Multimedia Learning and Cognitive Load Theory: 
Effects of Modality and Cueing

Huib K. Tabbers 
Rob L. Martens 
Jeroen J.G. Van Merriënboer 

Educational Technology Expertise Center (OTEC), 
Open University of The Netherlands, Heerlen, The Netherlands

The experiment described in this paper is part of a Ph.D. project that is co-funded by SPC Group, a multimedia company in ’s Hertogenbosch. The experiment was carried out at the University of Gent in Belgium and we would like to thank Martin Valcke and Tammy Schellens from the Faculty of Education for their support. Moreover, we would like to thank Patricia Coors for her technical assistance and Marijn, Tim and Myga for their excellent rating work.

This paper is a tentative version of an article.

Correspondence concerning this paper should be addressed to Huib Tabbers, Educational Technology Expertise Centre (OTEC), Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands. Electronic mail may be sent to huib.tabbers@ou.nl.
Abstract

The design of effective multimedia instructions is investigated within the framework of Cognitive Load Theory and Generative Theory of Multimedia Learning. According to these theories, replacing visual text with audio (modality effect) and adding visual cues to a picture (cueing effect) increase effectiveness by lowering the mental workload of the learner. The generalizability of these effects was tested using instructions with a non-technical subject matter, instructional design, and longer training times. For 1 hour, 151 students studied the material and afterwards completed two tests. Measures of mental effort were administered during instruction and tests. Results show that adding visual cues only had a small effect, while replacing the text with audio decreased the effectiveness, contrary to the expectations. A plausible explanation is that the instructions were user-paced, contrary to the system-paced instructions used in earlier research.
Multimedia Learning and Cognitive Load Theory: Effects of Modality and Cueing

The use of multimedia computers in education has led to the development of all sorts of instructional material in which verbal and nonverbal presentation modes are combined. Unfortunately, educational research has not yet given clear answers to the question how to design effective and efficient multimedia instructions. However, two recent theories that have yielded some promising results are Mayer's Generative Theory of Multimedia Learning (Mayer, 1997) and Cognitive Load Theory (Sweller, 1988; Sweller, van Merriënboer & Paas, 1998).

Both theories base their instructional design principles on the cognitive architecture of the learner and the way in which information is processed. In his Generative Theory of Multimedia Learning, Mayer (1997) describes how the learner builds mental representations of multimedia material. One important step in this process is the integration of both verbal and nonverbal information in working memory. For example, when instructions consist of a picture and an explanatory text, the learner has to switch back and forth between the two and integrate them mentally. However, this process is cognitively demanding, and at the expense of mental resources that could otherwise be allocated to the learning process. Cognitive Load Theory calls the unnecessary memory load caused by the presentation format of instructions extraneous load (Sweller et al., 1998). Changing the presentation format of instructions can lower this extraneous load and increase its effectiveness.

In their empirical research, Sweller and Mayer have demonstrated a number of ways in which changes in presentation format can increase the effectiveness of multimedia instructions. One of these changes is replacing written text with spoken text (called the modality effect) and another one is adding visual cues to a picture to prevent visual search (the cueing effect).

The modality effect implies that multimedia instructions are more effective when the verbal information is presented auditorily instead of visually. This has been demonstrated by a number of researchers who showed that replacing written or on-screen text with audio led to a lower mental effort during instruction (Kalyuga, Chandler & Sweller, 1999; Tindall-Ford, Chandler & Sweller, 1997), less time on subsequent problem solving (Jeung, Chandler & Sweller, 1997; Mousavi, Low & Sweller, 1995) and improved test scores (Kalyuga et al.; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Tindall-Ford et al.). Most authors explain their results by referring to the working memory model of Baddeley (1992). In this model, working memory has two modality-specific slave systems, one for visual and spatial information and one for acoustic information. When information is presented in two sensory modalities (visual and auditory) rather than one, both slave systems are addressed and total working memory capacity is increased. So relative to the available resources, the extraneous load of the multimedia instructions is reduced.

However, an alternative explanation of the modality effect is also available. Earlier research on multimedia instructions has shown that physical integration of verbal and nonverbal information also results in a lower extraneous load and better learning (Chandler & Sweller, 1991; Chandler & Sweller, 1992; Sweller, Chandler, Tierney & Cooper, 1990; Tarmizi & Sweller, 1988). Placing a textbox next to the part of the picture the text is referring to prevents the learner from splitting his or her attention between text and picture, and shortens the time needed to keep information active in working memory. Sweller et al. (1998) call this phenomenon the split-attention effect. The same effect has been found by Mayer in a series of experiments (see Mayer, 1997, for an overview) that showed that multimedia instructions were more effective when verbal and visual information were presented simultaneously rather than separately in time and space. Mayer calls this the contiguity effect. Now the modality effect can also be explained in terms of preventing split-attention by integrating verbal and nonverbal information. After all, with the text presented as audio the learner can listen to the narration and look at a picture at the same time. This idea is supported by the results of one experiment from Tindall-Ford et al. (1997), that showed that physically integrated and bimodal instructions were equally effective. On the other hand, experiments by Mousavi et al. (1995) and Moreno and Mayer (1999) demonstrated that a modality effect also occurred when picture and text...
were presented sequentially instead of simultaneously. So despite the temporal detachment of text and picture, bimodal instructions still proved to be superior to visual-only instructions. That means that although the modality effect can be explained in terms of preventing split-attention, at least some part of the effect must be the result of a different mechanism, like an increase in available memory resources.

Replacing visual text with audio does not always improve the efficiency and effectiveness of multimedia instructions. Jeung et al. (1997) did not find a modality effect using pictures with a high visual complexity. Only after adding visual cues to the pictures in the form of electronic flashing they found an effect in terms of shorter time on subsequent problem solving for the bimodal conditions. This cueing effect has also been found with visual-only instructions. In one experiment, Kalyuga et al. (1999) used color-coding to link on-screen text with corresponding parts of the picture. This resulted in better test scores when compared to instructions without visual cues. Preventing unnecessary visual search in visual-only instructions seems to have the same effect as physical integration, as it shortens the time needed to keep information active in working memory. With bimodal instructions however, the learner has to search the picture at the same time as he or she is listening to the text. When the complexity of a picture is high, this visual search causes a kind of temporal split-attention, because the text and the relevant parts of the picture are not perceived simultaneously any more. Adding cues to a picture prevents visual search and thus restores the temporal contiguity of text and picture. So the cueing effect seems to work differently for visual-only and bimodal instructions, preventing either spatial or temporal spatial split-attention. However, in both cases adding visual cues to a complex picture decreases the extraneous load and increases the effectiveness of multimedia instructions.

Almost all multimedia instructions used in the above-mentioned studies taught subjects from technical domains like geometry (Mousavi, Low & Sweller, 1995; Tarmizi & Sweller, 1988), Jeung et al., 1997), scientific explanations of how lightning develops (Mayer & Moreno, 1998; Moreno & Mayer, 1999) and electrical engineering (Kalyuga et al., 1999; Tindall-Ford et al., 1997). The instructions consisted of worked-out examples or contained pictures of a model or machine in combination with a textual explanation of its functioning. In most cases, students had to study the instructional material for only a few minutes, and were tested immediately afterwards. Furthermore, only the studies by Tindall-Ford et al. (1997) and Kalyuga et al. (1999) administered mental effort measures during instruction. The other studies did not measure cognitive load at all, which makes it difficult to interpret their findings in terms of Cognitive Load Theory.

The present study investigates both modality and cueing effects using measures of mental effort, to be able to draw conclusions about the cognitive load of the different presentation formats. Moreover, to test for the generalizability of earlier findings, instructions on a subject matter outside the technical domain are used and training time is longer than in earlier studies. For this purpose we developed multimedia instructions on the Four Component Instructional Design (4C/ID)-model of Van Merriënboer (1997). This model describes a design strategy for the training of complex cognitive skills. To be able to study modality and cueing effects four versions of the instructions were created, varying in both the modality of text and the use of visual cues. In the two visual conditions all texts were presented visually, either with or without cues in the diagrams. The two audio conditions had all texts accompanying diagrams presented auditorily, also either with or without cues in the diagrams.

The hypotheses that follow from Cognitive Load Theory and Mayer's Generative Theory of Multimedia Learning are that presenting the texts accompanying the diagrams as audio will decrease the extraneous load and increase the effectiveness of the instructions (the modality effect), and that adding visual cues to the diagrams will prevent visual search and also decrease extraneous cognitive load and increase the effectiveness of the instructions (cueing effect). To measure the effectiveness of the instructions, we looked at the extent in which students could recall elements of the 4C/ID-model (reproduction), and at the extent in which they could apply the model in a new situation (transfer). Furthermore, to be able to draw conclusions not only about the effectiveness but also about the cognitive efficiency of the different presentation modes, we looked at the mental effort spent on both the instructions and the tests.
Method

Participants

The participants were 151 second-year students from the Department of Education of the University of Gent in Belgium (age between 20 and 25 years, 37 males and 114 females). The experiment was part of a course on instructional design, but at the time of the experiment the students had not received any lessons on instructional design models yet. Forty participants were randomly assigned to the VN condition (visual text, no cues in diagram), 36 to the VC condition (visual text, cues in diagram), 36 to the AN condition (auditory text, no cues in diagram) and 39 to the AC condition (auditory text, no cues in diagram).

Materials

Instructions. The multimedia instructions on the 4C/ID-model of Van Merriënboer (1997) focused on how to develop a blueprint for a training program based on the skills hierarchy of a complex skill. The material was constructed as a web site with a linear structure that first offered two worked-out examples followed by a general explanation of the design strategy. After a short introduction to the 4C/ID-model, the first worked-out example showed the different stages in developing a blueprint for the training of the complex skill doing experimental research in a series of six diagrams, showing skill hierarchies and sequences of learning tasks (see Figure 1 for a screen example). The second worked-out example consisted of three diagrams showing the same process for designing a house, and finally the strategy of the 4C/ID-model was summarized in two diagrams. The explanation that accompanied each of the eleven diagrams was cut up into smaller explanatory texts of only a few sentences long. Each piece of text referred to a different part of the diagram. The students could scroll through the texts at their own pace.

Figure 1 One of the screens in the VN-condition showing a diagram of a learning sequence accompanied by a small piece of explanatory text at the top. By clicking on one of the arrows next to the page number students could move forwards or backwards through the text.

Four different versions of the web site were created that differed in the way the texts accompanying the diagrams were presented and in the use of visual cues in the diagrams. In the two audio
conditions, students could listen to the texts through a headphone and were able to replay them by clicking on a play-button, while in the two visual conditions texts were depicted above the diagrams. The visual cueing was done by coloring bright red those parts of the diagram the explanatory text was referring to.

Subjective mental effort scale. After the explanation of each diagram, a separate page followed with a subjective rating scale to measure the mental effort spent on the instructions. This nine-point scale ranged from very, very low mental effort to very, very high mental effort. When a student clicked on one of the nine options of the rating scale, the program automatically continued with the next diagram. The average score on the eleven mental effort scores (one for each diagram) was taken as a measure of mental effort during instructions. The subjective rating scale used in this study was developed as a measure of cognitive load by Paas and others (Paas, 1992; Paas & van Merriënboer, 1993; Paas & van Merriënboer, 1994; Paas, van Merriënboer & Adam, 1994). Paas and Van Merriënboer (1994) show that mental effort is just one dimension of cognitive load that is not only influenced by task-characteristics but also by subject characteristics like prior knowledge and subject-task interactions like motivation. We tried to exclude any of these effects by randomization of our subjects over the conditions.

Reproduction test. The reproduction test originally consisted of two multiple-choice tests of 20 and 30 items. All items were statements about the 4C/ID-model like “A macro-sequence in the 4C/ID-model is a series of subskills in a cluster”, or “According to the 4C/ID-model, the same subskills can be trained in more than one learning task”, and the students could choose between correct, incorrect or I don’t know. The 30-item test only contained verbal statements, while the 20-item test combined the verbal statements with small parts of diagrams. A right answer yielded one point, a wrong answer minus one point and not knowing zero points. After item analysis we summed the scores of 35 items taken from both tests into one total reproduction score (Cronbach’s alpha = .73).

Transfer test. The transfer test contained a short description of the skills an expert researcher applies when he or she is searching for literature, followed by the assignment to design a blueprint for the training of this complex skill according to the 4C/ID-model. To be able to score the results of the transfer test a scoring form was developed consisting of forty yes/no-questions that checked to what extent and how accurately the strategy prescribed by the model had been applied in the transfer task. Every yes scored one point, and the sum score was converted to a percentage, ranging from zero (no steps from the model taken) to 100 (all steps taken accurately). Three independent raters scored the transfer tests using the form. The tests of twenty-six students were scored by all three raters, showing an interrater agreement of .88 (calculated as a single measure intraclass correlation).

Procedure
The experiment was carried out in four sessions and in each session between thirty-five and forty students were tested simultaneously. The sessions took place in a classroom that had forty multimedia computers connected to the Internet through the university network, with ten computers for each experimental condition. The ones that delivered bimodal instructions had headphones attached to them. When the students entered the room they were randomly assigned to one of the computers. Each computer showed a browser-window (without any of the menu options visible) set on the first page of the introduction. The students first read some general information on how to navigate in the instructional material and how to fill in the mental effort scales. In the auditory conditions, students were also reminded that they had to use the headphones during the instructions. Furthermore, all participants were told that after the instruction they would have to make two multiple-choice tests that tested their knowledge of the 4C/ID-model and that they would have to design a training according to the model. When the students had finished reading, they were given the opportunity to ask the experimenter any questions about the procedure. Subsequently, the students logged in onto the actual website with the instructions by typing in their student number. They had a maximum of seventy minutes to study the instructional material, but if they finished earlier, they could do something for themselves like browsing the Internet. The server on which the instructional website
ran kept record of the time spent on the learning task and of the mental effort scores of each participant.

After the instruction phase the computers were shut off and the three paper-and-pencil tests were administered. In two of the sessions students first got the two multiple-choice tests followed by the transfer test, while in the other two sessions the order was reversed to prevent any sequencing effect. Maximum time for each multiple-choice test was ten minutes, and in the transfer test the students had thirty minutes to design a training. After each test students rated the mental effort spent on the test on a nine-point scale similar to the ones used in the learning material.

Results

The variables under analysis were training time, reproduction score, transfer score, and mental effort spent on instruction and on the tests. All scores were analyzed with two-factor analyses of variance (ANOVAs), with modality (visual vs. auditory text) and cueing (no cues vs. cues in the diagram) as the between-subjects factors. For all statistical tests, a significance level of .05 was applied. Table 1 shows the average scores on the dependent measures for all four conditions.

Table 1 Mean Scores on Dependent Measures for all Conditions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Condition VN (n = 40)</th>
<th>Condition VC (n = 36)</th>
<th>Condition AN (n = 36)</th>
<th>Condition AC (n = 39)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Time in Minutes</td>
<td>48.5  8.4</td>
<td>52.8  10.1</td>
<td>59.2  7.9</td>
<td>61.1  8.1</td>
</tr>
<tr>
<td>Mental Effort during Instructions</td>
<td>4.1  1.0</td>
<td>4.2  0.9</td>
<td>4.0  0.9</td>
<td>4.0  1.0</td>
</tr>
<tr>
<td>Reproduction Score</td>
<td>16.5  7.9</td>
<td>18.6  7.8</td>
<td>11.9  7.3</td>
<td>14.7  7.0</td>
</tr>
<tr>
<td>Mental Effort on Reproduction Tests</td>
<td>6.5  1.0</td>
<td>6.6  1.1</td>
<td>6.3  1.1</td>
<td>6.1  1.1</td>
</tr>
<tr>
<td>Transfer Score</td>
<td>70      20</td>
<td>69      20</td>
<td>60      17</td>
<td>68      19</td>
</tr>
<tr>
<td>Mental Effort on Transfer Test</td>
<td>6.6  1.5</td>
<td>7.0  1.2</td>
<td>6.3  1.4</td>
<td>6.5  1.4</td>
</tr>
</tbody>
</table>

Training time was not equal for all conditions, with participants in the audio conditions needing significantly more time than the visual conditions ($F(1, 147) = 45.22, MSE = 74.86, p < .001$), and the cued conditions needing more time than the not-cued conditions ($F(1, 147) = 4.93, MSE = 74.86, p < .05$).
Modality and Cueing effects in Multimedia Learning

Figure 2 shows the average scores on the eleven separate mental effort measures that were administered during the instructions. This pattern was nearly identical in all four conditions and followed the increasing complexity of the diagrams used in the instructions. The average score for mental effort during instructions was 4.1 (4 = rather low mental effort), and there were no significant effects modality or cueing on either the eleven separate measures or on the average score.

The results of the reproduction test showed a significant effect for the modality of text ($F(1, 147) = 11.82, MSE = 56.44, p < .001$), with the visual conditions doing better than the audio conditions ($M = 17.5$, and $M = 13.4$, respectively). The effect of adding cues to the diagram also reached statistical significance ($F(1, 147) = 3.85, MSE = 56.44, p < .05$), with a higher score for the cued conditions ($M = 16.5$) than for the non-cued conditions ($M = 14.3$). The mental effort score for the reproduction test showed a significant effect for the modality of instructions ($F(1, 147) = 4.98, MSE = 1.13, p < .05$). Students in the visual conditions ($M = 6.6$) had spent a little more effort on the test than their colleagues in the audio conditions ($M = 6.2$).

The scores on the transfer task showed a marginally significant effect for the modality of the text ($F(1, 147) = 3.23, MSE = 374.14, p = .07$) in the same direction as in the reproduction test ($M = 70$ for the visual conditions, $M = 64$ for the audio conditions). The mental effort score for the transfer test showed no significant differences.

Discussion

It is clear from our results that the strategies that should help to lower the extraneous load of our multimedia instructions and increase its effectiveness have not been wearing off as expected. In terms of effectiveness, the results do show a modality effect, but opposite to what has been found in earlier research. Students in the visual condition perform better on both the reproduction and the transfer test than the students in the audio conditions. Furthermore, the cueing effect is only significant in the reproduction test but not in the transfer test. Also in terms of efficiency, the results deviate from what has been reported in earlier research. The mental effort spent during the instructions is equal for all conditions, contrary to the results of Tindall-Ford et al. (1997) and Kalyuga et al. (1999). That could mean that the differences between the conditions were very small and that the measurements were too insensitive to notice them. However, the results also show that students’ mental effort
scores vary with the complexity of the diagram, which seems to indicate a sufficient sensitivity of the scale. An alternative explanation for the lack of difference in reported mental effort would be that students have compensated a lower extraneous load with a higher learning effort, resulting in an equal mental effort score. In that case, only the scores on the tests reveal which presentation mode has led to the most efficient use of mental resources. On the reproduction test it is remarkable that the higher scores for the visual conditions are coupled with higher mental effort scores. So it is possible that the students in the visual conditions have reached a better performance on the reproduction test by investing more mental effort. However, the visual conditions do show better scores on the transfer test without any differences in mental effort spent on the test. Moreover, students in the visual conditions have spent significantly less time on the instructions, which only strengthens the conclusion they that have outperformed their colleagues in the audio conditions.

Why do our results differ so significantly from those achieved with earlier research on multimedia learning and cognitive load, especially regarding the modality effect? First of all, the instructions used in our experiment had a subject matter outside the technological domain used in earlier experiments. Although theoretically this should not have any influence on the cognitive load imposed, it might be argued that instructional design strategies are more conceptual and less procedural in nature than for example the scientific explanations from Moreno and Mayer (1999). And visual text might be more suitable for presenting conceptual information than auditory text, as the learner has more time to reflect on the information. On the other hand, the instructions on the 4C/ID-model used quite straightforward procedural descriptions of the design process, and students in the auditory conditions had the opportunity to listen to a piece of text as many times as they wished, giving them ample opportunities to elaborate on the information.

The length of the instructions might be a more plausible factor explaining why bimodal instructions were not superior to visual-only instructions. In our experiment, students spent more than an hour studying the instructional material, which contrasts sharply with the few-minutes instructions used in earlier research on the modality effect. Differences in extraneous load that have an influence in short learning tasks may loose their influence as more time-related factors are becoming dominant in the learning process, such as concentration and span of attention. Listening might be even more tiresome or boring than reading, resulting in less motivation. This could explain why students in the audio conditions have spent less mental effort on the tests and achieved lower test scores. However, the mental effort measures during instruction do not indicate any differences between the conditions, but instead show the same gradual increase in mental effort for all conditions, which doesn’t point to any differences in motivation or concentration during the instructions.

Another point that can be made is that the instructions were not complex enough. Tindall-Ford et al. (1997) for example showed that the modality effect did not occur when the intrinsic complexity of the instructions was not high enough. In our study, the reported mental effort scores during instructions are below average, which could indicate that the instructions were not complex enough. On the other hand, the subject of instructional design was completely novel to our students, and the diagrams used in the instructions had a high level of complexity. It is therefore not very likely that our instructions were too easy. Moreover, the modality effect has been demonstrated before with low or average mental effort scores, as in Kalyuga et al. (1999).

None of the explanations given so far provides a clear and definite answer to the question why our results are different from those of earlier research. However, a closer look at the way in which instruction modes were experimentally compared in each study reveals two other factors that might account for the differences in results. First of all, Jeung et al. (1997), Mousavi et al. (1995) and Tindall-Ford et al. (1997) used visual-only instructions in which all explanatory text was printed next to the diagram and compared it to instructions in which the students only saw the picture and could listen to the explanation. That means that they not only replaced visual text with audio, but also drastically reduced the visual search necessary to be able to mentally integrate the text with the diagram. So in their experiments, the difference in effectiveness between bimodal and visual-only instructions can be largely attributed to the difference in visual complexity. In our experiment, the
explanatory texts in all conditions were all cut to smaller pieces, so that students in the visual conditions only had to search the picture and did not have to search the text as well. Although this reduces the difference in visual complexity with the bimodal conditions, this still does not explain why the students in the latter condition obtained even better test scores.

Mayer and Moreno (1998) and Kalyuga et al. (1999) also cut up their explanatory texts in smaller pieces and still found a modality effect. However, in their experiments the instructions were presented as system-paced animations. The time a student could study a picture and its accompanying texts was determined by the speed of the narration in the bimodal condition. In our study however, the students could scroll through the explanatory texts at their own pace. Now it could be argued that bimodal instructions are advantageous when animations are system-paced, while visual-only instructions are more effective when the student sets the pace. In animated instructions, the advantage of bimodal instructions is that picture and text can be perceived simultaneously, leading to a lower extraneous load than visual-only instructions in which the learner has to skip between text and picture in limited time. In student-paced instructions however, this advantage disappears because the learner with the visual-only instructions has more time to relate the text to the picture, which will lead to a lower extraneous load especially when visual search is not too high. Moreover, with visual texts it is much easier to jump back and forth through the text than with auditory texts that are linear by nature and are much less easy to skim through. Especially when texts are complicated or very long, this could make visual-only instructions more effective than bimodal instructions and reverse the modality effect. In future research, this hypothesis can be investigated by comparing system-paced with student-paced bimodal instructions.

Taken together, our results show that the guidelines for the design of multimedia instructions that follow from Cognitive Load Theory and Mayer’s Generative Theory of Multimedia Learning are not just generally applicable. Although we did find a small cueing effect in our experiment, we could not replicate the modality effect. Replacing visual text with audio even had a negative effect on learning contrary to what both theories would predict. A possible explanation for this finding is that a bimodal presentation is only advantageous when the system sets the pace of the instructions, while visual-only instructions are the preferred format if the learner has control. However, this is a speculative account that needs to be investigated in more detail. Further research into the conditions under which the modality and cueing effects occur may produce more specific rules for the design of multimedia instructions. By including mental effort measures in this research, not only the effectiveness but also the cognitive efficiency of the instructions can be determined, which will help to explain the findings in terms of the processes that take place in the learner’s mind.

References


