Visuospatial working memory in learning from multimedia systems

V. Gyselinck, M.-F. Ehrlich, C. Comoldi, R. de Beni & V. Dubois
Laboratoire de Psychologie Expérimentale, Université René Descartes Boulogne France et Dipartimento di Psicologia, Università di Padova, Italia

Abstract Multimedia systems involve the association of various types of information: verbal information presented visually or auditorily, static or dynamic pictorial information, and sound information. In a cognitive approach, integrating this information involves complex processes constrained by properties of the learner’s cognitive system, and especially by the capacity of working memory. This paper, reports on an experiment focused on the integration of verbal and pictorial information when students learn a series of physics concepts. The involvement of the visuospatial working memory was investigated by means of a dual-task paradigm. Results show that pictorial information enhances the learning process. They also suggest that the visual and the spatial components of visuospatial working memory should be considered separately. Finally, they emphasise the need to consider the limitations in cognitive resources available to the learner.

Keywords: Control group; Multimedia; Physics; Undergraduate

Introduction
Multimedia systems are developing rapidly and will continue to do so especially in instructional fields. Due to these systems, the instructional process may be made more flexible, rich and individualised. However, in order to be successful, the systems must be adapted to, and take into account, the cognitive limitations of the learner (Reinking & Bridwell-Bowles, 1996). People involved in the creation and use of teaching material must consider a series of related questions. For example: to what extent is the learner able to treat different types of information simultaneously? Which rules guide selecting the amount and nature of simultaneously presented information? What are the sources of individual differences in learning ability from multimedia systems?

Multimedia systems typically involve the association of various types of information: verbal information (words, sentences or short texts) presented visually or auditorily, pictorial information (illustrations, photographs, schemas) presented visually in a static or dynamic way, and sound information. Because of the limits of
the computer screen, the amount of information presented simultaneously is restricted. In the case of the description of complex phenomena, the presentation requires several successive screens. It follows that learners have to first integrate various types of information presented simultaneously on the same screen, and second to maintain this integrated information activated in memory while processing incoming information. In a cognitive approach, integrating this information involves complex processes constrained by properties of the learner’s cognitive system, and especially by the limited capacity of working memory.

This paper reports on an experiment focused on the integration of verbal and pictorial information, both presented visually. More specifically, the investigation looked at the comprehension and memory processes of students learning a series of physics concepts (static electricity, gas pressure, etc.) through the computer assisted presentation of a text associated with illustrations. It was hypothesised that the presence of illustrations facilitated learning and that the integration of text and illustrations involves the visuospatial working memory.

**Association of verbal and pictorial information**

A large body of literature has investigated the effects of illustrations on text processing (for a review, see Gyselinck & Tardieu, 1999). However, results from different investigations are sometimes contradictory. Furthermore, the case of illustrations associated with text presented via computer does not appear to have been considered sufficiently. Beyond the empirical evidence, several authors (Glenberg & Langston, 1992; Hegarty & Just, 1993; Kruley et al. 1994; Gyselinck, 1995; Gyselinck & Tardieu, 1999) have proposed to interpret the facilitative effect of illustrations in the framework of Johnson-Laird’s (1983) theory. According to this theory, the reader constructs an internal representation, called a mental model, that has a structure analogous to that of the situation described in the text. Illustrations depicting the content of the text they accompany should facilitate the construction of a mental model. Such a view is supported by empirical findings showing that illustrations are particularly useful when underlining the relationships between elements described in the text.

Consequently, an important question concerns the properties of pictorial information and its relationship to the text. Being interested in scientific materials, Hegarty et al. (1996) proposed to distinguish three categories of pictorial information: **iconic diagrams** (e.g. the drawing of a cross-section of the eye) which depict the structural features of a concrete system, **schematic diagrams** (e.g. an electrical circuit) which use graphic conventions to depict the components of abstract concepts and their organisation, and **graphs** which depict some quantitative facts. These three categories are not mutually exclusive, their boundaries can be difficult to define and their functions may depend on the properties of the subject matter. However, such a tentative classification helps to conceptualise the integration processes of the pictorial and verbal information. In addition, these two types of information can be made redundant by providing similar information by means of two different media or they can complement each other by providing different information about the same subject.

It appears that the processes involved in the integration of textual and pictorial information are very complex. Some data suggest that these processes, whose by-product is a mental model of the subject matter, are dependent on the learner’s
knowledge stored in long-term memory (Mayer & Gallini, 1990). The hypothesis behind the research described here is that they are also constrained by the capacity of the learner’s working memory and more particularly that of the visuospatial working memory.

**Working memory capacity**

Working memory has been defined by Baddeley (1986) as a system of limited capacity that ensures a double function of dealing with, and temporarily holding information. This form of active memory intervenes in all complex cognitive activities: language comprehension and production, new knowledge acquisition, reasoning and problem-solving, etc. (see Cornoldi & McDaniel, 1990; Ehrlich & Delafoy, 1990; Vecchi et al. 1995). In Baddeley’s model, the working memory comprises three parts: a central executive that coordinates two slave systems, the articulatory loop and the visuospatial sketchpad. The articulatory loop has the role of maintaining the phonological entries active under the control of an articulated process. The visuospatial sketchpad maintains spatial and visual information, thus ensuring the formation and manipulation of mental images.

Quite a number of studies have clearly highlighted the role of the central executive in text comprehension. The capacity of this central system, measured with the help of a particular test (reading span), is an important factor in the high level psycholinguistic operations underlying comprehension (Just & Carpenter, 1992). Nevertheless, as Gathercole & Baddeley (1993) have underlined in their book, although there have been a few studies (e.g. Pazzaglia & Cornoldi, 1999), the role of the articulatory loop and the visuospatial sketchpad in comprehension remains an open question. The present hypothesis is that the integration of verbal and pictorial information calls upon the three components of the working memory, and in particular the visuospatial sketchpad.

Kruley and colleagues (1994) reported some data supporting the involvement of the visuospatial sketchpad in the integration of texts and illustrations. They used a dual-task paradigm widely used in the studies aimed at demonstrating the functioning of a separate visuospatial sketchpad (Logie, 1995). Texts were presented auditorily in two conditions: text only or text accompanied by one picture (iconic diagram) that displayed the structural relationships between the parts of the objects described in the texts (e.g. the volcano). If the processing of the picture involves the visuospatial sketchpad it was expected that a visuospatial concurrent task would selectively impair performance when the text was accompanied by the picture. Results from several experiments were interesting but not conclusive. The expected effects were observed on concurrent task performance but not on comprehension. Furthermore the range of this study was limited by its methodological features, in particular the oral presentation of the texts accompanied by a single illustration.

In the experiment reported in this paper, the Kruley et al. (1994) rationale was used with materials describing basic physics phenomena presented on a computer screen. Each text contained nine sentences presented successively. Two formats of presentation were compared: sentence only or sentence accompanied by an illustration. The first goal was to extend evidence that illustrations facilitated the computerised learning conditions. The second goal was to study the involvement of the visuospatial sketchpad in the integration of texts and illustrations by means of a dual-task paradigm. Three concurrent tasks were used. While reading, students had
to perform either a visuospatial concurrent task, or a verbal concurrent task, or a control concurrent task. The comprehension of the phenomena described was tested by means of factual and inferential questions. A negative effect of the visuospatial and also the verbal concurrent tasks compared to the control task was expected in both formats of presentation. If the integration of verbal and pictorial information involves the visuospatial sketchpad, then this integration should be disrupted by the concurrent visuospatial task. Consequently, the negative effect of the visuospatial task should be stronger on the text-plus-illustrations format, decreasing the beneficial effect of illustrations. Such a decrease should not be observed with a concurrent verbal task.

Method.

Students
Forty-eight undergraduates participated voluntarily in exchange for course credit. An additional 12 students were excluded because of their high prior knowledge in Physics.

Material
Six texts dealing with basic notions of physics were used (electrolysis, gas pressure, gas volume, static electricity, osmosis, and state transitions of pure compounds). All texts conformed to the same structure of a title and nine sentences. For each sentence, a colour illustration was designed representing the elements mentioned in the sentence and the causal and temporal relationships between them. Each text was followed by six questions which tested the comprehension of the notion presented. There were two kinds of question: three questions tested factual information, explicitly given in the text, and the other three tested students’ ability to draw elaborate inferences from several sentences in the text. Each question required the selection of the correct response out of three possibilities. A further text, dealing with the functions of the heart was prepared for the practice phase. It was six sentences long and was prepared with associated illustrations and two questions.

To evaluate students’ prior knowledge in physics, a cloze-text, 25 sentences long was prepared. It was entitled ‘The Structure of matter’ and dealt with general notions in physics which were not directly dealt in the experimental texts. Twenty words were missing.

Procedure and design
Participants were tested individually and the time of the session was about 45 minutes. In the first phase, they were given a series of tests the first of which was the Corsi-blocks test measuring their visuospatial span (Milner, 1971; Orsini et al., 1987). In this test, the experimenter successively points to an increasing sequence of blocks arranged on a board. After each presentation, the student has to reproduce the sequence in the correct order. The value of his/her span corresponds to the length of the highest correct sequence he/she was able to reproduce. Students were also given a digit span test measuring their phonological short-term memory: in a similar way to the Corsi-blocks test, the experimenter gives an increasing series of digits orally, and the student has to repeat the sequence in the correct order. Students were then given a test measuring subjective vividness of mental images (the VVIQ test, Marks, 1973). Finally, they were presented with the cloze-text to measure their level of
knowledge in physics and they had to write the missing words in the empty places.

Two groups of participants, whose mean scores in the previous three tests were equivalent, were formed for the second phase of tests. They were: group T (the text-only group) who were presented only with the texts, and group T + I (text-plus-illustrations) who were presented with texts accompanied by illustrations. In the second phase of the experiment, all the students in each group performed the learning task in the three concurrent tasks conditions. The order of the conditions was counterbalanced across students, thus forming six subgroups in each format of presentation. Students first practised the learning task and the concurrent tasks with the practice text.

The students of the T group were presented with the text on a computer screen. The order of the six texts was randomly assigned to each student. The participant was instructed to read the text in order to understand it and to be able to answer various questions about it. Each sentence was exposed for 10 seconds and was then followed by the subsequent sentence. At the end of each text, the questions appeared one at a time, in a new random order for each text and each student. The participant had to read the question and then press a key in order to read the three alternative responses. The three alternatives were presented altogether on the screen, and the student had to press the key corresponding to the chosen alternative. The position of the right answer was randomised across the six questions. The student was instructed to answer as quickly and correctly as possible. Reading of the texts was associated with one of three conditions: visuospatial concurrent condition, verbal concurrent condition and control condition.

In the visuospatial concurrent condition, derived from Kruley et al. (1994), the students had to remember a configuration of dots within a rectangular grid (4 cells x 4 cells) while reading the successive sentences of the text. First, an initial dot display was shown for 667 ms; then the display disappeared and students had to read and to understand a series (1, 2 or 3) of sentences; then, a test display was presented until the student decided whether the display was the same or different from the one memorised. A new display to be memorised was then presented after the student had given his/her response by pressing one of two keys. Half of the test displays presented were different from those previously presented. This sequence of events was repeated four times for each of two successive texts.

In the verbal concurrent condition, the procedure was similar, except that students were presented with a series of three ‘non-words’ which they had to pronounce and to memorise. For the tests (half same, half different), only one ‘non-word’ was presented, until the student had decided, on the basis of its pronunciation, whether it pertained or not to the series memorised.

In the control condition, reading was interrupted four times as in the two other conditions. A 4×4 grid was shown to students with three ‘non-words’ inside. Immediately after the presentation, a new grid appeared on the screen and the student had to decide whether it was identical or not (on the basis of the ‘non-words’ used and on their location in the grid) to the one just presented. Thus, contrary to the visuospatial or the verbal condition, students did not have to remember the grid during reading.  

The difficulty of the concurrent tasks had been tested in a pre-experiment conducted with 18 students. The concurrent tasks were combined with a primary task assumed to impose no specific demands neither on the visuospatial sketchpad, nor on the phonological loop, but that required attention and could last the
In the T + I format, the procedure was the same, but the illustration was presented in the upper part of the screen whereas the sentence appeared in the lower part.

**Results**

*Comprehension data*

Figure 1 shows the mean percent correct on the multiple choice questions in each condition. An analysis of variance was conducted with format of presentation (text-only vs. text-plus-illustrations) as a between-students factor, conditions (control vs. visuospatial vs. verbal concurrent tasks) and type of questions (factual vs. inferential) as within-students factors. The main effects of format of presentation and of type of questions were significant (respectively $F_{1,46} = 6.23, p < 0.025$ and $F_{1,46} = 140.07, p < 0.0001$). The interaction between these two variables was also significant ($F_{1,46} = 6.92, p < 0.025$). Planned comparisons showed that the beneficial effect of pictures was significant for inferences ($F_{1,46} = 9.21, p < 0.005$) but not for factual questions ($F_{1,46} = 1.25$). However, no effect of conditions was observed, and no interaction with the format of presentation. The three way interaction between condition, format and type of question was not significant.

![Figure 1. Mean percentages of correct responses on comprehension questions as a function of condition, for the text-only and the text-plus-illustrations formats.](image)

In order to test the hypothesis about the role of the visuospatial condition, it was necessary to carry out a separate analysis taking only the visuospatial and the control conditions into consideration. In this analysis, however, no effect involving the condition was significant. Thus, the visuospatial concurrent condition did not impair comprehension performance, and contrary to expectation, the beneficial effect of illustrations was not decreased by the concurrent visuospatial task.

Finally, when the verbal concurrent condition and the control condition were same time as reading of the texts. The task chosen was a task of comparison of series of tempi, and the procedure used was the same as in the present experiment. Analyses revealed that for a same level of performance in the comparison task, accuracy on the concurrent tasks was equivalent ($m = 86.1\%$ for the visuospatial task, $m = 86.8\%$ for the verbal task, and $m = 86.1\%$ for the control task) which suggested that these three tasks had the same level of difficulty.
considered, no effect involving the condition was significant.

Thus, results showed that comprehension performance was facilitated by the presentation of illustrations but it was not impaired by either of the concurrent conditions.

Concurrent tasks data
Mean percent correct on the concurrent tasks is shown on Fig. 2. An analysis of variance was performed with format of presentation as a between-students variable and condition as a within-students variable. The main effect of conditions was not significant, nor was the main effect of format or the interaction between these two variables. When the visuospatial condition and the control condition were considered, the main effect of format was not significant, nor was the effect of condition and the interaction. When the verbal condition and the control condition were considered, the main effect of condition was significant ($F_{1,46} = 4.52, p < 0.05$): as is shown on Fig. 2, the performance on the verbal concurrent task was lower than performance on the control task. However, the main effect of format of presentation was not significant, nor the interaction between the two variables.

Complementary analysis: the role of the capacity of the visuospatial sketchpad
Recently, Pazzaglia & Cornoldi (1999) showed that the capacity of the visuospatial sketchpad affected the memorisation of a descriptive text. Students with a high spatial span, measured by means of the Corsi-blocks test, recalled the description of a city better than low spatial span students, both subgroups of students being equivalent on a digit span test. Thus, in the present study, we decided to explore the role of the spatial span in the integration of verbal and pictorial information. The scores on the Corsi-blocks test were used to contrast high and low spatial span students. Two subgroups were formed in the T group and the T + I group, contrasted on the spatial span, but with the same mean scores in digit span. In the T group, 11 students formed the high spatial span subgroup, and 13 students formed the low spatial span subgroup. The mean Corsi-blocks scores were $m = 6.36$ and $m = 4.38$, respectively, a difference which was statistically different ($t(22) = 8.21, p < 0.0001$) whereas the mean scores on the digit test were equivalent ($m = 7.63$ and $m = 7.77$, $m = 7.63$ and $m = 7.77$.)
t(22) = 0.23). In the T + I group, 13 students formed the high spatial span subgroup and 11 students formed the low spatial span subgroup. The mean scores on the Corsi-blocks test were, respectively, $m = 6.38$ and $m = 4.54$ and this difference was significant ($t(22) = 8.74$, $p < 0.0001$), whereas subgroups were equivalent on the digit span test ($m = 7.31$ and $m = 8.00$; $t(22) = 1.38$).

Comprehension data were submitted to a new analysis of variance, with the same factors as in the previous analysis, plus the spatial span factor as a between students factor. Given the small number of students in each subgroup, results should be considered cautiously. However, they provide interesting findings. The main effect of spatial span was not significant, but the three way interaction between this factor, format of presentation, and type of questions was significant ($F_{1,44} = 10.43$, $p < 0.005$). As is shown on Fig. 3, different patterns of results are observed as a function of spatial span. Planned comparisons showed that for the high spatial span students, the format of presentation was marginally significant ($F_{1,22} = 4.01$; $p = 0.05$) and the interaction between format of presentation and type of questions was strongly significant ($F_{1,22} = 17.77$, $p < 0.0005$). For the low spatial span students, neither the format of presentation, nor the format x type of questions interaction was significant. High spatial span students benefitted widely from the illustrations associated to texts for answering inferences, whereas low spatial span students did not show any beneficial effect. Figure 3 also shows that the beneficial effect of illustrations tends to be stronger in the visuospatial concurrent condition. However, no interaction involving the spatial span factor and the concurrent tasks condition factor was significant.

**Discussion**

The first goal of the present experiment was to extend evidence concerning the facilitative effects of illustrations in computer supported learning. The results clearly showed that the comprehension of texts presented on a computer is improved by the presentation of illustrations. In addition, this beneficial effect was more important for
inferential questions than for factual questions, which suggests a deeper understanding.

The second goal was to study the involvement of the visuospatial sketchpad in the integration of texts and illustrations by means of a dual-task paradigm. Unexpectedly, the visuospatial concurrent task did not produce any negative effect either on the comprehension performance or on the concurrent task performance. Compared to the control condition, it did not decrease the beneficial effect of illustrations on comprehension. On the contrary, the mean values tended to show (although not significantly) a higher effect of illustrations in the visuospatial concurrent task condition. The verbal concurrent task produced a negative effect only on the concurrent task performance. This effect was similar in both formats of presentation. This last result is not surprising, given that both formats involved verbal processing. Again, the positive effect of illustrations seemed higher in the verbal concurrent condition. In addition, a complementary analysis indicated that some results were dependent on the capacity of the visuospatial sketchpad. Students were contrasted into high and low spatial span subgroups. Examination of the results showed that only the high span students benefited from the presentation of illustrations associated to texts for answering inferences. This benefit was maintained and tended to be stronger when students performed the visuospatial concurrent task.

Thus, the beneficial effect of illustrations on the comprehension of the physics notions was observed in control condition and this effect was preserved despite of the concurrent tasks. The processing of illustrations did not compete with the visuospatial concurrent task. At first glance, these data fail to show that the integration of texts and illustrations involves the visuospatial working memory specifically. How to account for such data?

A first reason relates to the features of the control condition. It could be argued that the task used in this condition, because it requires a careful analysis of both verbal and visual materials interrupts and disturbs the process of integrating text and illustrations. However, the level of comprehension performance observed in this condition was similar to the performance observed in previous experiments in which a normal reading procedure was used (Gyselinck, 1995). In fact, the concurrent load conditions seem to enhance processing with respect to the control condition, especially in the high spatial span students. These results are congruent with Wenger & Payne’s (1996) observations that a concurrent spatial task may facilitate comprehension of a hypertext compared with a classic linear text. As the Text-plus-Illustrations format involved both verbal and visuospatial material, in this case the positive effect was found for both verbal and visuospatial concurrent memory tasks.

Another interpretation relates to the relevance of the visuospatial concurrent task with respect to the particular characteristics of the illustrations used in this experiment. It is now widely accepted that a visual pictorial component and a spatial component have to be distinguished in the visuospatial sketchpad of working memory (Logie, 1995; Pazzaglia & Cornoldi, 1999). Directly derived from the task of Kruley et al. (1994), the visuospatial concurrent task used in this research mainly involved the visual pictorial component. However, a close examination of the material used shows that the illustrations were very different from those used in Kruley et al. (1994). In the present experiment, illustrations depicted few elements and the result of the actions between them, requiring an exploration from left to right. Furthermore, students had to process a sequence of illustrations associated
with sentences. In such a situation, it could be assumed that spatial sequential processing should predominate. In addition, these results depended on the capacity of the visuospatial sketchpad as measured by the Corsi-blocks test which is mainly a spatial test. This tends to corroborate the interpretation that the failure of the visual concurrent task to produce a negative effect on performance could be due to its non congruence with respect to the requirements of the processing of the illustrations used.

In order to test this interpretation, a second experiment is currently in progress with the same material (text-only format and text-plus-illustrations format). Two other concurrent tasks are used, and in the control condition no task at all is to be performed. The concurrent tasks, widely used in working memory studies, involve either the spatial sequential component of the sketchpad (sequential tapping task) or the phonological loop (verbal articulatory task), while imposing low demands upon the central executive (Farmer et al. 1986). In this second experiment, the role of the individual spatial span as measured by the Corsi-blocks test is more systematically taken into consideration. In agreement with the present hypotheses, the first results suggest that the spatial concurrent task totally cancels the benefit due to the presence of illustrations, but does not impair text-only performance at all. In contrast, the verbal concurrent task upsets both formats of presentation similarly, maintaining the benefit of illustrations. Such a selective impairment due to the spatial concurrent task demonstrates the involvement of the spatial component of the sketchpad in the integration of text and illustrations. They also suggest that the involvement of visuospatial working memory in the integration of text and illustrations is dependent on the capacity of this subsystem as measured by the Corsi-blocks test.

In conclusion, these data show the advantages of illustrations in understanding and memorising scientific texts presented on a computer screen. These advantages appear to be related to the capacity of the illustrations to enhance enriched representations of the texts, possibly mental models (Gyselinck, 1995), necessary in order to make inferences. The ability to make inferences is critical to the learning process, giving an index of depth of encoding and the possibility of maintaining and generalising acquired information. Instructional multimedia systems promoting the integration of verbal and pictorial information therefore appear to be particularly important.

Data also suggest that the integration of texts and illustrations involves the visuospatial working memory. Students appears to be able to cope with the double task involving both the visual component of the concurrent task and the spatial sequential component of the processing of illustrations. However, students could not process the illustrations when the concurrent task loads also the spatial component. These data suggest that the conception and utilisation of multimedia systems must take into account the characteristics of the various types of information which are associated with respect to the kinds of processing they require. They also emphasise the need to consider the limitations in cognitive resources available to the learner. More specifically, the limitation of the visuospatial working memory appears to play an important role in the efficient integration of verbal and pictorial information.

References