British Journal of Educational Psychology (2009), 79, 371–387

The British Psychological Society

www.bpsjournals.co.uk

# Concrete and abstract visualizations in history learning tasks

Maaike E. Prangsma[1]*, Carla A. M. van Boxtel[2], Gellof Kanselaar[3] and Paul A. Kirschner[4]

[1]Amsterdam School of Communications Research, University of Amsterdam, Amsterdam, The Netherlands
[2]The Graduate School of Teaching and Learning, University of Amsterdam, Amsterdam, The Netherlands
[3]Research Centre Learning in Interaction, Utrecht University, Utrecht, The Netherlands
[4]Netherlands Laboratory for Lifelong Learning, Open Universiteit Nederland, Heerlen, The Netherlands

**Background.** History learning requires that students understand historical phenomena, abstract concepts and the relations between them. Students have problems grasping, using and relating complex historical developments and structures.

**Aims.** A study was conducted to determine the effects of tasks with abstract and/or concrete visualizations on the learning of historical developments and structures. The hypothesis was that students receiving visualizations would learn and retain more historical knowledge and concepts than those not receiving visualizations.

**Sample.** First-year pupils in vocational middle school ($N = 104$) worked in randomly assigned pairs.

**Methods.** After reading a text, the pairs were given a learning task in one of four conditions: Textual, Concrete visualized, Abstract visualized, and Combined.

**Results.** Post-test and retention test results showed no significant differences. There were some significant differences on the evaluation questionnaire.

**Conclusions.** Combining text and different types of visualizations in learning tasks does not necessarily enhance history learning. Possible explanations given are the ecological setting, the semiotics of the domain of history – that are not defined clearly – and the difficulty of unequivocally visualizing historical concepts.

Though modern history schoolbooks are often characterized by an impressive number of pictures, little research has been done into the workings and value of visual support within the domain of history. Distinctive features of history as a school subject are

*Correspondence should be addressed to Maaike Prangsma, University of Amsterdam, Amsterdam School of Communications Research, Kloveniersburgwal 48, 1012 CX Amsterdam, The Netherlands (e-mail: M.E.Prangsma@uva.nl).

DOI:10.1348/000709908X379341

historical phenomena (such as the events, structures and themes of an era), the temporal and causal relationships between them, and the concepts that describe these phenomena and concepts. Students have problems grasping, using, and relating complex historical developments and structures (Carretero, Asensio, & Pozo, 1991). Visualizations in other domains have been shown to have added value for learning compared to purely textual representations. Peeck (1993), for example, discusses several studies that show that presenting appropriate pictures alongside text increases understanding and memorization, and Mayer's (2001) Cognitive Theory of Multimedia Learning argues that visualizations can be powerful learning tools. These positive effects of visualizations are often explained by the Dual Coding Theory (Paivio, 1991), which assumes that information is processed through one of two channels – verbal or visual – and predicts that adding pictures to text will benefit learning in most cases, as pictures can be processed both verbally and visually, resulting in more elaborate encoding and the availability of more retrieval cues to the learner. Both Peeck and Mayer, however, focus mainly on the domain of the natural sciences. The humanities and social sciences – such as history and geography – remain underrepresented in the corpus of research on learning with visual and multimodal representations.

The type of information represented can strongly determine a representation's suitability for achieving its goal. O'Donnell, Dansereau, and Hall (2002) state that more research is needed on the match and mismatch between knowledge maps (e.g. causal or hierarchical schemas) and the macrostructure of the information they represent. It is possible that this effect can be extended to other types of visualizations. Schnotz and Bannert (2003) found that, for the success of visualizations of time-differences in geography, it is essential that the representation type used fits the knowledge type asked for. For example, in their study on different types of world maps showing time zones they found that it is easier to calculate time differences with a carpet diagram than with a circle diagram. Likewise, Butcher (2006) found that visual representations seem most successful when they are designed to support the specific cognitive processes needed for deep understanding; that is, when there is a match between representation type and knowledge type. Different domains have different needs. A timeline, for instance, will not often be found in geography schoolbooks, but it is very appropriate for history, because it visualizes the temporal relations necessary for building a coherent representation of the past.

Visual representation types have been classified in different ways. One common dimension is the classification of visualizations as abstract vs. concrete (e.g. Lohse, Biolsi, Walker, & Rueter, 1994). Concrete and abstract visualizations each have advantages and disadvantages. Concrete visualizations have a strong resemblance to objects in the real world, for example, photographs and realistic drawings. They may also be easier to interpret than abstract visualizations, as they require little understanding of abstract visual conventions. On the other hand, they seem less appropriate for visualizing structural and relational information. Concrete visualizations such as drawings also differ in the extent to which they show a realistic and detailed image of a phenomenon.

Original historical visualisations, such as photographs or paintings, can give a clear image of a historical phenomenon. For instance, building styles, specific tools, religious objects, or period costumes can help shape this image. However, the educational value of such concrete visualisations is often limited by problems of source reliability (after all, these visualizations are the creator's interpretation of objects or events), redundancy of information (e.g. decorative elements), the use of period-specific symbols (e.g. a dog in a medieval picture represents loyalty) and students' lack of experience with the general visual language of a particular period (e.g. frescos, icons or romantic paintings; Husbands, 1996; Sauer, 2000). Realistic drawings made especially for educational purposes might be a good alternative, as redundant information can be left out and complex information can be simplified. Within the domain of the natural sciences, such drawings combined with text seem to be beneficial for learners. Butcher (2006) investigated students learning about the heart and circulatory system using either a simplified drawing that highlights structural relations, or a more detailed and realistic visualization. She found that simplified drawings best support factual learning and information integration.

Abstract visualizations show information units in a way that does not resemble tangible objects, but rather focus on certain aspects of the information, often containing visual elements whose meaning is based on convention, such as arrows in a flowchart, or colours used to show altitudes on geographic maps. Some examples of abstract visualizations are causal diagrams – that focus on the causal relations between the components of the information unit – and flowcharts – where each item has a conventional meaning ('file', 'defer', etc.). O'Donnell *et al.* (2002) describe several advantages of more schematic representations, which they call knowledge maps. These maps can focus attention on the macrostructure of a body of information. Using knowledge maps resulted in higher recall of main issues in comparison with using text, and the maps seem to be especially supportive for students with weak verbal skills. Other studies on the use and construction of concept maps have also shown that abstract visualizations can support learning, both of concepts and of relations (Fischer, Bruhn, Gräsel, & Mandl, 2002; Robinson, Robinson, & Katayama, 1999; Van Boxtel, Van der Linden, Roelofs, & Erkens, 2002; Van Drie, Van Boxtel, Erkens, & Kanselaar, 2005). Understanding causal relations plays an important part in history learning, but these relations are also very complex: There are usually multiple causes for a single event, their importance or presence is often not immediately obvious, and in addition to their direct effect the causes also influence each other. A causal schema that visualizes these relations might help create a clearer overview for the learner (Barnes, 2002).

Given the importance of combining the 'Building-blocks' of historical knowledge (i.e. knowledge of historical phenomena, relations between those phenomena, and concepts describing phenomena and relations) and the affordances and limitations of concrete and abstract visualizations that 'fit' these types of knowledge, the idea occurs that a combination of concrete and abstract visualizations of these different elements (phenomena, concepts, and relations) can support the acquisition of historical knowledge. It is thought that historical phenomena can be better understood if one can form an image of it *and* relate it to other phenomena (Carretero, Jacott, Limón, López-Manjón, & León, 1994; Husbands, 1996; Leinhardt, 1993). Fasulo, Girardet, and Pontecorvo (1998) argue that a picture can only show a snapshot of a series of events. To make such an image meaningful, it needs to be framed by a temporal plot and its context of antecedents and consequences. For example, 'manorialism' in the Early Middle Ages can only be fully understood in the context of the fall of the Western Roman Empire. Students need to understand how each phenomenon is related to other phenomena, both chronologically (i.e. temporal relations) and in terms of cause and effect (i.e. causal relations; Masterman & Rogers, 2002). Although several studies have shed light on the effects of either schematic (abstract) or pictorial (concrete) visualizations on learning, little is known about the effects of combining such representations highlighting different aspects of the topic content, especially in learning assignments.

The mediating function of multimodal representations is determined – among others – by the nature of activities with the representations (Peeck, 1993). Theory seems to suggest that assembling and constructing multimodal representations – as opposed to simply receing them – more strongly encourages articulation of ideas and content, discussion, and deep processing (Cox, 1999). Most research on learning with visual or multimodal representations deals with representations that are given to learners. Often, the participants are instructed simply to 'study' the materials, and are not required to perform any other activity, such as sequencing those that represent the components of a process. These types of tasks can be viewed as different positions along a continuum, with representations merely presented to the learners at one end, and representations constructed by the learners themselves (e.g. having them draw) at the other end (Van Meter & Garner, 2005). Both mainstream theories, Dual Coding Theory (Paivio, 1991) and the Cognitive Theory of Multimedia Learning (Mayer, 2001), are based on research with presented representations. Somewhere between those two extremes on the continuum are representations where some parts are provided to the learners and some parts are constructed by the learners themselves – for example when learners have to add captions to pictures.

A study by Prangsma, Van Boxtel, and Kanselaar (2008) focused on differences between textual and multimodal tasks. These tasks were based on the idea that assembling and constructing visual or multimodal representations – as opposed to simply presenting them – would more strongly encourage deep processing through active involvement with the content. Moreover, this effect should be even stronger for collaborative group work, because there the visualizations also function as communicative support, encouraging discussion and articulation of ideas and content (Cox, 1999; Reimann, 2003). The visualizations that students built-in the study by Prangsma *et al.* (2008) contained both concrete pictures and abstract schemas at the same time. The pictures each visualized a single historical phenomenon (e.g. trade by barter) and the pictures and text had to be incorporated by the students in causal schemas and timelines. The study showed that integrating multimodal representations in a timeline led to significantly better learning outcomes than working on a textual task, but only in the short run. Since that study dealt with different combinations of abstract and concrete visualisations, one might suspect that differences in abstractness and concreteness of the visualisations might have played a role. However, this could not be distilled from the data gathered. Therefore, the study described here was designed to pull apart the three modes of representation – text, concrete pictures and abstract schemas – to try to find out what the effect was of each type separately as well as in combination. In addition, the tasks were reduced to more closely resemble the tasks used by Bodemer, Ploetzner, Bruchmüller, and Häcker (2005), who conducted an experimental study in which students had to relate textual and pictorial information about the working of a tire pump to each other by dragging and dropping captions on to elements of a drawing of a tire pump on a computer screen. This type of integration activity significantly improved learning when the learning material was more complex.

### *Problem definition*

This study was set up to research the differences in learning effects of different types of visualizations. It deals with differences between learning with textual tasks, with concrete (i.e. realistic) pictures added to the textual tasks, with text in an abstract causal map showing the relations between historical phenomena in the task, and with a combination of text and pictures in a causal map. The main question addressed by this research is: Does combining text and different types of visualizations – abstract and concrete – in history tasks enhance the acquisition of knowledge of a historical phenomenon, including the concepts and relations linked to it?

## Method

### *Participants*

The participants in this study ($N = 104$) were pupils from six different first-year classes in vocational middle school with three different teachers in two different schools (pupils aged 12 to 13). The majority of Dutch pupils in secondary school (approximately 60%) attend this type of school. About 24% of the pupils at this level have problems reading their school textbooks (Hacquebord, 2004).

### *Experimental tasks*

Working in pairs during one history lesson (approximately 45 min), participants carried out a task on the Early Middle Ages – 500 to 1000 AD in Western European history – and specifically on the effects of the fall of the Roman Empire. Each student pair was provided with a 328-word text. The text had a Flesch-Douma Readability Ease Score of 70, which is normal to fairly easy (Douma, 1960), and it contained 24 different concepts, including substantive concepts, such as *trade by barter, Viking, agriculture*, *administrator*, *lord*, and *serf*, and methodological concepts, such as *cause* and *change*. To make sure that all participants had read the text before starting on the assignment, the pairs were instructed to read the text out loud to each other. The text was the same for all four conditions, did not include visualizations, and was available during the task. After reading the text, the pairs were given a task in one of four conditions: Textual, Concrete, Abstract, and Combined. The task sheets with visualizations were designed according to the four principles for multimedia learning deemed appropriate for the materials used (Mayer, 2001). The tasks contained both words and pictures and/or schemas (multimedia principle), extraneous material was excluded (coherence principle), all elements of the task sheets were given to the participants simultaneously rather than successively (temporal contiguity), and depending on the condition text was organized in a causal schema and/or placed near a corresponding picture (spatial contiguity principle). Participants were asked to insert appropriate concepts, and thus finish sentences about events, phenomena, and relationships. In short, the focus of this study is on photorealistic drawings (i.e. concrete visualisations) of historical phenomena and concepts, and on causal maps (i.e. abstract visualizations) that represent relations between phenomena.

The tasks in all four conditions included the same 12 fill-in-the-blank sentences about the main issues in the text. The fill-in-the-blanks were informationally equivalent to the information in the text, but they were not identical to text sentences. The concepts to be filled in were present in the text, but appropriate synonyms were also allowed in some cases.

The tasks in the Textual condition contained just that: Text in the shape of the 12 fill-in-the-blank sentences. The Abstract condition contained the same fill-in-the-blank sentences, but in the form of a causal schema. The Concrete condition included eight pictures to illustrate the 12 fill-in-the-blank sentences. Reading from left to right, the sentences in this condition were in the same order as those in the text condition. The tasks in the Combined condition combined all three elements: the fill-in-the-blank sentences and pictures were integrated in the causal schema. In both conditions with pictures, each picture represented all or part of the concepts that had to be filled in by the students. Thus, students were required to combine the verbal texts with the presented visualizations. The black-and-white drawings used for the answer sheets were produced specifically for the experiments by a professional schoolbook illustrator and showed simplified representations of historical concepts, such as *Vikings*, *trade by barter*, and *agriculture*. The task sheets for all four conditions are shown in Figures. A1–A4 in the Appendix.

All participants worked in pairs for two reasons. First, this was done to encourage active processing through discussion (Erkens, Jaspers, Prangsma, & Kanselaar, 2005; Roschelle, 1992). Also, the tasks were based on the tasks in the previous study discussed in the Introduction, which dealt with dyads because there it enabled us to study the learning process through the student dialogue.

### *Preparatory assignment*

To give all participants the same starting-point for the experimental task in the study, and to give them the required background knowledge about the topic, a preparatory assignment was given. Participants – having just rounded off a chapter on Antiquity – were asked to draw pictures to illustrate a text on the Fall of the Roman Empire. The text was divided into three sections. The first section was about the situation in the second century: a large Roman empire divided in provinces, governed by an emperor and with a strong army defending the borders with fortresses and soldiers. The second section included information about the weakening of the empire (due to incompetent emperors, the division of the empire in an eastern and a western part, invasions and overthrowing of the last of the Roman emperors). The third and final section concluded with the situation as it existed in approximately the year 500 AD, when the Western Empire was gone and only the Eastern Empire continued. The participants were then asked to make three drawings – one for each section of the text – that together would give an accurate representation of the Fall of the Roman Empire. The task was piloted in two history classes.

### *Tests*

Participants completed the same individual test three times (pre-test, post-test, and retention test) which required them to indicate whether given statements were true or false. Each test consisted of the same 28 true–false items. Since, the instructional text was only 328 words in length, it was difficult to construct three parallel tests with a sufficient number of items each. Together, the 28 items covered the full range of phenomena, relations and concepts in the text and the assignment. To avoid test-effect as much as possible, the order of the questions was reversed for the post test. Some examples of test items are: ‘Viking raids were giving people trouble’ (true), ‘Roman administrators stayed to govern the territory’(false), ‘Most people lived from trade’(false), and ‘Almost everyone lived on agriculture’(true).

Reliability of the true–false test was determined using Cronbach’s alpha. Prior knowledge was low, so pupils had to resort to guessing on the pre-test, resulting in random answers, and thus in a Cronbach's alpha of .53. Cronbach's alpha for the post test, though not high (.68), was acceptable. Cronbach's alpha for the retention test was .73.

In addition to the true–false test, the post test was preceded by a free recall test, in which participants were asked to write down everything they could remember from the text and the task for the period 500 to 1000 AD. They were encouraged to write full sentences, and when giving loose concepts at least try to explain them.

The free recall test was analysed by counting the number of different historical concepts used and the number of correct propositions stated. First, all responses were divided into segments. A new segment was assumed where: (1) There was a signal word, e.g. *but, so, because, and, then,* and in some cases also for *when*, and *that*, and (2) either a new sentence was started (full stop), or the clause could function as an independent sentence (i.e. containing at least a verb and a subject). As a consequence, clauses with ellipsis were not segmented, so that for example, 'The farmers gave away part of their crop and received protection in return' was coded as a single segment. Informationally irrelevant segments such as 'I can't remember anything else' or 'We did two tasks and I worked with Jerry' were excluded from the analyses.

For the analysis of concepts, the concepts were underlined, and the number of different concepts was counted for each participant. The list used for concept coding was based on the text and task and contained a total of 24 different concepts. No distinction was made between correctly and incorrectly used concepts. Loose concepts (i.e. not part of a proposition – see the next paragraph) did count towards the number of concepts, although they were not counted as propositions.

For the analysis of propositions, the segments were coded as proposition or non-proposition. A proposition was defined as a statement about a historical phenomenon, relation or concept. Segments that only contained loose concepts (e.g. 'Romans') were considered non-propositions. Finally, the propositions were coded as either correct or incorrect. A correct proposition was defined as a statement about a historical phenomenon, relation or concept that is at least partly based on the contents of the text or task, and that does not conflict with historical reality as it is normally interpreted, for example, 'Serfs gave away part of their crop in exchange for protection', 'Germanic are a people'. An incorrect proposition was defined as a statement that conflicts with information in the text, that is historically incorrect (such as an anachronism), that is too general and not specific for the particular period and/or situation, or that refers to other periods than the one dealt with in the text and task, for example, 'Farmers looked for protection from a serf', 'Roads and bridges are for transportation'. Propositions were counted as incorrect when they dealt with the historical period preceding the one discussed in the text, but were correct historically. Propositions were counted as correct when they dealt with the historical period preceding the one discussed in the text, but were also mentioned in the text.

Inter-rater reliabilities were calculated for each step for a random selection of 15 post tests. The inter-rater reliability (Cohen's kappa) for concept coding was very good (.86), for distinguishing propositions it was good (.76), and for correctness of propositions it was also good (.77). Examples of propositions from the free recall test are shown in Table 1.

### *Evaluation questions*

We also included questions to evaluate, the participants' perception of the task difficulty and enjoyability and of their perception of their own achievement. Concrete

**Table 1.** Examples of pupil propositions from the free recall test

| Correct propositions | Incorrect propositions |
|---|---|
| Without the Romans it was not safe in the city | Western Europe was conquered first. [This happened before 500 AD.] |
| The Vikings started plundering | The Roman army fell apart. [This happened before 500 AD.] |
| Roads and bridges were not looked after anymore | People lived from trade. [Quite the opposite: people resorted to agriculture.] |
| Serfs gave part of their crop to the lord in exchange for protection | Serfs worked for the lord to get part of the crop |

visualizations such as the pictures used in our study may motivate students to focus attention on important parts of the text and the accompanying task. The pictures may also reduce the difficulty level of the task, because the pictures represent part of the concepts that have to be filled in by the students. Salomon and Leigh (1984), on the other hand, found that a representational format that is perceived as being easier results in lower mental effort, leading to less deep processing. The questionnaire was given to the participants after completion of the task. This questionnaire consisted of three questions each on a four-point scale: (1) Did you or didn't you enjoy the task? (very enjoyable – not enjoyable at all) (2) Did you find the task easy or difficult? (very easy – very difficult) and (3) Did you learn a lot or very little from the task? (a lot – nothing).

### *Setting and procedures*

The teachers assigned the pupils to three levels (low, intermediate, and high), based on their history grades on their report cards. These levels were used to distribute pupils over conditions within each class, and then to divide them into dyads with contiguous ability ranges (low + intermediate and intermediate + high) and intermediate dyads, because these combinations have been shown to result in an optimum balance between symmetry to enhance communication and relations and asymmetry to keep the dialogue going (Saleh, Lazonder, & De Jong, 2005). Low-intermediate and high-intermediate dyads were evenly distributed over the conditions within each class. Participants with missing data (e.g. due to missed tests) were not included in the final sample. Table 2 shows the final distribution of pupils from different teachers over conditions.

**Table 2.** Final sample size: Distribution of participants over conditions and teachers

| | Condition | | | | |
|---|---|---|---|---|---|
| | Textual | Abstract | Concrete | Combined | Total per teacher |
| Teacher A | 10 | 7 | 8 | 12 | 37 |
| Teacher B | 10 | 5 | 9 | 7 | 31 |
| Teacher C | 10 | 13 | 6 | 7 | 36 |
| Total per condition | 30 | 25 | 23 | 26 | 104 |

For each teacher, two classes were included in the sample.

The experiment began in the first lesson after completion of the regular lessons on Antiquity. To ensure that all participants had a similar starting-point for the period before 500 AD, the preparatory assignment on the fall of the Roman Empire was administered before the pre-test. The pre-tests were administered one to six days before the start of the experiment. The evaluation questionnaire was given to the participants directly after completion of the task. The post test was administered directly after the questionnaire: After the free recall test was collected, the true–false part of the post test test was handed out. The retention test was administered about 6 weeks after the experiment. The participants did not receive feedback on their tasks or tests during the entire period of the study. Between the post test and the retention test, regular classes were taught about the Early Middle Ages, such as the spread of Islam and Christianity, but not about the specific topic of the experiment.

### *Hypotheses*

Significantly higher post test scores were expected for the Abstract and Concrete conditions than for the Textual condition, because the concrete and abstract visualizations are expected to stimulate more elaborate encoding and can function as anchors for remembering the information. We expected the Combined condition to have significantly higher scores than the other three conditions, because this condition supports the formation of a clear image of historical phenomena, it makes causal relations more salient and it provides the most anchors. Furthermore, we expected that students in both conditions with concrete visualizations would perceive the task as easier and more enjoyable than students in the other conditions.

## Results

There were no significant differences between conditions on the pre-test score ($F(3,100) = 0.39$, $p = .76$, $\eta^2 = .01$). Since the data for the post test were not normally distributed, a Kruskall-Wallis test was used. The analysis showed that there were no significant differences between any of the conditions on the post test score ($\chi^2(3) = 2.48$, $p = .48$), nor between conditions on the retention test score ($F(3,100) = 0.10$, $p = .96$, $\eta^2 = .00$). Additional analyses showed that the scores increased between the pre-test and post test for all conditions, as well as between the pre-test and the retention test, meaning that performance of all conditions improved. The difference between the post test and retention test scores did not show a significant decline in learning results for any of the conditions. The descriptive results are shown in Table 3.

**Table 3.** Means and standard deviations for the pre-test, post-test and retention test scores

| | | Pre-test | | Post-test | | Retention test | |
|---|---|---|---|---|---|---|---|
| Condition | *N* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Textual | 30 | 17.50 | 3.45 | 22.73 | 3.35 | 23.97 | 3.18 |
| Abstract | 25 | 16.96 | 3.59 | 23.72 | 3.71 | 23.48 | 3.27 |
| Concrete | 23 | 16.57 | 3.91 | 22.78 | 3.00 | 23.74 | 3.53 |
| Combined | 26 | 16.58 | 3.94 | 23.88 | 3.34 | 23.65 | 3.21 |
| Total | 104 | 16.93 | 3.68 | 22.03 | 3.35 | 23.72 | 3.24 |

The free recall tests were analysed for use of historical concepts and propositions. An ANOVA showed that the four conditions did not differ significantly in the number of different concepts ($F(3,100) = 1.56$, $p = .20$, $\eta^2 = .04$), nor in the number of correct propositions ($F(3,100) = 2.29$, $p = .08$, $\eta^2 = .06$). Table 4 shows the mean number of different concepts and the mean number of correct propositions for each condition.

**Table 4.** Means and standard deviations for the number of different concepts and for the number of correct propositions in the free recall test

| | | Concepts | | Correct propositions | |
|---|---|---|---|---|---|
| Condition | *N* | *M* | *SD* | *M* | *SD* |
| Textual | 30 | 6.67 | 2.89 | 5.10 | 3.39 |
| Abstract | 25 | 6.96 | 3.19 | 4.32 | 2.91 |
| Concrete | 23 | 5.35 | 2.39 | 3.74 | 2.34 |
| Combined | 26 | 6.65 | 2.64 | 5.85 | 3.21 |

A Kolmogorov-Smirnov test showed that the data of the evaluation questions were not distributed normally. Table 5 shows the results for the questions for each condition. On average, the participants in all four conditions had a neutral opinion on the enjoyability of the task ($M = 2.11$, $SD = 0.52$), as a Kruskall-Wallis test showed no significant differences ($\chi^2(3) = 5.75$, $p = .12$). However, the conditions did differ in their evaluation of the difficulty level ($\chi^2(3) = 14.22$, $p = .00$) and their estimation of how much they had learnt ($\chi^2(3) = 8.19$, $p = .04$). A series of Mann-Whitney tests with Bonferroni correction showed that the condition with concrete visualisations rated their task as significantly easier than the Textual condition ($U = 187.00$, $p = .002$) and the condition with abstract visualizations ($U = 148.50$, $p = .001$). At the same time, when judging how much they had learnt, the Concrete condition thought they had learnt more than the Textual condition ($U = 214.00$, $p = .007$) thought they had.

**Table 5.** Means and standard deviations for the three questions in the evaluation questionnaire

| | Enjoyability | | Difficulty | | Learning gains | |
|---|---|---|---|---|---|---|
| Condition | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Textual | 2.23 | 0.43 | 2.00 + | 0.53 | 2.37 + | 0.67 |
| Abstract | 2.08 | 0.57 | 2.00 + | 0.41 | 2.16 | 0.55 |
| Concrete | 1.91 | 0.53 | 1.50 – | 0.51 | 1.86 – | 0.56 |
| Combined | 2.16 | 0.55 | 1.77 | 0.59 | 2.27 | 0.60 |

Variables with a + and – sign in the same column differ significantly at $p < .05$.
Enjoyability: 1=very enjoyable, 4=not enjoyable at all. Difficulty: 1=very easy, 4=very difficult. Learning gains: 1 = a lot, 4=nothing.

## Discussion

Returning to the hypothesis, the results lead to the conclusion that tasks requiring the combination of text and different types of visualizations – abstract and/or concrete ones – do not necessarily enhance history learning more than textual tasks alone.

More specifically, the abstractness or concreteness of the visualizations did not seem to play a role. A number of factors may have influenced this outcome such as the lack of difference between conditions, the complexity of the task, the students' lack of experience with visual tasks, the setting, the nature of the domain, or its semiotics.

First, the predicted effects of the use of visualizations may have failed to materialize as a result of the similarity between the materials used in the four different conditions in the study. All four conditions received the same carefully written one-page text. It is possible that reading this short text, in itself, was enough to be able to perform well on the tests, and that this reduced the added value of the tasks – even though the concrete pictures and abstract schemas were designed to support the learning process. While the visualizations in our study were carefully designed to suit the content, it is possible that the visualizations were simply not needed.

Second, it may have been the case that the task content used in our research was not complex enough to elicit the expected results. Bodemer, Ploetzner, Bruchmüller, and Häcker (2005) found that integration of multimodal representations was more effective only for learning from complex information. Perhaps the text that was given effectively explained the complexity of the information, or the fill-in-the-blanks task provided enough opportunity for actively processing the information without using the visualizations. If we had chosen different content requiring different visualizations – in particular abstract ones such as maps or timelines – this might have given quite different results. This also raises the question as to whether the abstract/concrete distinction is really a useful one for research on learning through integrating textual and visual information.

A third factor that may have affected the outcomes of our study is the extent to which students actually process the visualisations and/or integrate the textual and visual information. Participants in the Abstract, Concrete and Combined conditions were not explicitly stimulated to use the schema and/or pictures provided or to talk about them. It is possible that pupils are fixed in their habits, in their approach to doing history and dealing with history tasks. In other words, maybe 'old habits die hard'. Pupils may typically focus on the textual information, and not use the visual information that is presented to their full advantage unless explicitly told to do so or without being explicitly told how to do this. Such explicit instruction and practice with using visual information was not included in the research design. Compared with the tasks used in this study, the tasks used in our previous study required learners to be more active with the visualisations (e.g. to sequence the pictures) and to more actively integrate the textual and visual information (i.e. to write complete captions and not simply fill-in-the-blanks). This may explain why learners in our previous study who completed tasks with visualizations outperformed learners in the textual condition, whereas we did not find such a difference in this study. Several researchers suggest that simply providing visual representations is not sufficient and that active integration or processing is needed (e.g. Ainsworth, 1999; Scevak & Moore, 1998). On the other hand, in studies in other domains simply adding pictures to text without any constructive activity to process the information or relate text and pictures was beneficial to learning (Peeck, 1993).

Fourth, the setting of our study may have played an important role. The research was done in a classroom setting, where the tasks were incorporated in the normal history curriculum. This ecologically valid setting may have influenced the results through interfering circumstances, such as classroom dynamics and attention span differences. The experiments that led to Mayer's Cognitive Theory of Multimedia Learning (2001), on the contrary, were carried out in lab-like situations, in individual settings, and with a very different type of participant – mainly psychology undergraduate 'volunteers' instead of 12 to 14-year-olds in prevocational education – the lower levels of secondary education. While this study was not intended as a replication of Mayer's work it does raise the possibility that there may be some problems with the ecological validity of some of the assumptions underlying Mayer's theory.

Further explanations for the findings can be placed under two themes: the nature of the domain, and the semiotics of the visualisations. Both were different in this study than in most studies that did confirm the superiority of multimodal tasks. First, the nature of the subject domain might offer a possible explanation for the findings in this study. Previous studies on learning with visualizations by other researchers were mainly done in the domain of science. The nature of phenomena dealt with in science is often very different from the nature of historical phenomena (Voss & Wiley, 2006). Science is a well-structured domain, and its phenomena can often be understood in terms of 'how things work' – such as how lightning occurs or how a pump works: these are processes that can be shown in quite straightforward representations. History, on the other hand, is ill-structured and highly interpretative in nature (Wineburg, 1999). Historical concepts are often ill-defined, with no generally agreed upon definitions and with different meanings when applied to different historical situations (Limón, 2002). Historical phenomena – such as manorialism and serfdom – can be understood in terms of human behaviour in specific contexts, and are not easily unequivocally represented in concrete or even combined visualizations. In history learning, just a picture or schema with no or very little text may serve as an anchor, but perhaps it does not serve as a suitable replacement for just textual information.

Linked to this is the idea that the grammar – or semiotics – of visualisations in the domain of history has no unambiguous principles or rules as there are in science. The visual grammar of a domain like science, even though it deals with complex phenomena, is more standardized – e.g. open and closed pipelines are represented by a solid versus a dashed line. Historical phenomena, relations and concepts – such as war, democracy or justice – on the other hand, cannot be easily described with standard visualizations. De Westelinck *et al.* (2005) found limitations of Mayer's theory of multimedia learning (Mayer, 2001) in a study within the domain of educational psychology in longer learning periods and with more abstract and semiotically ambiguous concepts, as the study discussed here did for history. Just as in educational psychology, in history even seemingly simple symbols like arrows can be interpreted in different ways, for example as indicating dynamic relations such as causality, or merely as temporal relations, or as static relations describing the structure of phenomena (O'Donnell *et al.*, 2002). In addition, even the basic governing principles such as cause and effect can be quite unpredictable and cannot be captureds in general laws (Voss & Wiley, 2006). Simply adding arrows is not enough for describing causal relations, as a single event can trigger an avalanche of different possible consequences. Even when the context is known, for example the Early Middle Ages, pictures are also often open to multiple interpretations. This also makes it hard for students to learn to understand the form of representation – the importance of which is underlined by Ainsworth (2006): There are no fixed formats in history, for example for visualizing time, war or manorialism. Perhaps, the participants in this study did not use the visualizations simply because they did not have the skills to interpret them.

A bright spot in the results deals with the evaluation of the visualizations. Students in the pictorial visualization condition ('Concrete') rated the materials as easier and felt that they had learnt more from the task than the students in the textual condition rated their materials. Such positive appreciation of the materials should not be underestimated. The goal of educational innovation is not only to make learning more *efficient* so that learners learn the same amount of material in a shorter time-span, and/or make learning more *effective* so that learners learn better, but also to make learning more *pleasant* (for example by being easier or giving the feeling of learning more) such that the affective learning experience is more satisfying and learners will want to learn (Kirschner, 2004). Educational research on learning with visualizations tends to focus only on determining how specific tools, environments, or student characteristics affect either the effectiveness and/or efficiency of learning, and not on motivational or emotional aspects of representation types. Although concrete visualizations were not found to have an effect on performance in the research reported here, they still seemed to have some positive consequences for the way students appreciated the tasks. Future research should therefore pay attention to affective aspects of learning with different types of visualizations.

In this discussion, we have considered several possible explanations to clarify the absence of differences between conditions. Those explanations lead us in very different directions, and this shows that the causes are complex and varied. Future research should focus on gaining insight into these different aspects of the use and interpretation of different types of visualizations in the humanities and social sciences, for example in domains such as history or geography. In particular, more qualitative studies should shed light on the way pupils at different educational levels understand and interpret different types of visualizations of historical phenomena and relations and the specific conditions under which such visualizations can enhance history learning.

## Acknowledgements

This research was supported by a grant from The Netherlands Organization for Scientific Research (411-01-004). Wim Euverman created the artwork used in the learning tasks. The preparatory assignment was co-developed by Marloes van Rijn as part of her Master's thesis.

## 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*, 183–198.

Barnes, S. (2002). Revealing the big picture: Patterns, shapes and images at key stage 3. *Teaching History*, (107), 6–12.

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

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*, 182–197.

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 instruction: Vol. 3. European research in an international context* (pp. 27–48). Oxford: Pergamon.

Carretero, M., Jacott, L., Limón, M., López-Manjón, A., & León, J. A. (1994). Historical knowledge: Cognitive and instructional implications. In M. Carretero & J. F. Voss (Eds.), *Cognitive and* *instructional processes in history and the social sciences* (pp. 357–376). Hillsdale, NJ: Lawrence Erlbaum.

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

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*, 555–573.

Douma, W. H. (1960). *De leesbaarheid van landbouwbladen: Een onderzoek naar en een toepassing van leesbaarheidsformules*. [Readability of Dutch farm papers: A discussion and application of readability-formulas]. Wageningen, The Netherlands: Landbouwhogeschool.

Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior*, *21*, 463–486.

Fasulo, A., Girardet, H., & Pontecorvo, C. (1998). Seeing the past: Learning history through group discussion of iconographic sources. In J. F. Voss & M. Carretero (Eds.), *International review of history education: Vol. 2. Learning and reasoning in history* (pp. 132–153). London: Woburn Press.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction*, *12*, 213–232.

Hacquebord, H., Linthorst, R., & Stellingwerf, B. (2004). *Voortgezet taalvaardig: Een onderzoek naar tekstbegrip en woordkennis en naar de taalproblemen en taalbehoeften van brugklasleerlingen in het voortgezet onderwijs in het schooljaar 2002–2003*. [Advanced language proficiency: A study into text comprehension and vocabulary and into the language problems of first-year pupils in secondary education during the school year 2002–2003] Groningen, The Netherlands: Expertisecentrum Taal, Onderwijs en Communicatie (Etoc).

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

Kirschner, P. A. (2004). Design, development, and implementation of electronic learning environments for collaborative learning. *Educational Technology Research and Development*, *52*, 39.

Leinhardt, G. (1993). Weaving instructional explanations in history. *British Journal of Educational Psychology*, *63*, 46–74.

Limón, M. (2002). Conceptual change in history. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 259–289). Dordrecht, The Netherlands: Springer.

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

Masterman, E., & Rogers, Y. (2002). A framework for designing interactive multimedia to scaffold young children's understanding of historical chronology. *Instructional Science*, *30*, 221–241.

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

O'Donnell, A. M., Dansereau, D. F., & Hall, R. H. (2002). Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review*, *14*, 71–86.

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*, 227–238.

Prangsma, M. E., Van Boxtel, C. A. M., & Kanselaar, G. (2008). Developing a 'big picture': Effects of collaborative construction of multimodal representations in history. *Instructional Science*, *36*, 117–136.

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

Robinson, D. H., Robinson, S. L., & Katayama, A. D. (1999). When words are represented in memory like pictures: Evidence for spatial encoding of study materials. *Contemporary Educational Psychology*, *24*, 38–54.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*, *2*, 235–276.

Saleh, M. A., Lazonder, A., & De Jong, T. (2005). Effects of within-class ability grouping on social interaction, achievement, and motivation. *Instructional Science*, *33*, 105–119.

Salomon, G., & Leigh, T. (1984). Predispositions about learning from print and television. *Journal of Communication*, *34*, 119–135.

Sauer, M. (2000). *Bilder im Geschichtsunterricht. Typen, Interpretationsmethoden, Unterrichtsverfahren* (2nd ed.). [Images in history teaching: Types, interpretation methods and didactics] Seelze, Germany: Kallmeyersche Verlagsbuchhandlung.

Scevak, J. J., & Moore, P. J. (1998). Levels of processing effects on learning from texts with maps. *Educational Psychology*, *18*, 133–155.

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

Van Boxtel, C. A. M., Van der Linden, J., Roelofs, E., & Erkens, G. (2002). Collaborative concept mapping: Provoking and supporting meaningful discourse. *Theory into Practice*, *41*, 40–46.

Van Drie, J., Van Boxtel, C., Erkens, G., & Kanselaar, G. (2005). Using representational tools to support historical reasoning in computer-supported collaborative learning. *Technology, Pedagogy and Education*, *14*, 25–42.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review, 17*, 285–325.

Voss, J. F., & Wiley, J. (2006). Expertise in history. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 1746–2424). Cambridge, MA: Cambridge University Press.

Wineburg, S. (1999). Historical thinking and other unnatural acts. *Phi Delta Kappan*, *80*, 488–499.

*Received 19 December 2006; revised version received 23 September 2008*

## Appendix

Task sheets for all four conditions (Figs. A1–A4)

**500 to 1000 AD - The Western Roman Empire disappears: What changed?**

There are no more ___________ who organise everything.

There are no more ___________ to protect people.

Upkeep on ___________ stopped.

Therefore ___________ became more difficult.

After 500 AD most people lived from ___________

And they made everything they needed themselves.

There was also ___________

It became ___________

___________ traveled around looting.

Farmers looked for protection with a ___________

He employed ___________

The farmers gave ___________ to the lord in exchange for ___________

These farmers are called ___________

**Figure A1.** Task sheet for the textual condition.

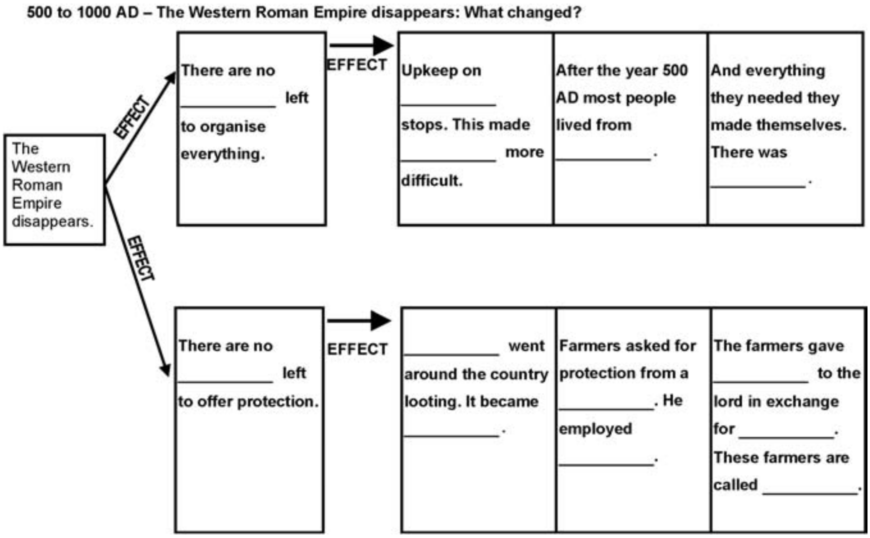


**Figure A2.** Task sheet for the abstract condition.

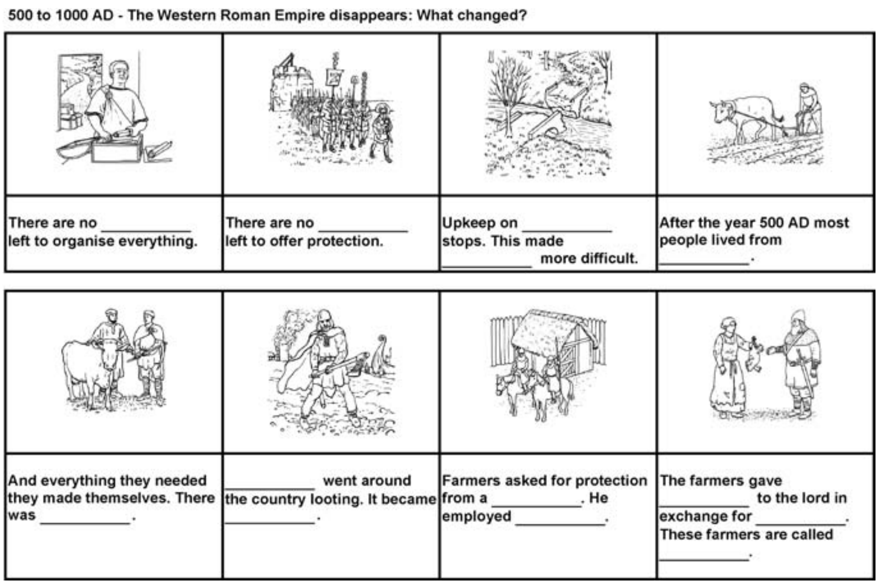


**Figure A3.** Task sheet for the concrete condition.

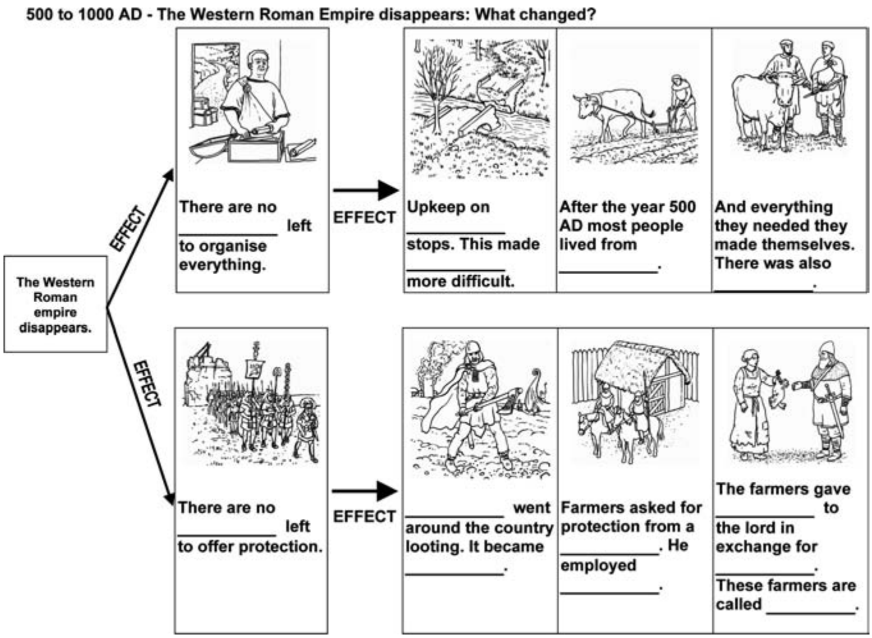


**Figure A4.** Task sheet for the combined condition.