compared to the List condition, and the Control condition also scored higher compared to the Diagram for this category.

The mean frequencies of utterances that were part of an *elaborated* historical reasoning episode are given in Table 6. Most elaboration was related to the asking and answering of questions. An ANCOVA revealed a significant difference between the conditions for the total amount of elaboration ($p \leqslant 0.01$). Both the Matrix and Control condition scored significantly higher on Elaboration compared to the Diagram and List condition. With regard to the type of elaboration the analyses showed a significant difference for Question ($p \leqslant 0.01$) and Reasoning ($p \leqslant 0.05$). Both the Control and Matrix condition scored higher on Question compared to the Diagram and List condition. The Control condition scored also higher on Reasoning compared to the Diagram and List.

Table 6 also presents the results for the degree of co-construction in historical reasoning. An ANCOVA revealed a significant effect Co-construction ($p \leqslant 0.01$). Simple contrast analysis indicated that both the Control and Matrix condition scored higher on Co-construction compared to the Diagram and List condition. In addition, it was expected that the conditions with a representational tool would show more co-elaborated historical reasoning than the Control condition. However, the results indicated the contrary. Analysis of variance showed a significant difference between the conditions for the amount of co-elaborated historical reasoning ($p \leqslant 0.01$; see Table 6). The Control condition showed significantly more co-elaborated historical reasoning compared to the Diagram, List and the Matrix. Moreover, the Matrix showed more co-elaborated historical reasoning compared to the Diagram.

5.2. Constructed representations

In the experimental conditions, the students collaboratively constructed a representation. The question now arises as to whether the three representational formats resulted in differences in the number of represented arguments, the balance between arguments pro and contra, the number of sources used and the equality of participation in the construction of the representation. The results of an ANCOVA, presented in Table 7, confirm an effect of the conditions on the total number of arguments ($p \leqslant 0.01$), the number of arguments pro ($p \leqslant 0.01$) and arguments contra ($p \leqslant 0.01$). Simple contrast analyses revealed that in both the List and in the Matrix condition more arguments were used compared to the Diagram condition. The Matrix scored significantly higher on Arguments Pro compared to both the List and Diagram, whereas the List scored higher than the Diagram. With regard to Arguments Contra both List and Matrix scored higher compared to Diagram.

Table 7
Mean scores and standard deviations of the representation for the experimental conditions and results of analysis of variance ($N = 48$)

	Diagram ($N = 16$)	List ($N = 14$)	Matrix ($N = 18$)	F	p
Total arguments	13.5^{-} (3.7)	17.6^{+} (3.1)	18.6^{+} (2.0)	13.60	0.00^{**}
– Arguments pro	8.9^{-o} (2.4)	10.9^{-x} (1.6)	12.5^{+} (1.2)	17.10	0.00^{**}
– Arguments contra	4.6^{-} (1.9)	6.7^{+} (2.5)	6.1^{+} (1.7)	4.64	0.01^{**}
Balance pro and contra	4.3^{-} (2.3)	4.2^{-} (2.8)	6.4^{+} (2.1)	4.78	0.01^{**}
Sources used	14.3^{-o} (4.8)	20.9^{-x} (4.5)	25.1^{+} (0.7)	36.89	0.00^{**}

Note. Means are unadjusted means. Post hoc analysis: $^{+}$ sign. > $^{-}$; ^{x} sign. > ^{o}.
$^{*}p \leqslant 0.05$.
$^{**}p \leqslant 0.01$.

Additionally, the balance between the arguments pro and contra was analyzed (see Table 7). The balance was computed as the difference between the number of arguments pro and the number of arguments contra. A larger score means less balance. An ANCOVA revealed a significant difference between the conditions. Simple contrast analyses indicated that the Matrix showed less balance compared to the Diagram and List. Table 7 also shows the number of sources used in the representations. In sum, 26 sources were available. In the Matrix condition almost all sources were used and in the Diagram condition students used the least number of sources. An ANCOVA revealed a significant difference between the conditions (see Table 7) and simple contrast analyses showed that this difference was significant between all three conditions.

How did the students experience working with the representational tool? In the questionnaire, the students were asked to give their opinion about working with the representational tool. About 80% of the students who worked with one of the representational tools ($N = 96$) thought it a useful way of working. They thought that the tool helped them to select important information for the essay and to structure this information. About 15% (5% was missing) was less positive and considered the construction of the representation as extra work, or preferred their own way of working instead of the structure offered by the representation they used. The students who worked in the Control group were asked whether they preferred the way they worked, or the way the other group of students in their class worked (they used the Matrix). One third of the 30 students who filled out the questionnaire did actually prefer to work with the matrix tool, for they thought it would be a useful way to select and organize the information.

5.3. Learning outcomes

5.3.1. Essays

As to whether the construction of different representations resulted in differences in learning outcomes, the results of the collaboratively written essays were first exam-

Table 8
Mean scores, standard deviations and maximum scores for the essay ($N = 65$)

Quality of the essay	Diagram ($N = 16$)	List ($N = 14$)	Matrix ($N = 18$)	Control ($N = 17$)	Maximum score
Time	4.8 (0.8)	5.4 (0.7)	5.0 (0.8)	5.4 (0.9)	6
Concept*	6.9^{-} (1.2)	6.4^{-o} (1.1)	6.9^{x} (1.3)	7.3^{+} (1.6)	12
Change	6.1 (1.1)	6.5 (0.9)	6.8 (1.2)	6.6 (1.1)	9
Explanation	4.4 (0.8)	4.1 (1.6)	4.6 (1.3)	4.4 (0.9)	9
Standpoint	8.5 (2.5)	10.2 (2.0)	8.8 (2.2)	9.2 (2.2)	15
Source*	3.9^{-} (1.1)	4.1^{-} (1.0)	3.9^{-} (1.1)	4.4^{+} (1.2)	6
Structure	1.9 (0.8)	1.9 (0.7)	1.9 (0.7)	1.9 (0.6)	3
Total	*36.7 (5.1)*	*38.6 (4.9)*	*38.1 (5.3)*	*39.2 (4.6)*	*60*

Note. Means are unadjusted means. Post hoc analysis: $^{+}$ sign. > $^{-}$; ^{x} sign. > ^{o}.
* $p \leqslant 0.05$.

ined. In Table 8 the scores for the categories that were used to describe the quality of the essays are presented.

A MANCOVA, with the mean pair scores on the pretest as covariate, revealed a significant difference between the conditions for the quality of the essays ($F(21, 158) = 1.70$; $p \leqslant 0.05$), for the categories Concept ($p \leqslant 0.01$) and Source ($p \leqslant 0.05$). Simple contrast analyses showed that the Control condition scored higher on Concept compared to both Diagram and List. The Matrix scored also higher compared to List. With regard to the category Source scored the Control significantly higher compared to all three other conditions. The results did not indicate differences on the categories Change and Standpoint as was expected.

It was expected that the conditions in which the students had constructed an external representation would show higher scores for the essays than the control group. A possible explanation for the fact that this expectation was not met, might be that the students who did not have to construct a representation could spend more time on writing the essay. That the students in the Control group talked more in the chat about aspects related to the writing of text (such as Text construction and Revision), could point into this direction. To check this assumption, the time spent on the different tools in the CSCL environment was distilled from the log files (see Table 9).

Table 9
Mean time (in min) and standard deviations spent in the Chat, Notes, and Essay and the results of a one-way ANOVA ($N = 65$)

	Diagram ($N = 16$)	List ($N = 14$)	Matrix ($N = 18$)	Control ($N = 17$)	F	p
Chat	76.3 (34.8)	71.9 (22.9)	80.5 (18.4)	85.2 (25.8)	0.50	0.68
Notes	43.4^{-} (25.6)	61.0^{-} (35.6)	55.6^{-} (31.3)	93.2^{+} (36.4)	6.01	0.00**
Essay	49.0^{-} (26.1)	47.1^{-} (27.5)	55.6 (24.4)	68.8^{+} (20.9)	2.79	0.05*

Post hoc analysis: $^{+}$ sign. > $^{-}$; ^{x} sign. > ^{o}.
* $p \leqslant 0.05$.
** $p \leqslant 0.01$.

If the aforementioned explanation is true, the control group would have spent significantly more time at the writing of the essay. A one-way ANOVA revealed that the conditions differed significantly on this category (see Table 9) and that the Control condition indeed spent significantly more time on writing the essay compared to the Diagram and List (Dunnett's C post hoc test). Moreover, the Control condition spent also more time on working in the Notes box compared to the other three conditions. In general, the Notes box was used by the individual student (it is not a shared tool) to summarize important information from the sources. However, the Notes box was also used to write (parts of) paragraphs for the essay, when the partner was writing in the text editor (in which they could not work at the same time). The fact that the students in the Control condition spent more time making notes, could actually mean that they spent more time on the writing of the essay, not only in the shared text editor, but also in their personal Note box. However, additional analyses on the content of the notes should prove this.

The representational tool was added to support students in selecting and organizing information from the sources in order to write the essay. This assumption presupposes that the items mentioned in the representation will subsequently be used in the essay. To test this assumption, the number of items in the representation and in the essay was counted, as was the amount of overlap between them. The results of a one-way ANOVA, presented in Table 10, and the post-hoc tests (Dunnett's C) showed that both in the List and in the Matrix more items were presented compared to the Diagram. The students in the List condition also presented more items in their essays, compared to the students using the Diagram. The number of overlap items turned out to be significantly higher for both the List and Matrix condition in comparison to the Diagram condition. The percentage of items in the representation that were also mentioned in the essay was also calculated. The List showed the highest percentage and a significant difference with Diagram. In addition, the percentage

Table 10
Mean frequencies and standard deviations of number of items in representation and essay, number of overlap, percentages overlap on total representation item, and on total essay items and the results of a one-way ANOVA ($N = 48$)

	Diagram ($N = 16$)	List ($N = 14$)	Matrix ($N = 18$)	F	p
Representation	13.1^{-} (3.7)	16.9^{+} (3.0)	18.1^{+} (3.6)	13.03	0.00*
Essay	10.8^{-} (2.5)	13.9^{+} (3.2)	11.2 (3.3)	4.77	0.01**
Overlap	7.6^{-} (2.5)	11.3^{+} (2.4)	9.9^{+} (2.5)	8.89	0.00**
Overlap: % of total items in representation	53.8^{-} (16.4)	68.3^{+} (14.9)	55.2 (14.5)	4.07	0.02*
Overlap: % of total items in essay	69.9^{-} (14.9)	82.8^{+} (14.4)	84.7^{+} (10.0)	6.14	0.00**

Post hoc analysis: $^{+}$ sign. > $^{-}$; ^{x} sign. > ^{o}.
* $p \leqslant 0.05$.
** $p \leqslant 0.01$.

of overlap on the total of items in the essay was calculated. It turned out that the Matrix and List scored significantly higher compared to the Diagram. In other words, the students in the Matrix and List added less new items in their essays (that were not already presented in the representation), and that students in the Diagram added more new items.

5.3.2. Pretest and posttest

In Table 11 the results of the individual pretest and posttest on the main items are presented, as well as the maximum scores of the items. A paired samples T-test showed that the students improved on all items of the test ($p \leqslant 0.05$). The only exception was that the students in the List condition did not improve on their scores on the Source item (interpretation and evaluation of the trustworthiness of two historical sources).

As was mentioned before, the total score on the pretest turned out to be different for the conditions ($F(3) = 16.17$; $p = 0.00$). The post hoc test revealed that the score in the Diagram and List conditions was higher than the score in the Matrix and Control condition. Therefore the *individual* score on the pretest was used as a covariate in the analyses of the posttest. A MANCOVA with the pretest score as a covariate showed a significant effect of the conditions on the posttest scores ($F(21, 353) = 1.76$; $p \leqslant 0.05$). Significant differences at a 0.05 level were found for the variables: Concepts (defining and giving examples of historical concepts), Standpoint (providing argument for and against the statement that the sixties were revolutionary) and Source (interpretation of sources and evaluation of their trustworthiness). Simple contrast analyses revealed that both List and Control scored higher than the Matrix on Concepts and that the Control group scored higher than the Diagram and List on Standpoint and Source. Since the students in the Matrix condition reasoned more about processes of change and continuity in the chat, the Matrix condition was expected to score higher on the items about Change. However, this expectation was not met. Nor was the expectation met that the Diagram and List condition would score higher on Standpoint.

Table 11
Mean scores and standard deviations of pretest and posttest for the conditions ($N = 130$)

	Diagram ($N = 32$)		List ($N = 28$)		Matrix ($N = 36$)		Control ($N = 34$)	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Associations	3.1 (1.4)	5.9 (1.0)	3.1 (1.6)	5.3 (1.4)	2.6 (1.9)	5.4 (1.1)	2.0 (1.3)	5.5 (1.2)
Time	6.4 (1.6)	7.1 (1.5)	5.9 (1.8)	7.5 (1.3)	5.5 (2.0)	7.3 (1.5)	4.9 (2.5)	7.5 (1.8)
Concept	8.1 (2.7)	11.2 (2.4)	8.1 (2.7)	12.1 (2.0)	5.9 (2.5)	9.6 (2.6)	5.4 (2.9)	11.2 (4.4)
Change	2.5 (1.3)	4.3 (1.4)	2.6 (1.5)	4.6 (1.5)	1.9 (1.4)	4.2 (1.3)	1.9 (1.3)	4.5 (1.3)
Explanation	2.6 (1.9)	5.1 (2.7)	2.5 (1.8)	5.1 (2.3)	1.4 (1.7)	5.0 (2.5)	0.8 (1.1)	5.9(2.4)
Standpoint	1.7 (1.2)	3.3 (1.9)	2.1 (1.2)	3.4 (1.4)	1.7 (1.4)	3.5 (1.4)	1.1 (1.2)	3.7 (1.4)
Source	3.4 (1.7)	4.2 (1.7)	3.6 (1.7)	4.1 (1.6)	3.0 (1.8)	4.4 (1.7)	2.7 (1.8)	4.6 (1.9)
Total	*27.8 (4.3)*	*41.1 (7.2)*	*27.9 (5.5)*	*42.1 (5.9)*	*21.8 (6.5)*	*39.4 (7.6)*	*18.7 (8.4)*	*43.0 (7.6)*

Note. Means are unadjusted means.

5.4. Correlational analyses

An important expectation of this research was that the collaborative construction of an external representation will result in more co-elaborated historical reasoning in the chat protocols and in higher scores for the essay and the posttest. As was shown above, analyses of variance did not confirm this assumption. On the contrary, the Control condition, which performed the task without a representational tool, showed more co-elaborated historical reasoning. However, this leaves the question unanswered whether more co-elaborated historical reasoning goes together with higher scores for the essay and higher scores on the posttest, and whether the students who constructed a representation of higher quality also wrote a better essay and performed better on the posttest. Therefore correlations were computed for the variables: co-elaborated historical reasoning, scores of the constructed representation, scores of the essay, and the scores on the posttest. Notice that the scores on the test are individual scores, whereas the scores on the other three variables are group scores. To calculate the correlations between the test scores and the other variables, the pair score was ascribed to the individual students. This results in an enlargement of the N. The results of the analyses showed no consistent outcomes for the four conditions. Two significant correlations were found. In the Diagram condition there was a significant correlation between the number of arguments in the representation and the posttest scores ($r = 0.49$, $p = 0.005$, $N = 32$). In the Matrix condition the number of utterances that belonged to co-elaborated episodes correlated significantly with the score of the essay ($r = 0.66$, $p = 0.003$, $N = 18$). More co-elaboration went together with higher scores for the essay. In the other conditions this result was not found.

5.5. Prior experience with CSCL

Some students participating in the Control and Matrix condition had prior experience with CSCL (they had participated in another research-project with a similar kind of software tool), whereas none of the students in the Diagram and List condition had this kind of experience. To determine whether this experience may have influenced the outcomes of this study, the students of the Matrix and Control condition were divided in two new groups, the first group had experience with CSCL ($N = 46$; 23 pairs) and the other did not ($N = 24$; 12 pairs). All pairs consisted of students both with experience or both without experience. As mentioned before, the Matrix and Control condition did not differ in the scores on the pretest, and therefore t-tests for independent samples were used to test significant differences between the two groups ($p \leqslant 0.05$). It turned out that no significant differences were found for the following variables: (a) co-elaborated historical reasoning in the chat ($M_{\text{experience}} = 20.0$, $\text{SD}_{\text{experience}} = 23.3$, $M_{\text{no experience}} = 15.8$, $\text{SD}_{\text{no experience}} = 25.0$), $t(33) = 0.48$, $p = 0.63$, (b) the quality of the essay ($M_{\text{experience}} = 39.4$, $\text{SD}_{\text{experience}} = 4.0$, $M_{\text{no experience}} = 37.3$, $\text{SD}_{\text{no experience}} = 6.3$), $t(33) = 1.20$, $p = 0.24$, and (c) the scores on the individual posttest ($M_{\text{experience}} = 41.1$, $\text{SD}_{\text{experience}} = 7.9$, $M_{\text{no experience}} = 41.2$, $\text{SD}_{\text{no experi-}}$

$_{ence}$ = 7.7), $t(68) = 0.04$, $p = 0.96$. In other words, students with experience in CSCL did not score differently compared to students without previous CSCL experience. So, experience with CSCL cannot be used to explain why the Control and Matrix condition scored higher on the main variables.

6. Conclusions and discussion

In this article, the results of a study on the effects of the construction of external representations on the collaborative construction of historical knowledge in a CSCL environment are reported. The analyses focused on the collaborative process (historical reasoning, elaboration and co-construction in the chat), the constructed products (representation and essay) and individual learning outcomes (pre- and posttest). The results of this study indicated that a collaborative writing task in a CSCL environment is useful for promoting historical reasoning and the learning of history. All students learned from the task, as the results of the pretest and the posttest indicated. Moreover, the chat discussions and the constructed representations and essays reflected historical reasoning, although the amount of historical reasoning in the chat was less than expected. The main function of the chat in these kind of complex tasks seems to be the coordination of activities (e.g., Erkens, Jaspers, & Prangsma, 2001). Because typewritten utterances involve a lot of effort, students might confine themselves to what is minimally necessary for the coordination of the task instead of engaging in extended content-specific discussions.

It was hypothesized that the addition of a representational tool in the CSCL environment would result in more co-elaborated historical reasoning in the chat discussion. This expectation was not confirmed. An explanation might be that both the representational tool and the chat are shared tools with (among others) the function of sharing information. By adding an argument in the representational tool, the argument is communicated to the other student and becomes part of the shared context. Therefore, it is not necessary to also share this argument via chat. The suggestion made by Suthers and Hundhausen (2003), that when one wishes to modify a shared representation one feels the obligation to discuss this first with the partner(s), was not confirmed here. The difference in communication between the two studies, face to face in the study of Suthers et al., and chat in this study, might explain this different finding. It might be too much effort to communicate each addition or modification in the chat, whereas this might be easier in face to face communication. Thus, co-elaborated historical reasoning does not only take place in the chat discussion, but also *through* the use of the representational tools. This suggests that the representational tool does not only function as a cognitive tool that can elicit elaborative activities, but also as a tool *through which* students communicate and elaborate.

Furthermore, it was expected that the students who constructed an external representation would score higher on the essay. This expectation was also not confirmed by the results; the Control condition scored as well as, and sometimes even better than the experimental groups. A possible explanation for this outcome might be that

the students in the Control condition could spend more time on the writing of the essay, since they did not have to construct a representation. The analyses of the Task acts in the chat protocols, that indicated that the students in the Control condition were more focused on the writing of the essay, point in this direction. Additional analyses on the time spent in the different tools confirmed this explanation. The students in the Control condition did not only spend significantly more time in the Text-editor, but also in the individual Notes-box. The Notes-box was probably used to write parts of the essay when the other student was working in the text editor, which were then cut and passed in the essay. Further analyses on the content of the notes should reveal whether the Notes-box was indeed used for writing parts of the essay or for other purposes, and for which purposes the Notes-box was used in the other conditions.

Another possible explanation for the fact that the experimental groups did not score higher on the essays compared to the control group, might be that students had not had enough experience with constructing an external representation, let alone enough to use this representation for writing the essay. The large overlap between items of the representation and the essay in the List condition, indicated that the list was easiest to use for text writing. The diagram seemed to be most difficult to use; there was only a small amount of overlap between items of the representation and the essay, and a lot of new items were used in the essays that were not represented in the representation. Moreover, the standard deviations for almost all the variables were high, which indicates that the variation between dyads was large. Probably, a lot of other aspects are of influence on students' behavior and learning results, such as motivation, text writing skills, and experience with CSCL. This last was ruled out by the analyses presented on prior experience.

The comparison between the three different representational formats shows that the representational notation effects the type of interaction in the chat. This finding is in line with the findings of Suthers and Hundhausen (2003). The results of the analyses show some important advantages and disadvantages of each form of representation. It was expected that the matrix would have more potential to support domain-specific reasoning, in this study especially reasoning about historical changes. This hypothesis was confirmed. The students in the Matrix condition talked most about historical changes, a component of historical reasoning that was most important for the task that we used. Moreover, in this condition the amount of co-elaborated historical reasoning correlated positively with the total score of the essay. In the matrix most arguments were represented and references were made to almost all available sources. This finding is in line with results of the study of Suthers and Hundhausen (2003) in which the matrix group represented the most evidential relations. In this study, the matrix seems to have prompted students to fill in all available changes and continuities for which they used almost all available sources. However, this did not result in higher scores on the aspect Change in the essay and the posttest. It might be possible that the scoring of the essay was not detailed enough to catch the differences. Additional analyses of the essays showed that the students in the Matrix condition more often categorized the historical changes as political, cultural and economic, whereas students in the others

conditions did less. However, this difference was not caught by our scoring of the essay.

The results also indicated that a diagram is less suited for representing a lot of information. The constructed diagrams contained significantly less arguments and references to sources than the matrices and lists. A main advantage of an argumentative diagram lies in the fact that it is possible to organize the arguments graphically and to interrelate the arguments with links. However, a diagram might become too complex and too hard to organize when a lot of information has to be represented. It seemed that students in the Diagram condition had difficulties in selecting the most important information they needed for the writing of the essay. First, there was not much overlap in items in the representation and the essay and second they added a lot of new items in their essays. The results of this study also showed an advantage of constructing a diagram. In line with the expectations, in the diagrams students reached more balance between arguments pro and contra. Unfortunately, this result was not reflected in the essays or posttest.

To conclude, this study shows that a collaborative writing task in a CSCL environment is a useful task to promote historical reasoning and the learning of history. Moreover, the representational format seems to influence aspects of the collaborative learning process in a CSCL environment. The representational format seems especially of influence on aspects of domain-specific reasoning. However, this did not result in differences in learning outcomes. The representational tools did not elicit more domain-specific discussions in the chat, but were mainly used as tools through which information was communicated and shared. Continued work in this area needs to give us more insight into the support that representational tools can give, especially with respect to the domain-specific reasoning that is asked for in the task. Furthermore, more research about the role of experience with a CSCL environment and of experience with constructing and using external representations is needed.

References

Boscolo, P., & Mason, L. (2001). Writing to learn, writing to transfer. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Writing as a learning tool. Integrating theory and practice* (pp. 83–104). Dordrecht: Kluwer Academic Publishers.

Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction, 9*, 343–363.

Crook, C. (1998). Children as computer users: the case of collaborative learning. *Computers in Education, 30*(3/4), 237–247.

De Jong, T., Ainsworth, S., Dobson, M., Van der Hulst, A., Levonen, J., Reimann, P., et al. (1998). Acquiring knowledge in science and mathematics: the use of multiple representations in technology-based learning environments. In M. W. Van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.), *Learning with multiple representations* (pp. 9–40). Oxford, UK: Pergamon.

Erkens, G. (2002). *MEPA. Multiple episode protocol analysis (version 4.8) [computer software]*. Utrecht, The Netherlands: Utrecht University.

Erkens, G., Jaspers, J., & Prangsma, M. (2001). *Coordinated activity in collaborative learning*. In Paper presented at the Ninth European Conference on Learning and Instruction, Fribourg, Switzerland.

Erkens, G., Kanselaar, G., Prangsma, M. E., & Jaspers, J. G. M. (2003). Computer support for collaborative and argumentative writing. In E. De Lorte, L. Verschaffel, N. Entwistle, & J. van

Merriënboer (Eds.), *Powerful learning environments: unravelling basic components and dimensions* (pp. 157–176). Amsterdam: Pergamon, Elsevier Science.

Gere, A. R., & Stevens, R. S. (1989). The language of writing groups: how oral response shapes revision. In S. W. Freedman (Ed.), *The acquisition of written language. Response and revision* (pp. 85–105). Norwood, NJ: Ablex.

Giroud, A. (1999). Studying argumentative text processing through collaborative writing. In J. E. B. Andriessen & P. Coirier (Eds.), *Foundations of argumentative text processing* (pp. 149–179). Amsterdam: Amsterdam University Press.

Jaspers, J., & Erkens, G. (2002). *VCRI. Virtual Collaborative Research Institute (version 1.0) [computer software]*. Utrecht, The Netherlands: Utrecht University.

Kanselaar, G., De Jong, T., Andriessen, J., & Goodyear, P. (2000). New Technologies. In P. J. R. Simons, J. L. van der Linden, & T. Duffy (Eds.), *New Learning* (pp. 55–82). Dordrecht: Kluwer Academic Publishers.

Kirschner, P. A. (2002). Can we support CSCL? Educational, social and technological affordances for learning. In P. A. Kirschner, W. Jochems, P. Dillenbourg, & G. Kanselaar (Eds.), *Three worlds of CSCL: Can we support CSCL?* (pp. 7–47). Heerlen: Open Universiteit Nederland.

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11*(3), 203–270.

Kumpulainen, K., & Mutanen, M. (1999). The situated dynamics of peer group interaction: an introduction to an analytic framework. *Learning and Instruction, 9*(5), 449–473.

Leinhardt, G., Stainton, C., Virji, S. M., & Odoroff, E. (1994). Learning to reason in history: mindlessness to mindfulness. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 131–158). NJ: Lawrence Erlbaum Associates.

Lethinen, E., Hakkarainen, K., Lipponen, L., Rahikainen, M., & Muukonen, H. (2001). *Computer supported collaborative learning: A review*. Retrieved September 4, 2001 from http://www.kas.utu.fi/clnet/clnetreport.html.

Lipponen, L. (2002). Exploring foundations for computer-supported collaborative learning. In G. Stahl (Ed.), *Computer support for collaborative learning: foundations for a CSCL community. Proceedings of CSCL 2002, Boulder, Colorado* (pp. 72–871). .

Schegloff, E. (1991). Conversation analysis and socially shared cognition. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 150–171). Washington: American Psychological Association.

Stahl, G. (2002). Contributions to a theoretical framework for CSCL. In G. Stahl (Ed.), *Computer support for collaborative learning: foundations for a CSCL community. Proceedings of CSCL 2000, Boulder, Colorado* (pp. 62–71). .

Suthers, D. D., & Hundhausen, C. D. (2001). Learning by constructing collaborative representations: an emperical comparison of three alternatives. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *Proceedings European Perspectives on Computer-Supported Collaborative Learning* (pp. 577–584). Maastricht: Universiteit Maastricht.

Suthers, D. D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *The Journal of the Learning Sciences, 12*(2), 183–218.

Tynjälä, P., Mason, L., & Lonka, K. (2001). Writing as a learning tool: an introduction. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Writing as a learning tool. Integrating theory and practice* (pp. 7–22). Dordrecht: Kluwer Academic Publishers.

Van Boxtel, C. (2000). *Collaborative concept learning. Collaborative learning tasks, student interaction, and the learning of physics concepts*. Unpublished doctoral dissertation, Utrecht University, the Netherlands.

Van Boxtel, C. (2004). Dialogic physics learning: studying student interaction from three different perspectives. In P. Renshaw & J. L. van der Linden (Eds.), *Dialogic Learning* (pp. 125–144). Dordrecht: Kluwer Academic Publishers.

Van Boxtel, C., & Van Drie, J. (2003, August). *Collaborative reasoning as a key concept for analyzing classroom discourse*. In Paper presented at the 10th European Conference on Learning and Instruction, Padova, Italy.

Van Boxtel, C., & Van Drie, J. (2004, July). *Historical reasoning: a comparison of how experts and novices contextualise historical sources*. In Paper Presented at the History Education International Research Network Conference, Ambleside, United Kingdom.

Van der Linden, J. L., Erkens, G., Schmidt, H., & Renshaw, P. (2000). Collaborative learning. In P. R. J. Simons, J. L. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 37–54). Dordrecht: Kluwer Academic Publishers.

Van Drie, J., Van Boxtel, C., & Van der Linden, J. L. (in press). Historical reasoning in a computer-supported collaborative learning environment. In A. M. O'Donnell, C. E. Hmelo, G. Erkens (Eds.), *Collaborative learning, reasoning and technology*. Hillsdale NJ: Lawrence Erlbaum Associates.

Veerman, A. (2000). *Computer-supported collaborative learning through argumentation*. Unpublished doctoral dissertation, Utrecht University, the Netherlands.

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

Veldhuis-Diermanse, E. (2002). *CSCLearning? Participation, learning activities and knowledge construction in computer-supported collaborative learning in higher education*. Unpublished doctoral dissertation, Wageningen Univeristy, the Netherlands.

Voss, J. F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research, 27*(3), 255–265.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87–122.