Developing a 'big picture': Effects of collaborative construction of multimodal representations in history

Maaike E. Prangsma · Carla A. M. Van Boxtel · Gellof Kanselaar

Abstract   Many pupils have difficulties with the abstract verbal information in history lessons. In this study we assessed the value of active construction of multimodal representations of historical phenomena. In an experimental study we compared the learning outcomes of pupils who co-constructed textual representations, visual-textual representations, or visual-textual representations integrated in a timeline. 85 pupils in pre-vocational secondary education, aged 12–13, worked in dyads on a series of four history tasks. All pupils took a pre-test, post-test and retention test. Results show that working on visual-textual representations integrated in a timeline leads to higher short-term results than co-constructing textual representations. Dialogue analyses for two dyads working in the condition with visual-textual representations integrated in a timeline indicate that the extent to which pupils verbally integrate textual and visual information differs for the four different tasks.

Keywords   Visual representations · Multimodal representations · Collaboration · History learning

Introduction

Think back to your history classes in secondary school. Your knowledge of secondary school history is likely to be fragmentary and unstructured. Shemilt (2000) reports in an evaluation of the Schools History Project in Great Britain: ''few fifteen year-olds are able to map the past; even fewer can offer a coherent narrative [....] for many, the 'event-space' within which [historical narrative frameworks] form and grow is *incoherent* and lacking in order or meaning'' (p. 86). Lee (2005), who interviewed students about historical change,

M. E. Prangsma · C. A. M. Van Boxtel · G. Kanselaar (✉)
Department of Educational Sciences, Faculty of Social Sciences, Utrecht University, PO Box 80140, 3508 TC, Utrecht, The Netherlands
e-mail: g.kanselaar@uu.nl

reports that the students did not show convincing signs of access to an overall framework of the past.

Acquiring a chronological frame of reference is vital in learning and understanding history. However, many pupils have difficulties acquiring a coherent overview of significant historical events and developments, and they confuse phenomena and concepts. Many of the phenomena in history are abstract in nature, and they are related in complex ways. Think, for example, of forms of government, such as democracy.

The prominence of abstract verbal information in history can be a problem for relatively weak readers in lower secondary education (Hacquebord 2004). These pupils are expected to benefit particularly from efficient use of multimodal representations. Much of history can be visualised through different types of representations, such as pictures, cartograms, process and structure diagrams, and, naturally, timelines. The aim of this study is to assess the value of *active construction* of multimodal representations in supporting learners trying to acquire knowledge of historical phenomena in a chronological frame of reference.

Learning with multimodal representations

Research has given us insights into conditions for effective use of pictorial representations in addition to verbal ones, as well as into the processes of mental model construction through multimedia learning. Combining text and visualisations requires translating visual information into verbal information and vice versa, and then relating them to each other. These multimodal representations allow the learner to focus on different aspects of the topic being tackled and on connections between topics and aspects, thus promoting deep learning (Ainsworth 1999). Positive effects have been found within the domain of science and technology and in the context of individual use of presented multiple representations. Most research on learning with multiple representations is based on Paivio's Dual Coding Theory and on Mayer's Cognitive Theory of Multimedia Learning (Paivio 1991; Mayer 2001).

Dual Coding Theory (DCT) assumes that information is processed through one of two channels: the verbal channel or the visual channel. It predicts that adding pictures to text will benefit learning in most cases, as pictures can be processed both verbally and visually. This will result in more elaborate encoding, and the learner is provided with more retrieval cues (Paivio 1991). Mayer's Cognitive Theory of Multimedia Learning (CTML) strongly builds on DCT and makes three main assumptions: (1) information is processed through two separate subsystems for verbal and nonverbal information, (2) meaningful learning involves conscious processing, and (3) there is a limit to the capacity of working memory (Moreno and Valdez 2005). The assumptions lead to a set of seven principles for multimedia learning: spatial and temporal contiguity, coherence, modality, redundancy, individual differences and the multimedia principle (Mayer 2001). In his research, Mayer has found substantive evidence to support his theory (Mayer 2003; Mayer and Sims 1994).

However, although DCT and CTML offer valuable insights, they do have their limitations, most notably in (1) their applicability to different types of representations, (2) their generalisability to different domains, and (3) the tendency of research to focus on presented representations rather than construction of representations by learners. These three limitations will now be briefly discussed.

First, DCT and CTML have been tested on a limited number of different types of representations, mainly process diagrams showing 'how things work' (Mayer 2003; Moreno and Valdez 2005; Lohse et al. 1994). Cox (1999) remarked that a diagram is not

always worth ten-thousand words, because its worth depends on the type of diagram, which meanings it represents, who produced or uses it, and the nature of the task. Schnotz and Bannert (2003) suggest that the structure of depictive representations (as opposed to descriptive representations) directly influences the structure of the mental model constructed from it. In a study with different types of graphics they found that the structure of the representation offered to learners was reflected in the structure of their mental models of the topic. This implies that depictive representations need to be chosen with extreme care to match the class of phenomena, task and the intended mental model. It seems reasonable, then, to make further distinctions between different subtypes of depictive representations. Jones et al. (1988, 1989) present a taxonomy of types of visual representations corresponding to different types of text structures: spider map, series of events chain, continuum/scale, compare/contrast matrix, problem/solution outline, network tree, fishbone map, human interaction outline and cycle. Lohse et al. (1994), on the other hand, developed a structural classification on the basis of ten scales of characteristics – e.g., concrete-abstract, spatial-nonspatial – which resulted in eleven types of visual representations: structure diagrams, process diagrams, maps, cartograms, tables, graphic tables, pictures, icons, time charts, network charts, and graphs. Their taxonomy is research-based, practical and exhaustive and we will use it in our description of the representations in our research.

Second, previous studies have been limited mainly to the domains of secondary school mathematics and physics. Research replicating Mayer's work in the domain of Educational Sciences and Pedagogy (De Westelinck et al. 2005) did not confirm Mayer's multimedia principle and spatial contiguity principle, and De Westelinck et al. propose that different fields of knowledge raise different possibilities for the use of multimodal representations. The role of multimodal representations for history learning remains largely unclear.

Third, most of the research concerned with learning with multiple representations focuses on presented representations rather than on active construction of representations by learners. The mediating function of multiple representations is determined – among other things – by the nature of the activities elicited by the representations (Peeck 1993). So far, research has paid relatively little attention to activities of individual and collaborative construction or adaptation of multimodal representations (Scaife and Rogers 1996). Current trends in the field of learning and instruction stress the importance of active knowledge construction and collaborative learning. Cox (1999) stated that the process of translating information from a linguistic representation to a visual representation might be more effective than translation from one representation to another within the same modality. This idea is consistent with DCT (Paivio 1991). A meta-analysis of Horton et al. (1993) reports a modest positive effect of construction of concept maps on student achievement. Several researchers suggest that multiple representations support deeper understanding when students integrate information from different types of representations (e.g., Ainsworth 2006). Bodemer et al. (2005) found that asking students to actively relate textual components to components of a visualisation had a beneficial effect on learning when the learning material was particularly difficult and complex. However, most studies on learning with multiple representations do not discuss the extent to which learners actively relate textual and visual information.

In addition to fostering knowledge construction, multimodal representations can also function as communicative support in collaborative learning (Reimann 2003). Roth and Roychoudhury (1994) argue that concept mapping as a collaborative activity encourages communication and negotiation of meaning. Concept mapping engages students in discourse on relevant conceptual relationships. The required group product makes students

focus on pivotal principles in the domain, and thus stimulates abstract discussion. Collaborative concept mapping is an open task with no predetermined answers, and this provokes negotiation. The product serves as a visible representation that can facilitate discourse on abstract concepts and relationships. Students can refer to the concept labels and the propositions in the emerging representation while verbalizing their ideas and negotiating meaning. Recent research on collaborative construction of representations has rendered positive results for both learning processes and learning outcomes (Suthers and Hundhausen 2003; Van Drie et al. 2005; Van Boxtel et al. 2000).

The potential of multimodal representations for the acquisition of a chronological frame of reference

Our research focuses on a domain that is relatively unexplored regarding research on learning with multimodal representations: the domain of history. A chronological frame of reference is the knowledge base that is used when reasoning about the past. It consists of knowledge about: (1) historical phenomena, (2) temporal and causal relations, and (3) concepts describing phenomena and relations. Research from Beck and McKeown (1994) has shown that pupils have difficulty developing a coherent chain of events, and that the schemas pupils use are too general to offer ready slots to fit the specific information that they might have gleaned. In addition, the specific information is too sparse to be useful in connecting it to more general information. Furthermore, pupils have particular difficulty forming a notion of complex historical developments and structures (Husbands 1996; Carretero et al. 1991). Making these developments, structures, temporal and causal relations visible through pictures and diagrams can render abstract phenomena and relations more explicit.

The first component of a chronological frame of reference consists of different historical phenomena: the events, structures and themes of an era. Leinhardt (1994) makes a distinction between different types of phenomena that are central to instructional explanations in history classes: events, structures, themes and metasystems. Events are narratives of the actions of people and institutions, limited in time and space, such as revolutions and conquests. Structures are the more constant social elements with descriptive features, for example social class structure or systems of government. Themes are the clarifying notions at the core of historical understanding of people and nations over time, such as tensions between North and South in the United States. Metasystems include the metacognitive tools of history, for example analysis and perspective taking. Events, structures and themes are specific classes of historical phenomena that may require different types of representations. Events, for example, are often represented by narratives, such as a narrative of the rise and fall of the Roman Empire or a narrative of the French Revolution (Husbands 1996). Such narratives can be textual, but they can also be visually represented, for example in a timeline or a comic strip.

The second component of a chronological frame of reference is knowledge of relations between historical phenomena: temporal relations and causal relations. Temporal relations can be represented by a timeline (Hoodless 1996). Constructing timelines can help to sequence events, and to develop awareness of duration and 'key dates' or landmarks (Stow and Haydn 2001). Dawson (2004) emphasizes both the active construction of timelines (instead of looking at completed ones) and the inclusion of images rather than just words and dates. However, a timeline with dates or periods and textual descriptions of historical phenomena only visualizes temporal relationships. It does not show the underlying causal

relationships between the events and phenomena. Moreover, a timeline only visualises temporal relations. Other representations (such as historical pictures, animations, matrices comparing periods, and causal diagrams) might combine with a timeline to facilitate visualization of phenomena and casual relationships. Spatial representations can make it easier to understand relations (Larkin and Simon 1987), and pictures may elucidate historical figures, situations and landscapes. Such a combination of diagrams and pictures is also in line with Friedman (1982), who found that in history children under ten preferred to work with verbal or pictorial representations over spatially organised diagrams, such as timelines. Combining diagrams, pictures and text requires explicit linking of the different representations. The historical phenomena being visualised need to be contextualised in time. Continuity, change and causality cannot be recognised unless the temporal relationships are clear. In addition, causal relationships do not just require insight into temporal relations, but also into the types of phenomena to be explained. Connecting different types of information (that are represented in different representational formats) may thus be crucial in history learning.

The third component of a chronological frame of reference is knowledge of concepts used to describe phenomena and relations. Using historical terminology is an important part of history learning, and it involves both methodological concepts such as change, continuity and causes, as well as substantive concepts, such as feudalism and Enlightenment. Understanding the big picture requires generalisation through a range of abstract concepts (Hunt 2000). Domain specific concepts are tools to question, think about, describe, analyse, synthesise and discuss historical phenomena (Van Boxtel and Van Drie 2004). Therefore, historical concepts are an important component of multimodal representations used to display historical phenomena and relations. In line with the spatial contiguity effect (Mayer 2003), it can be expected that students learn more deeply when relevant historical concepts are placed near the corresponding pictures.

The focus of our research is on the active *construction* of multimodal representations in *collaborative learning tasks* in *history*. We address the following question: What are the effects of the type of constructed representation on the acquisition of a chronological-conceptual frame of reference? We expect to find that construction of a combination of visual and textual representations has a positive effect on the acquisition of a chronological-conceptual frame of reference, and that construction of a combination of visual and textual representations integrated in a timeline has an even larger positive effect, compared to construction of textual representations. The hypotheses are in line with the idea that active construction of visual–textual representations and the integration of verbal and pictorial information helps pupils to develop a 'big picture' of historical phenomena that can be more easily remembered and transferred to new tasks.

Method

Participants and design

The pupils in our study ($N = 143$) were from six different first year classes in three different schools, with one history teacher for each school (pupils aged 12–13). The classes were all in pre-vocational secondary education (VMBO), which a majority of Dutch pupils in secondary school (some 60%) attend. The language proficiency of these pupils is relatively low. History as a school subject is only part of the compulsory curriculum for the first two

years for these pupils, so there is little time for developing a chronological frame of reference.

We conducted an experiment with dyads working in one of three conditions: textual representations (Text); visual-textual representations (Visual); and visual-textual representations integrated in a timeline (Timeline). Scores for the multiple-choice section of the pre-test were used for heterogeneous dyad composition: the pupils were divided into three groups with low, intermediate and high scores. Pupils from the group with the lowest scores and pupils from the group with the highest scores were teamed up with pupils from the group with intermediate scores. In classes where the intermediate scoring pupils outnumbered the low and high scoring ones, intermediate-intermediate dyads were also formed. Due to absence of a number of pupils at the start of the first task lesson, some classes did not contain enough intermediate scoring participants, so several low-low and high-high dyads were also formed. In total, the final sample contains 4 participants from low-low dyads (2 in the Visual condition and 2 in the Timeline condition), and 4 participants from high-high dyads (2 in the Text condition and 2 in the Visual condition). All dyads, except two, consisted of either two girls or two boys.

Given the space, time and attention required for the different conditions we chose to assign the dyads within each class randomly to one of two (instead of all three) conditions. In four out of six classes, the two conditions were put in different classrooms. We made sure that low/intermediate and high/intermediate scoring dyads were evenly distributed over the two conditions. All conditions were given the same amount of time to finish the tasks.

Due to a number of causes, 41% of the original sample had to be discarded: 32% of the Text condition, 49% of the Visual condition, and 37% of the Timeline condition. Pupils or dyads were discarded if one or more of the following reasons applied to them: absence during one or more task lessons; absence during the pre-test or the post-test; one or more tasks were missing or not finished; working individually (due to odd numbers). Pupils were included in the final sample in the following cases: if only the last task was partly missing; absence during the retention test; dyad partner absent during only one lesson. The high proportion of discarded dyads was mainly due to loss of concentration among pupils in school A, who were participating in Ramadan (a Muslim religious festival) during the period the experiment took place. As a result, few dyads in this school managed to finish all tasks. As a result, there is less diversity in the ethnic background of the entire sample (about 9% with a foreign background) than in the general population of Dutch pre-vocational secondary education (24%).

The final sample sizes are shown in Table 1. Although the participants worked in dyads, the table shows some odd totals. This is because participants were individually assessed, and from some dyads only one of the partners could be included for the reasons mentioned above.

Table 1 Number of pupils per condition for each school/teacher in final sample

School/teacher	Text	Visual	Timeline	Total
A	3	1	0	4
B	12	24	10	46
C	15	6	14	35
Total	30	31	24	85

Experimental tasks

The experimental tasks were based on our experiences with a pilot study (see below). Working in pairs, pupils carried out a series of four tasks on the Early Middle Ages during three consecutive history lessons. The period of 500–1000 AD of Western European history – the Early Middle Ages – was selected for several reasons. First, the period includes the full range of phenomena that are dealt with in the history curriculum. Also, the period is representative of the difficulty level of, and the types of relations in other periods in the history curriculum. In addition, the existing curriculum for this period includes different types of developments and social structures that are closely related. It marks a turning point in Western European history with the decline of the Roman Empire and the subsequent development of manorialism, and these two developments can only be understood in relation to each other. A number of important socioeconomic, political and religious changes took place during the Early Middle Ages. The period includes some very abstract concepts, and life during this period was in many ways very different from our pupils' lives. On top of that, there is very little original visual material available from the Early Middle Ages that can help pupils in shaping a notion of this period. The specific task content was chosen on the basis of the 2001 report of the (Dutch) Committee of History and Social Studies that proposes ten eras with their specific aspects for the history curriculum in Dutch schools. In this proposal, the Early Middle Ages are called ''the time of monks and knights'', and its specific aspects are the spread of Christianity, manorialism, and the rise and spread of Islam. The four tasks each had different types of content.

The tasks that were used in the Visual and the Timeline condition were designed according to Mayer's (2003) principles for multimedia learning. According to these principles students learn more when words and pictures are combined (multimedia effect), when extraneous material is excluded (coherence), and when words are placed near a corresponding picture (spatial contiguity). The representations for the four tasks can be categorised as follows according to the taxonomy by Lohse et al. (1994): (1) process diagram (decline of the Roman Empire), (2) network chart (effects of the fall of the Roman Empire), (3) structure diagram (manorialism), and (4) cartograms (spread of Christianity and Islam). Appendix A shows examples of completed answer sheets for Task 1 and Task 3 in the Visual condition. All pictures used for the answer sheets were made especially for the experiments by a professional illustrator. The picture for the third task was in full-colour, while for the other tasks we used black-and-white drawings. The task products in the Timeline condition were linked to each other in a large timeline the size of two sheets of flip chart paper (about 60 × 140 cm). The tasks in the Text condition covered the same content, but only textual answers were required, and no pictures were provided on the answer and instruction sheets. An overview of the tasks is shown in Table 2.

The accompanying texts were the same for all three experimental conditions and these were two to three pages in length each, including appropriate illustrations. Important concepts were printed in bold typeface, and pupils were encouraged to use these concepts in their answers. The tasks were closed off with a summary question, such as: ''Finish the sentence below with two causes and use your answers above and the text: The Roman Empire disappeared because ... and because ... ''. The pilot study mentioned at the beginning of this section ($N = 22$) was done to determine how much time was needed, and to try out the tests and the tasks in two different experimental conditions (Text and Timeline).

Table 2 Contents of the collaborative tasks, and corresponding activities in the experimental groups

Knowledge type and contents	Conditions		
	Text (control group)	Visual group	Timeline group
Task 1 Development: The decline of the Roman Empire	Fill-in-the-blank	Completing a storyboard:	Completing a storyboard:
	Ordering sentences	– Ordering pictures	– Ordering pictures
	Summary question	– Adding concepts and captions	– Adding concepts and captions
		Summary question	Summary question
			Adding text and colour to timeline
Task 2 Causal relations: The effects of the fall of the Western Roman Empire	Fill-in-the-blank	Completing a causal network:	Completing a causal network:
	Ordering sentences	– Choosing pictures	– Choosing pictures
	Summary Question	– Adding concepts and captions	– Adding concepts and captions
		Summary question	Adding arrows to timeline
			Summary question
Task 3 Structure: Manorialism	Answering questions	Describing a historical image:	Describing a historical image:
	Summary question	– Answering questions linked to details in the picture	– Answering questions linked to details in the picture
		Summary question	Summary question
			Adding arrows to timeline
Task 4 Development: Religion: spread of Christianity and Islam	Fill-in-the-blank	Completing two cartograms:	Completing two cartograms:
		– Fill-in-the-blank in tags linked to specific points on the maps	– Fill-in-the-blank in tags linked to specific points on the maps
		– Colouring the spread of religions in the maps	– Colouring the spread of religions in the maps
		Summary question	Summary question
			Adding text and colour to timeline

Tests

The pupils were given the same individual knowledge test three times: a pre-test, a post-test, and a retention test. We included a retention test to investigate long-term effects. The test consisted of three subtests that were administered separately to prevent questions giving away each other's answers. To ensure coverage of the learning goals, the test was based on a test matrix, which in turn was based on a matrix of the learning goals. Subtest A was a free association spider on the Early Middle Ages. Subtest B consisted of open questions, including both textual and visual questions (e.g., with pictures, maps, or timelines). Subtest C consisted of different types of multiple-choice questions (e.g., multiple

response questions, choosing which came first in time). A total of 11 items were devoted to questions on chronology: 3 items in subtest B, and 8 items in subtest C. An overview of the test is shown in Table 3. Some examples of test questions are shown in Appendix B.

We used Cronbach's alpha to determine test reliability. Table 4 shows Cronbach's alpha for the pre-test, post-test and retention test. Prior knowledge was low, so pupils had to resort to guessing on the pre-test, resulting in low homogeneity. Cronbach's alphas for the post-test and retention test were not high but were still considered acceptable. One item from subtest B was excluded from further analyses, because it had zero variance in the pre-test, and little variance in the post-test and retention test. This item asked about the Arab Empire, while most pupils seemed unable to disconnect the concept *empire* from the Romans. Table 4 shows the test reliability after deleting this item.

The three subtests were scored on the basis of the task text content. The interrater reliability between two raters for subtest B on 74 randomly chosen tests from different classes (28% of the total number of scored subtests B) was .89 (Cohen's kappa).

Setting and procedure

The pre-tests were administered four or seven days before the start of the experiment. The total time taken by pupils varied from 30 to 40 min. Pupils were given brief instructions about the research study and the tasks both during the pre-test lesson and at the start of the experiment. For all classes, the experiment was started in the first history lesson after administering the pre-test.

The classes were divided over two classrooms by condition. One of the researchers monitored one condition, and an assistant monitored the other condition, while the teacher switched classrooms from time to time. Apart from a short introduction before splitting up in conditions, there was no classroom instruction, nor did the teacher, researcher or assistant give feedback on content, only on completion (''Is it finished, yes or no?''). The participants started each task by reading a text on the task topic; the same text was used for all conditions. After reading, the dyads were given the task sheets and instructions. When a dyad had completed one task, they were given the next. The pupils had three lessons (about

Table 3 Overview of test and subtests

Test unit	No. of items	Maximum score	Response format
A	1	no maximum	Association spider
B	8	8	Short answer
C	18	18	Multiple-choice
Total	27	no maximum	

Table 4 Results of reliability analysis for pre-tests, post-tests, and retention tests

Test	Cronbach's Alpha*
Pre-test ABC	.31
Post-test ABC	.58
Retention test ABC	.66

* Standardised item alpha

150 min) to finish all four tasks. The conversation by the dyads was audio taped with recorders placed on their desks.

The post-test was administered either at the start of the first history lesson after rounding off the experiment, or the next day at the start of a lesson by the class mentor. The retention test was administered 33–49 days after the post-test ($M = 37$).

Analysis of student dialogue

From the final sample of 85 participants, two Timeline dyads (4 participants) were selected for a closer look at the discourse, in particular to get an idea of the occurrence of integration of text and representations during work on the different tasks. The dyads were selected from all Timeline dyads from the final sample. The chosen dyads had the most complete protocols (i.e., protocols were available for all four tasks). Also, the participants in these two dyads showed a strong increase in their scores between pre-test and post-test, so we expected to find indicators of relating textual to visual information in the student dialogues. Dyad A consisted of two boys: Allan (pre-test 14, post-test 23, retention test 20) and Adrian (pre-test 14, post-test 24, retention test 26). Dyad B consisted of two girls: Bridget (pre-test 7, post-test 14, retention test 11) and Betty (pre-test 12, post-test 27, retention test 20).

For each of the two dyads selected, the dialogues were typed out, coded and analysed with utterances as the unit of analysis. The protocols were analysed in several steps. First, the utterances were coded for their basic topic: Content, Procedure, Social talk or Other. The focus in coding and analysing the interaction processes was on the content related part of the discourse. Content propositions included utterances about historical phenomena and relations, about pictures, or about the answers to be given on the answer sheet. Procedural propositions referred to physical characteristics of the task materials or to the spelling of the answer, or they were utterances for regulating the collaboration or the behaviour of the partner. Social talk included all utterances by the dyad partners that were irrelevant to the task. The category Other included utterances by other participants, by the teacher, or by the experimenter. Unintelligible utterances were also assigned to the Other category. The Procedure, Social talk and Other utterances were not investigated further, but served as a valuable context for interpreting the Content propositions. Examples of these categories are shown in Table 5.

The next step was to indicate which utterances were passages taken directly from the text, the instructions, or the answer sheet. As the participants did not originate these

Table 5 Examples for the Social talk, Procedural, and Content coding used for coding the dialogues

Code	Examples	
Social talk	''Are you going to the party on Friday?''	''You know who called last night?''
Procedural utterances	''It's your turn now.''	''Do you have sticky tape?''
Content utterances	''There was little trade''	''The bridge has collapsed''
	''There was less trade''	''When the Romans left, bridges collapsed''
	''Trade''	''Viking boats are cool''

utterances, they were not coded within the Content category. Again, these utterances provided a context for interpreting the Content propositions. The steps thus described were also used to code and analyse a larger set of 20 dyad dialogues of Task 2 (see Prangsma et al. 2007). The interrater reliability between two coders was calculated for four dialogue protocols (totalling 1060 utterances) and amounted to .74 (Cohen's kappa).

Finally, the Content propositions were coded for Integration, which was understood when information from the text was related to (a part of) the multimodal representations (e.g., schemas, pictures, the timeline itself) on the answer sheets. The dialogues were coded by the first two authors, and any disagreements were discussed and decided upon by them. Some dialogue fragments with integrative utterances are shown in Table 6.

Hypotheses

We compared the learning outcomes of pupils who co-constructed textual representations, visual-textual representations, and visual-textual representations integrated in a timeline.

Table 6 Dialogue fragments with integrative utterances

Speaker	Relating text and timeline – Colouring and labelling the timeline (Task 1)	Integration
Allan	Colour it yellow. The part that goes with Antiquity	0
Allan	Of all Antiquity it is	1
Adrian	But this is the Early Middle Ages and Antiquity is this	1
Allan	Then that is from 500 ... from 400 to 500 I think	1
Adrian	Yes	0

Speaker	Relating text and pictures – Describing and sequencing historical phenomena (Task 1)	
Allan	Look, the Romans	1
Allan	This and this I think	1
Allan	But it is all ...	0
Adrian	But here it is split, that is almost at the end	1
Adrian	And here is the fighting	0
Adrian	And that comes then too, and I think it is like this	1

Speaker	Relating text and pictures – Describing causal relations between depicted phenomena (Task 2)	
Betty	Ehm, but is this an effect of that, ehm that the effect of that?	1
Betty	Ehm, but this is actually, this is actually the effect of this one.	1
Betty	Because there was no protection anymore ...	0
Teacher	Yes	0
Betty	... there was looting.	0

Speaker	Relating text and geographic map – Colouring and labelling the map (Task 4)	
Betty	Here it says you have to colour, ehm, red that were Christian during the Early Middle Ages. That's this, right?	1
Bridget	That is ... Christian part	1

The visual-textual representations that pupils constructed, integrated historical concepts, historical phenomena and relations (the components of a chronological-conceptual frame of reference). The considerations above lead to the following hypotheses. First, we expect that the pupils who co-construct visual-textual representations that combine pictures, diagrams and historical concepts will gain more historical knowledge than pupils who co-construct textual representations, because information will be processed both verbally and visually, and verbal and visual information will be integrated. Second, we expect that pupils who integrate different visual-textual representations into an overall representation (a timeline) will gain more historical knowledge than pupils who co-construct textual representations and pupils who co-construct visual-textual representations without integration, because in this condition temporal relations are also visualised. Third, we expect that these differences will endure over a longer time span.

Results

Tests

We present the results of our analyses of pupils' performance in the three conditions.

Table 7 shows the mean scores and standard deviations on the pre-test, post-test and retention test in the three conditions. Pre-test data were examined to identify initial differences in prior knowledge scores between the conditions. An ANOVA indicated that the three conditions did not differ significantly from each other in their mean pre-test scores ($F(2, 82) = .32$, $p = .72$). Although the pre-test turned out to have problems of reliability, the three conditions did not differ significantly on pre-test results. There is no obvious reason to suspect that the groups differed at pre-test.

Before comparing the test results of the three conditions, we investigated a preliminary question: Did the participants actually learn something from the tasks? For each condition, we checked this with paired samples T-tests between pre-test, post-test and retention test. The paired samples T-tests showed that in each condition the pupils' score increased significantly from pre-test to post-test ($p = .00$) and from pre-test to retention test ($p = .00$). There was a significant decrease from post-test to retention test for both the Visual ($t(31) = 4.34$, $p = .00$ (two-tailed)) and the Timeline condition ($t(21) = 3.23$, $p = .00$ (two-tailed)), but not for the Text condition ($t(29) = 1.79$, $p = .08$ (two-tailed)).

As we found no significant differences between the conditions on the pre-test, we went on to check for differences between conditions on the post-test and retention test scores. The ANOVA for the post-test showed a significant difference between the three conditions ($F(2, 82) = 3.66$, $p = .03$). Post-hoc analysis (Bonferroni) showed that this is attributable to the significant difference between the Timeline and the Text condition, with a mean

Table 7 Means and standard deviations for test scores in the three conditions

	Text			Visual			Timeline		
	N	M	SD	N	M	SD	N	M	SD
Pre-test	30	12.13	2.56	31	11.55	2.71	24	11.79	3.32
Post-test	30	15.83	4.08	31	17.58	4.40	24	19.21	5.32
Retention test	30	14.87	4.54	29	15.17	4.44	22	16.09	5.49

difference of 3.38 (p = .03). The Timeline group performed significantly better on the post-test than the Text group. On average, the scores of the Visual group on the post-test were 1.75 higher than the Text group scores, but this difference was not significant (p = .58). We did not find significant differences between conditions for the retention test scores, which means there is no difference in long-term effects ($F(2, 78)$ = .43, p = .65).

We suspected that the difference in post-test scores between the Text and Timeline groups might be attributable mainly to questions related to temporal relations, so we did an exploratory analysis for the section of the test with the 11 time-related questions. After all, there were no significant differences for the Visual group, and that leaves the built-in emphasis on temporal relations as the main distinguishing factor between the Text and Timeline tasks. The ANOVA results for the time-related questions in the post-test – taken both from subtest B and subtest C – again showed a significant difference between conditions ($F(2, 82)$ = 4.71, p = .01), and again, the post-hoc tests (Bonferroni) showed a significant difference only between the Text and Timeline groups (p = .01) with a mean difference of 1.45. These results suggest that the difference between the Text and Timeline groups might be attributable to this section of the test. Again, the difference disappears in the long run ($F(2, 78)$ = .26, p = .77). Analyses of just the non-time-related questions shows no significant differences between conditions (post-test: $F(2, 82)$ = 1.89, p = .16; retention test: $F(2, 78)$ = .46, p = .64).

Although the participating teachers (there was only one teacher per school) had pupils from all conditions, and pupils and dyads were carefully (but not completely randomly) assigned to conditions, we decided to check whether the distribution of conditions over classes could have interfered with the results. We used independent samples T-tests and found that the significances and directions of the significances for the separate classes were the same as for the entire sample.

We found no significant differences, then, between the Visual and the Text conditions, nor between the Visual and Timeline conditions. On the other hand, we did find a significant difference in time-related questions between the Text and Timeline conditions. It seems then that the visual support with pictures and schemas in general might not be the distinguishing factor we expected, whilst the combination of pictures and schemas with specific visualisation of time does seem to make a difference, at least in the short run.

Products and process

To gain more insight into the learning processes, we first took a closer look at the use of the multimodal representations in the group products to see how the pupils in the Visual and Timeline conditions dealt with the pictures incorporated in them. The products revealed that especially in the first and fourth tasks some dyads came up with interpretations that did not match their intended meaning or the information in the text. In the first task the pupils completed a storyboard about the decline of the Roman Empire by ordering drawings and adding concepts and captions. One of the drawings in this task showed Roman soldiers who walk away from a ruin. Some of the dyads did not relate this drawing to the departure of Roman soldiers from the provinces back to Italy after the Western-Roman Empire fell. These dyads came up with descriptions such as ''the armies revolted'', ''wandering of nations'', and ''they had built roads and bridges''. One dyad described the drawing that showed the division of the Roman Empire as showing that ''the Romans had conquered almost everything they wanted to conquer''. One picture contained a map of the Eastern Roman Empire with its emperor on the right, and an empty space and an empty chair on the

left (a representation of the fact that there was no Western Roman Empire anymore). One dyad associated this picture with trade and wrote the following caption: ''There were large distances between countries, so for trade as well''.

In the fourth task it appeared that many pupils found it difficult to colour the spread of Christianity and the spread of Islam on a map. Pupils were asked to first colour the spread of Christianity until 500 AD in map A, and then colour the spread of Christianity from 600 to 1000 AD and the spread of Islam from 600 to 1000 AD in map B. The pupils were provided with a map that combined these changes. Many dyads did not execute the task as it was intended. Some of the dyads simply copied the colours of the given map both in map A and map B – which was the wrong answer – others coloured more than was asked for in the instruction or did not colour the correct parts of the map. Thus, it seems that pupils had difficulties reading a complex map and using it to construct two new maps. We may have over-estimated pupils' map reading and construction skills.

Our premise was that the Timeline condition would do better than the Text condition on the tests because these participants had to integrate the textual and verbal information within the tasks. In addition, they had to integrate the information from the different tasks into a larger whole – a timeline. Table 8 shows the results of a closer examination of the number of utterances by task for each of the two dyads, as well as the percentage of integrative utterances in the sections of the dialogues where the participants are working on the timeline itself – as opposed to just one task sheet. The differences between the two dyads might be explained by their pre-test levels: Dyad A consisted of two intermediate scoring pupils, whilst Dyad B consisted of an intermediate and a low scoring pupil.

A closer look at integration of text and representations in Timeline dyads A and B showed that Task 1 – which involved completing a process diagram on the decline of the Roman Empire by sorting pictures and adding text – elicited the most integrative utterances

Table 8 Number of utterances by task for each dyad, and percentage of integrative utterances in the timeline sections of the dialogues

	Task 1 Storyboard		Task 2 Causal network		Task 3 Image		Task 4 Cartogram	
	Dyad A	Dyad B	Dyad A	Dyad B	Dyad A	Dyad B	Dyad A	Dyad B
Procedural	72	50	56	33	41	21	61	49
Social talk	53	4	53	8	36	6	58	7
Content, of which:	197	140	109	57	81	31	146	131
–Integration frequency[a]	45	37	28	14	12	3	29	13
–Integration %[b]	22.84	26.43	25.69	24.56	14.81	9.68	19.86	9.92
–Integration in timeline frequency[c]	11	5	0	0	9	3	0	2
–Integration in timeline %[d]	5.58	3.57	0.00	0.00	11.11	9.68	0.00	1.53
Total number of utterances	322	194	218	98	158	58	265	187

[a] Integration frequency: The total number of content utterances that show integration

[b] Integration %: The percentage of content utterances that show integration

[c] Integration in timeline frequency: The total number of content utterances that show integration and refer to the timeline

[d] Integration in timeline %: The percentage of content utterances that show integration and refer to the timeline

in both dyads (45 and 37, respectively). On the other hand, Task 3 – which involved adding captions to parts of a structure diagram on manorialism – elicited very little integration (12 and 3, respectively), and most or even all (Dyad B) of these integrative utterances are found when the participants are occupied with linking Task 3 to the timeline (11.11% and 9.68% of all Content utterances, respectively).

Since we found a significant difference between the Text and Timeline conditions on the post-test for the timeline-related test questions, we expected to find a substantive amount of integration during the part of the task that involved the overall timeline. Table 8 also shows the frequency and percentage of integrative utterances in the timeline sections of the dialogues. The numbers differ widely between the tasks. Whilst for Task 3 almost all of the little integration we found concerns the timeline, for Task 4 we found very little timeline integration, and for Task 2 we found no timeline integration talk at all. Again, Task 1 shows slightly higher numbers than the other three tasks. It is possible that the combination of the type of representation, the types of activities, and the type of timeline activities involved specifically in Task 1 – a process diagram, sorting pictures and colouring and labeling the timeline – encouraged integration more than the representation types and activities of the other three tasks.

Conclusion and discussion

The aim of this study was to investigate the effects of the type of constructed representation on the acquisition of a chronological frame of reference. We compared the learning outcomes of pupils who – after reading a text on the Early Middle Ages – worked in one of the following conditions: (1) co-construction of textual representations; (2) co-construction of visual-textual representations; and (3) co-construction of visual-textual representations integrated in a timeline. Our hypotheses were partly confirmed. The pupils who integrated the visual-textual representations in a timeline outperformed the pupils in the textual representation condition on the post-test, but not on the retention test. Considering the fact that the pupils in the Timeline condition scored higher particularly on the time-related items of the test (marking an event on a timeline or ordering events chronologically), we assume that the timeline representation resulted in higher learning outcomes, because it also made temporal relations explicit. An explanation for the disappearance of this effect in the long run (as measured with the retention test), might be that knowledge about temporal relations is more difficult to retain. Furthermore, a timeline that is studied or constructed may be difficult to remember and to reconstruct from memory in a new task. A suggestion for further research is to examine the influence of different representational formats on long-term learning effects as well as on the learning process.

Contrary to our expectations pupils who co-constructed the visual-textual representations did not show significantly higher scores on the post-test and retention test than pupils in the textual representation condition. This may suggest that collaborative construction of visual-textual representations without integration in a timeline does not help pupils more in developing a 'big picture' of historical phenomena (events, structures, and themes) than the construction of textual representations. However, we need to be careful in drawing this conclusion. First, although the visual-textual representation group did not show significantly higher post-test scores than the textual representation group, the means point in the direction of our hypothesis. Second, although the pictures were meant to give a more concrete representation of abstract phenomena and abstract relations, a closer look at the

group products revealed that some dyads had difficulty with understanding some of the pictures that were used, as well as with reading and producing maps showing historical developments. De Westelinck et al. (2005) found the same problem with the iconic sign system used in their research. A possible drawback of pictures in comparison with texts might be that it is more difficult to predict the kind of associations that are evoked. This also makes it difficult to control construction activities with these representations. Some situations and developments in history are difficult to represent with a low degree of ambiguity in a drawing or diagram and might be better represented by means of a story, film or a computer animation. This is a problem that springs from the unclear semiotics of the domain of history. Even when the context of a picture is known, for example the Early Middle Ages, pictures are often open to multiple interpretations. Thus, misinterpretation of the pictures may have resulted in inaccurate or inappropriate mental representations. The natural ambiguity of the domain also makes it hard for students to learn to understand the forms of representation – the importance of which is underlined by Ainsworth (2006): there are no fixed formats in history. Further research can shed light on the associations that are evoked by different types of pictures.

When we look at Mayer's seven principles for multimedia design (2007), we find that only one is partially confirmed by our data: the multimedia principle. This principle states that a combination of text and visual representations results in higher learning outcomes than just text. Simply a combination of text and visual representations did not have the effect predicted by Mayer's principle, but we did find that a combination of text and visual representations integrated in a wider overview had more positive effects than just text. However, the circumstances of Mayer's experiments have so far been quite different from those in our research: (a) his experiments were very short, whereas ours lasted four sessions, (b) his participants were college students, whereas ours were pupils in prevocational secondary education, (c) his experiments took place mainly in lab settings, whereas ours were set in normal classrooms, embedded in the curriculum, and (d) the topics of his tasks were very physical in nature, whereas ours were quite abstract and more difficult to represent, and (e) his tasks involved presented representations, whereas ours required construction activities by the learners. The principles of CTML might not work in quite the same way under these different circumstances, as was the case in the study by De Westelinck et al. (2005).

The dialogue analyses seem to support the idea that learning with multimodal representations is to a large extent mediated through active relating of information from different representations (in this study: textual and visual information). The discussions by the two dyads – both of which have learned relatively a lot from the tasks – show quite a few utterances that reflect active integration of textual and visual information. However, the analyses also showed large differences between tasks. Although all four tasks were designed to encourage the pupils to relate and integrate the information from the text and the visual representations, it seems that Task 1 (completing a storyboard about the fall of the Roman Empire) elicited a lot of active integration, whilst Task 3 (describing an image representing manorialism) elicited very few integrative utterances. Possibly, Task 1 stimulated active integration more strongly because the integration was scaffolded more: first the pictures are discussed using historical concepts from the text, and then the pictures are put in the correct order. Both activities require that textual information (historical concepts, causal and temporal relations) are related to visual information. In Task 3, pupils answered questions about elements of an image, to which the correct answers were concepts from the text. Thus, it was perhaps not strictly necessary to connect the text to the image, and connecting it to the questions was sufficient, even though the questions did

refer to the image. Out of the four tasks offered, Task 3 required the least varied activity from the learners, and fewer relations had to be drawn within the representation than in the other three tasks. It is possible that this task was too easy, that too much information was given away by the guiding questions, or that the picture was not really needed to find the answers to these questions. Future research should focus on the extent to which pupils working on multiple representations connect the information from the different representations—in different task types, representation types, and domains - and how the occurrence or non-occurrence of such connecting activities is related to learning outcomes.

The multimodal representations that pupils worked on in this study combined visualisations of historical phenomena (e.g., the split of the Roman Empire, trade by barter) and relations (causal, temporal). We cannot draw conclusions about the effects of such a combination from the results of this study. In a follow-up study, we will investigate whether representations that combine visualisation of both phenomena and relations can contribute more to the acquisition of historical knowledge than visualisation of only the phenomena or the relations. In conclusion, there are indications in our study that integration and visualisation of historical phenomena, concepts, and causal and temporal relations helps pupils to acquire knowledge of an era.

Appendix A

Model answers for Task 1 and Task 3 (Figs. 1, 2)

Part 1: The Roman Empire disappears

What were things like around the year 400? Concepts that go with this period: 1) *Antiquity* 2) *Roman army* 3) *Emperor*	Concept: *transmigrating people* What happened? *Peoples from Eastern Europe transmigrated into the Roman Empire*	Concept: *split of Roman Empire* What happened? *The Roman Empire was split into a Western part and an Eastern part*	Concept: *Western Empire gone* What happened? *The emperor of the Western part is dethroned. The Western Roman Empire disappears*	Concept: *the Romans leave* What happened? *The Roman army and administrators move back to Italy*

The Roman Empire disappeared because *the western part had weak administrators* and because *Germanic people conquered Rome*

Fig. 1 Model answers for Task 1 in the Visual condition

Part 3: Manorialism
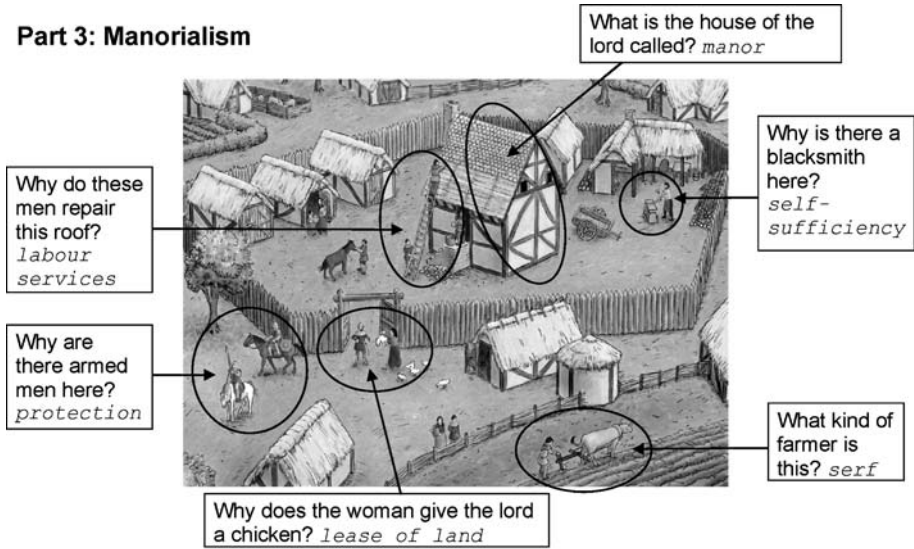
**Fig. 2** Model answers for Task 3 in the Visual condition

Appendix B

Example test questions

B.3

It is the year 54. The monk Bonifatius travels through Europe. He came from Ireland, but now he is in Friesland. In the picture you can see Bonifatius with his **bible** in his hand. What was the monk Bonifatius doing in Northern-Europe?

B.5

Around what year did the Early Middle Ages begin? And when did they end? Colour the period in the timeline below. Write the name 'Early Middle Ages' above it.

1	500	1000	1500	2000
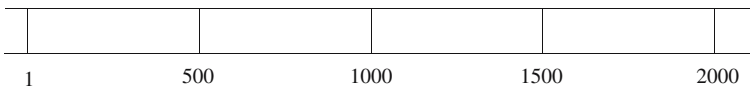
C.1

The great Roman Empire disappeared. What caused this to happen?

Tick the **two** correct answers.
❑ The empire was increasingly badly governed.
❑ There was no longer any long-distance trade.
❑ Germanic peoples wandered into the empire.
❑ The Roman Empire became Christian.
❑ The Arabs marched into the empire.
❑ Roads and bridges were no longer cared for.

C. 8

What came first? Each time, circle what came **first**.
What came first: **Roman Empire** or **manorialism**
What came first: **wandering of nations** or **rise of Islam**
What came first: **Christianity in the Netherlands** or **disappearance of Roman Empire**

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33*, 131–152.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*, 183.

Beck, I. L., & McKeown, M. G. (1994). Outcomes of history instruction: Paste-up accounts. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 237–256). Hillsdale, NJ: Lawrence Erlbaum.

Bodemer, D., Ploetzner, R., Bruchmüller, K., & Häcker, S. (2005). Supporting learning with interactive multimedia through active integration of representations. *Instructional Science, 33*, 73–95.

Carretero, M., Asensio, M., & Pozo, J.I. (1991) Cognitive development, historical time representation and causal explanations in adolescence. In M. Carretero, M. Pope, R.J. Simons & J.I. Pozo (Eds.) *Learning and inswdtruction: Vol. 3. European research in an international context* (pp. 27–48). Oxford: Pergamon.

Committee of History and Social Studies. (2001). *Verleden, heden en toekomst (Past, present and future).* Enschede, The Netherlands: SLO, Netherlands Institute for Curriculum Development.

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

Dawson, I. (2004). Time for chronology? Ideas for developing chronological understanding. *Teaching History, 117*, 14–24.

De Westelinck, K., Valcke, M., De Craene, B., & Kirschner, P. (2005). Multimedia learning in social sciences: limitations of external graphical representations. *Computers in Human Behavior, 21*(4), 555–573.

Friedman, W. J. (1982). Conventional time concepts and children's structuring of time. In W. J. Friedman (ed.), *The developmental psychology of time* (pp. 171–208). New York: Academic Press.

Hacquebord, H. (2004). Taalproblemen en taalbehoeften in het voortgezet onderwijs: Leerlingen- en docentenvragenlijsten als instrumenten voor taalbeleid (Language problems and language needs in secondary education: student and teacher questionnaires as instruments for language policy). *Levende Talen, 5*(2), 17–28.

Hoodless, P. (1996). *Time and timelines in the primary school (Vol. 69).* London: The Historical Association.

Horton, P. B., McConny, A. A., Gallo, M., Woods, A. L., & Hamelin, O. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Science Education, 77*(1), 95–111.

Hunt, M. (2000). Teaching historical significance. In J. Arthur, & R. Phillips (Eds.), *Issues in history teaching* (pp. 39–53). London: Routledge Falmer.

Husbands, C. (1996). *What is history teaching? Language, ideas and meaning in learning about the past.* Buckingham, UK: Open University Press.

Jones, B. F., Pierce, J., & Hunter, B. (1988, 1989). Teaching students to construct graphic representations. *Educational Leadership, 46*(4), 20–25.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*(1), 65–99.

Lee, P. (2005). Historical literacy: Theory and research. *International Journal of Historical Learning, Teaching and Research, 5*(1), 25–40.

Leinhardt, G. (1994). History: a time to be mindful. In G. Leinhardt, I. L. Beck, & C. Stainton (Eds.), *Teaching and learning in history* (pp. 209–256). Hillsdale, NJ: Lawrence Erlbaum.

Lohse, G. L., Biolsi, K., Walker, N., & Rueter, H. H. (1994). A classification of visual representations. *Communications of the ACM,* 37(12), 36–49.

Mayer, R. E. (2001). *Multimedia learning.* Cambridge, UK: Cambridge University Press.

Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction, 13*(2), 125–139.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389–401.

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research & Development, 53*(3), 35–45.

Paivio, A. (1991). *Images in mind: The evolution of a theory*. London: Harvester Wheatsheaf.

Peeck, J. (1993). Increasing picture effects in learning from illustrated text. *Learning and Instruction, 3*(3), 227–238.

Prangsma, M. E., Van Boxtel, C. A. M., Kanselaar, G., & Kirschner, P. A. (2007). *Multimodal representations in collaborative history tasks: A look at student dialogues*. Manuscript in preparation.

Reimann, P. (2003). Multimedia learning: beyond modality. *Learning and Instruction, 13*(2), 245–252.

Roth, W.-M., & Roychoudhury, A. (1994). Science discourse through collaborative concept mapping: New perspectives for the teacher. *International journal of science education, 6*(4), 437–455.

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work?. *International Journal of Human-Computer Studies, 45*(2), 185–213.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*(2), 141–156.

Shemilt, D. (2000). The caliph's coin The currency of narrative frameworks in history teaching. In P. N. Stearns, P. Seixas, & S. Wineburg (Eds.) *Knowing, teaching, and learning history: National and international perspectives* (pp. 83–101). New York: NYU Press.

Stow, W., & Haydn, T. (2001). Issues in the teaching of chronology. In J. Arthur, & R. Phillips (Eds.), *Issues in history teaching* (pp. 83–97). London: Routledge Falmer.

Suthers, D. D., & Hundhausen, C. D. (2003). An empirical study of the effects of representational guidance on collaborative learning. *Journal of the Learning Sciences, 12*(2), 183–219.

Van Boxtel, C. A. M., & Van Drie, J. (2004). Historical reasoning: A comparison of how experts and novices contextualise historical sources. *International Journal of Historical Learning, Teaching and Research, 4*(2), 89–97.

Van Boxtel, C. A. M., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, *10*(4), 311–330.

Van Drie, J., Van Boxtel, C.A.M., Jaspers, J., & Kanselaar, G. (2005). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior, 21*(4), 575–602.