# The modality effect in multimedia instructions

**Huib K. Tabbers (huib.tabbers@ou.nl)**
**Rob L. Martens (rob.martens@ou.nl)**
**Jeroen J. G. van Merriënboer (jeroen.vanmerrienboer@ou.nl)**
Open University of the Netherlands: Educational Technology Expertise Centre; P.O. Box 2960
NL-6401 DL Heerlen, The Netherlands

**Abstract**

The influence of presentation format on the effectiveness of multimedia instructions was investigated. According to Cognitive Load Theory (Sweller, Van Merriënboer & Paas, 1998) and Mayer's theory of multimedia learning (Moreno & Mayer, 1999), replacing visual text with audio will decrease working memory load and improve learning (modality effect). This hypothesis was tested in two experiments in which students studied multimedia instructions on an instructional design model. The students reported the mental effort spent on the instructions, and made a retention and a transfer test after the instructions. The results show that replacing text with audio is only effective when multimedia instructions are system-paced.

## Introduction

Guidelines for the design of multimedia instructions are often based on intuition and practical experience rather than on the results of experimental research (Park & Hannafin, 1994). However, two recent lines of research that have yielded some interesting results are the work by John Sweller and his colleagues on Cognitive Load Theory (Sweller, 1988; Sweller, van Merriënboer & Paas, 1998), and the experiments carried out by Richard Mayer and his colleagues on multimedia learning (for an overview, see Moreno & Mayer, 1999). Both researchers claim that multimedia instructions consisting of verbal and pictorial information, like for example a picture of a machine and a text about its functioning, place a high demand on working memory resources, because the learner has to switch between text and picture in order to integrate them mentally. An interesting finding in their research is that this memory load can be reduced by presenting the verbal information auditorily instead of visually. They call this phenomenon the *modality effect* or *modality principle*. The explanation they give is based on the working memory model of Baddeley (1992). In his model, working memory has two modality-specific slave systems: one for processing 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 memory load of the multimedia instructions is reduced, leaving more space for the actual learning process.

Sweller and Mayer have demonstrated the superiority of audio over written or on-screen text in a number of experiments. For example, Jeung, Chandler and Sweller (1997) and Mousavi, Low and Sweller (1995) showed that students receiving multimedia instructions with audio spent less time on subsequent problem solving compared to students receiving visual-only instructions. Furthermore, students in experiments by Kalyuga, Chandler and Sweller (1999) and Tindall-Ford, Chandler and Sweller (1997) reported less mental effort during instruction and attained higher test scores, while in the studies by Mayer and Moreno (1998; 1999) students had higher scores on retention, transfer and matching tests. In one experiment, Moreno and Mayer (1999) even used instructions in which the animation and the accompanying text were presented sequentially instead of simultaneously. Despite the temporal detachment of text and picture, bimodal instructions still proved to be superior to visual-only instructions. This shows that the modality effect seems to be at least for some part the result of an increase in available memory resources.

Based on these results, Sweller and Mayer strongly advocate the use of audio in multimedia instructions. However, one limitation of the above-mentioned studies is that they all deal with short multimedia instructions on well-defined technical subjects like geometry (Mousavi et al., 1995; 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). This raises the question how powerful the modality effect actually is. Can it also be demonstrated with multimedia instructions that are outside the technical domain and are of greater length? This question is dealt with in the first experiment of this study.

The second issue that can be raised given the evidence so far, is that the results can be explained in more than one way. For example, Jeung et al. (1997), Mousavi et al. (1995) and Tindall-Ford et al. (1997) used visual-only instructions in which the complete 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 reduced the visual search necessary to link the right parts of the text with the right parts of the diagram. So in their experiments, the difference in effectiveness between bimodal and visual-only instructions could also be attributed to the difference in visual complexity.

On the other hand, Mayer and Moreno (1998; 1999) and Kalyuga et al. (1999) cut 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. The learners in the bimodal condition could use this limited period of time effectively because they could look at the picture and listen to the text at the same time. The learners in the visual-only condition on the other hand had to spend part of their time in a process of skipping back and forth between text and picture in order to integrate them mentally. We question if the modality effect will still appear if you give the students in the visual condition more time to relate the text to the picture. This issue is dealt with in the second experiment.

## Experiment 1

The aim of our first experiment was to see if we could replicate the modality effect using longer multimedia instructions in a non-technical subject domain. For this purpose we developed web-based instructions on an instructional design model. The material mainly consisted of diagrams with explanatory texts. Jeung et al. (1997) showed that replacing visual text with audio does not always improve the effectiveness of multimedia instructions, especially when using pictures with a high visual complexity. They argued that the visual search needed to find the part of the picture the text is referring to increases the memory load. After adding visual cues to the pictures in the form of electronic flashing they regained the modality effect. In our experiment we used colour coding as a means of preventing unnecessary visual search.

The hypothesis that follows from Cognitive Load Theory and Mayer's work on multimedia learning is that presenting the texts accompanying the diagrams as audio will decrease the working memory load of the instructions. Therefore we divided the students in two groups, one receiving bimodal instructions (the audio group) and one receiving visual-only instructions (the visual group), and measured the mental effort spent on the instructions. Paas and Van Merriënboer (1994) argue 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 groups, so that differences in mental effort scores could be attributed to the differences in presentation format.

It is also possible that the freed working memory resources are used for the learning process itself and no differences in mental effort are reported. Therefore we also looked at the extent in which students could recall elements of the design model in a retention test, and at the extent in which they could apply the model in a new situation with a transfer test, to see if there was any difference in performance. Finally, we also measured the mental effort spent on both tests.

### Method

**Participants** The participants were 41 students from a Teacher Training College for Primary Education in Heerlen, the Netherlands (20 second-years and 21 third-years; age between 18 and 24; 11 males and 30 females). They had applied on a voluntary base and were paid forty guilders for their participation. During their studies, the students hadn't had any lessons on instructional design models. Twenty participants were randomly assigned to the visual group and 21 to the audio group.

**Materials** We developed web-based 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. The instructions focused on the question how to develop a blueprint for a training program based on the skills-hierarchy of a complex skill. The instructional website started with a short textual introduction to the model. Subsequently, the design strategy of the 4C/ID-model was demonstrated in a series of eleven diagrams representing skill hierarchies and sequences of learning tasks. These diagrams formed two worked-out examples and a general explanation of the strategy. The first example consisted of six diagrams that showed the different stages in developing a blueprint for the training of the complex skill *doing experimental research* (see Figure 1a and 1b for screen examples). The second worked-out example consisted of three diagrams showing the same process for the complex skill *designing a house*, and finally the general strategy of the 4C/ID-model was explained in the last two diagrams.

All eleven diagrams were accompanied by a textual explanation on how the model was applied in the specific situation. These explanatory texts were cut up into smaller pieces of only one or two sentences long, in such a way that each piece of text referred to a specific part of the diagram. Moreover, these parts were coloured bright red in the diagram to prevent any unnecessary visual search. So while studying a diagram, only the accompanying text and the colour coding changed, not the diagram itself.

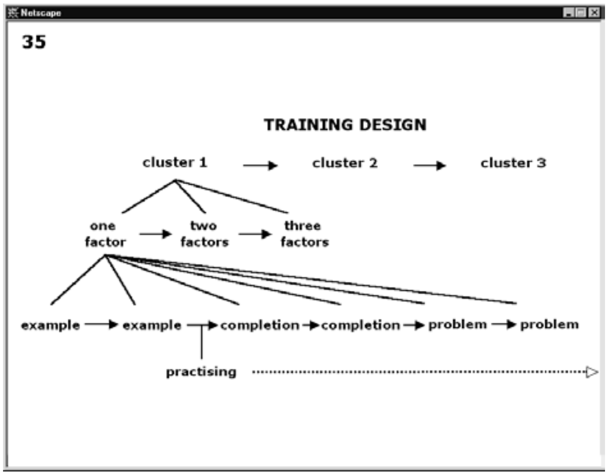


Figure 1a: Screen example of the audio version of the multimedia instructions.

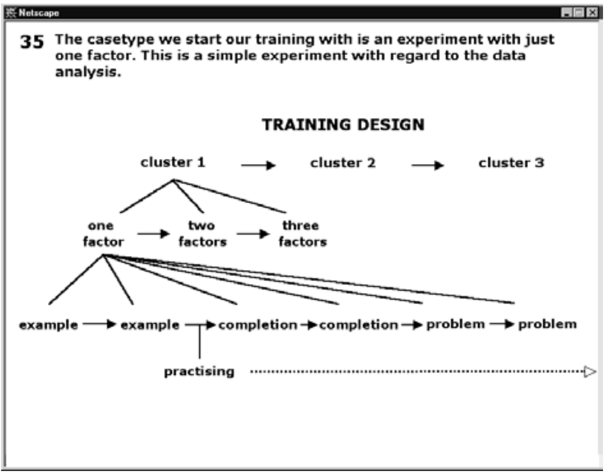


Figure 1b: Screen example of the visual text version of the multimedia instructions.

Two versions of the instructional website were created that differed in the way the texts accompanying the diagrams were presented. In the audio version (Figure 1a), students could listen to the pieces of explanatory text that accompanied a diagram through a headphone. Three seconds after the audio had finished playing, the colour coding in the diagram changed and the next piece of audio started. In the visual text version (Figure 1b) the pieces of explanatory text were depicted right above the diagrams. After the same period of time as in the audio condition, the colour coding in the diagram changed and a new piece of text appeared above the diagram. The time it took to study all eleven diagrams was about 30 minutes

After each diagram, a separate page followed with a nine-point rating scale on which the students could rate the mental effort they had spent on the instructions. This scale was developed by Paas and others (Paas & van Merriënboer, 1994; Paas, van Merriënboer & Adam, 1994). When a student clicked on one of the nine options, the program automatically continued with the next diagram. The average score on the eleven rating scales was taken as a measure of mental effort during instructions.

The retention test consisted of two paper-and-pencil tests, one of 30 and one of 20 multiple-choice items. The 30-item test contained only verbal statements, while the 20-item test combined verbal statements with small parts of diagrams. 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*. Each right answer yielded one point. The retention score was calculated by taking the sum of the scores on all fifty items (Cronbach's alpha = .74). A nine-point rating scale similar to the ones used in the instructions followed both multiple-choice tests. The average score on both scales was taken as a measure of the mental effort spent on the retention test.

The transfer test was also a paper-and pencil test that contained a short description of the skills an expert researcher applies when he or she is doing a literature search. The assignment was to design a blueprint for the training of this complex skill according to the 4C/ID-model on a blank answering form. After this test again a nine-point rating scale had to be completed as a measure of the mental effort spent on the transfer test. To be able to score the results of the transfer test a scoring form was developed consisting of twenty-eight *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 ranged from zero (no steps from the model taken) to 28 (all steps taken accurately). After the experiment, two independent raters scored the transfer tests using the form, showing an inter-rater agreement of .95. The average rating score was taken as the transfer score.

**Procedure** The experiment was carried out in eight sessions of about two hours, and in each session between one and seven students were tested simultaneously. These sessions took place in a multimedia lab that had seven computers connected to the Intranet of the Open University. Three computers 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 a webpage displaying some general information about the experiment. When the students had finished reading, the experimenter told all the students to log in onto the actual instructions by typing in a password. All students started at the same time and studied the instructions all by themselves. The server on which the instructional website ran kept record of the mental effort scores of each participant.

After the instruction phase the three paper-and-pencil tests were administered. For each multiple-choice test the students got ten minutes, and for the transfer test they got thirty minutes.

### Results

The variables under analysis were mental effort spent on the instructions, on the retention test and on the transfer test, and retention and transfer score. All scores were analysed with one-tailed t-tests. For all statistical tests, a significance level of .05 was applied. Table 1 shows the average scores on the dependent measures for the experimental groups.

Table 1: group means on dependent measures (standard deviations in brackets)

| | audio | visual text |
|---|---|---|
| mental effort instructions | 4.3 (0.8) | 4.9 (0.9) |
| mental effort retention test | 6.2 (0.8) | 6.4 (1.2) |
| mental effort transfer test | 6.4 (1.4) | 7.1 (1.1) |
| retention score (0-50) | 31.4 (6.1) | 29.8 (5.4) |
| transfer score (0-28) | 9.6 (6.2) | 10.3 (5.4) |

The reported mental effort during instructions showed a significant effect for the modality of text (t(39) = -2.19, $p < .05$). Students in the audio group had spent less effort than their colleagues in the visual group. The mental effort spent on the retention test showed no differences between the groups ($t(39) = -0.53$, $p > .10$). However, the mental effort scores in the transfer test did show a significant difference between the groups ($t(39) = 3.42$, $p < .05$), with the students in the audio group spending again less effort than their colleagues in the visual groups.

Although the audio group did a little better than the visual group on the retention test, this effect was not significant ($t(39) = 0.88$, $p > .10$). Also no significant difference was found between the groups on the transfer test ($t(39) = -0.40$, $p > .10$).

### Discussion

The results of the first experiment show that the modality effect can be replicated with longer multimedia instructions on a non-technical subject like instructional design. Students in the audio group report lower mental effort scores during the instructions as a result of decreased memory load. This is confirmed by the fact that in both the retention and transfer test the students in the audio group score just as good as the students in the visual group. Moreover, getting the same result in the transfer test has cost them less mental effort. The modality effect is not as strong as in the experiments by Kalyuga et al. (1999) and Tindall-Ford et al. (1997), who found both lower mental effort and better test scores. However, the fact that students in the audio condition reach the same test results with less mental effort still points at the superiority of audio over visual text in multimedia instructions.

## Experiment 2

In our second experiment we wanted to investigate the question if the modality effect can still be found if the students in the visual group get more time to relate the verbal information to the diagram. Therefore we not only varied the modality of the text, but also the pacing of the instructions. The system-paced groups were identical to the two groups in the first experiment, while students in the user-paced groups could set the pace of the instructions for themselves. This way we compared four groups: audio-user, audio-system, visual-user, and visual-system.

### Method

**Participants** The participants were 81 second-year students from the Department of Education of the University of Gent in Belgium (age between 18 and 30 years; 8 males and 73 females). The experiment was part of a regular course on instructional design, but at the time of the experiment the students had not received any lessons on instructional design models yet. Eighteen participants were randomly assigned to the audio-user group, another 18 to the audio-system group, 24 to the visual-user group, and 21 to the visual-system group.

**Materials** The multimedia instructions were the same as in the first experiment, only two extra user-paced versions were created. In the audio-user version (Figure 2a), the students were able to replay the sentences they had just heard by clicking on a small *play*-button, while in the visual-user version (Figure 2b) students could reread the text as many times as they wanted to. To continue with the next piece of text students in both groups had to click on a forward arrow.

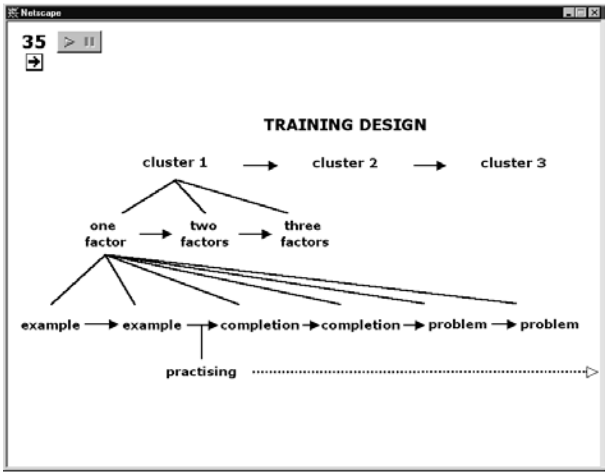


Figure 2a: Screen example of the audio-user version of the multimedia instructions.

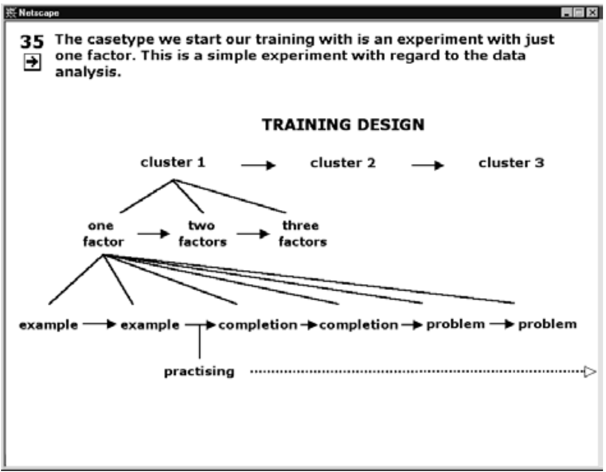


Figure 2b: Screen example of the visual-user version of the multimedia instructions.

The measurements were the same as in the first experiment, only this time all tests were presented on the computer. Moreover, the retention test consisted of 40 items taken from the retention test of the first experiment. The sum of the 40 items formed the total retention score (Cronbach's alpha = .68). After the experiment, two independent raters scored the transfer tests, showing an inter-rater agreement of .92.

**Procedure** The experiment was carried out in four sessions of about two-and-a-half hour, and in each session between fifteen and twenty-four students were tested simultaneously. These sessions took place in a classroom that had twenty-four multimedia computers connected to the Internet through the university network, with six computers for each experimental group. The procedure was almost identical to the first experiment. Only this time, students could immediately continue with the tests whenever they had finished studying the instructions. The server on which the instructional website ran kept record of the time spent on the learning task (in minutes), of the mental effort scores and of the retention score of each participant.

## Results

The variables under analysis were training time, mental effort spent on the instructions, on the retention test and on the transfer test, and retention and transfer score. Except for training time, all scores were analysed with two-factor analyses of variance (ANOVAs), with modality (audio vs. visual text) and pacing of the instructions (system pacing vs. user pacing) as the between-subjects factors. For all statistical tests, a significance level of .05 was applied. Table 2 shows the average scores on the dependent measures for all four groups.

Table 2: group means on dependent measures (standard deviations in brackets)

| | audio-user | audio-system | visual-user | visual-system |
|---|---|---|---|---|
| time on instructions | 33.7 (3.3) | | 37.8 (5.5) | |
| mental effort instructions | 4.3 (1.0) | 4.1 (0.7) | 4.0 (1.0) | 4.2 (1.0) |
| mental effort retention test | 6.5 (1.3) | 6.7 (1.0) | 6.5 (1.3) | 6.7 (1.0) |
| mental effort transfer test | 7.3 (1.1) | 7.3 (1.5) | 7.5 (1.1) | 7.1 (1.5) |
| retention score (0-40) | 26.5 (4.8) | 28.9 (3.7) | 28.6 (3.6) | 25.3 (5.6) |
| transfer score (0-28) | 17.8 (4.4) | 17.7 (4.1) | 16.8 (4.8) | 14.1 (5.6) |

With regard to the time spent on the instructions, only the two user-groups were compared, because in the system groups time was equal for all students (about 30 minutes). It showed that the students in the visual-user group had spent significantly more time on the instructions than the students in the audio-user group ($t(40) = -2.7$, $p < .01$, two-tailed).

There were no significant differences between the groups on mental effort during instructions. The same goes for the mental effort spent on the retention test, and for the mental effort spent on the transfer test.

The results on the retention test showed a significant interaction effect ($F(1,77) = 7.99$, $MSE = 20.16$, $p < .01$). In the two system groups, the audio group did better than visual text, while in the user groups this effect was reversed, with visual text outperforming the audio group. The scores on the transfer task showed a significant main effect for the modality of the text

($F(1,77) = 4.67$, $MSE = 23.07$, $p < .05$), with the students in the audio groups scoring higher than the students in the visual groups ($M = 17.8$, vs. $M = 15.5$, respectively). Inspection of the separate group means shows that especially the students in the visual-system group did worse that their colleagues in the audio groups. However, this interaction was statistically not significant.

### Discussion

The results show that in the system-paced groups, a modality effect is found in terms of improved learning outcomes, but not in mental effort. This is a little different from the results in the first experiment in which the audio group spent less mental effort but did not have better test scores. This reversal might be accounted for by the fact that the second experiment was part of a regular course, and that the students in the audio condition were more prepared to invest the freed memory resources in the learning process itself, resulting in higher test scores with equal mental effort.

However, in the two groups in which the students set the pace of the instructions, no modality effect is found at all. Not only do the students in the visual-user group perform almost equally well on the transfer test, on the retention test they even outperform the students in the audio-user group. The visual-user group has taken more time to study the instructions, which confirms our idea that the modality effect in the system-paced condition is at least partly the result of a lack of time to relate the text to the diagrams in the visual-system group.

## General Discussion

The results of both experiments show that replacing on-screen text with audio will only increase the effectiveness of multimedia instructions if the student has no control over the pacing of the instruction and the pace is set by the time of the narration. In that case we find either lower mental effort or better test results, even with a subject matter from a non-technical domain like instructional design. However, with more time (or the possibility to let the student determine the pace) visual-only instructions can be just as effective as bimodal instructions.

From a theoretical point of view, the results seem to indicate that the modality effect as demonstrated in earlier experiments can be accounted for in other terms than an increase in memory resources. One possible explanation is the lack of time to relate verbal to pictorial information in visual-only conditions. One of the things we will do in our future research is get a closer look at what actually happens when students are studying multimedia instructions by measuring eye-movements and look for different patterns in visual search.

## Acknowledgements

This research project is partly funded by SPC Group, a multimedia company from ‘s Hertogenbosch, The Netherlands.

## References

Baddeley, A. (1992). Working Memory. Science, 255, 556-559.

Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*, 329-343.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning. *Journal of Educational Psychology, 91*, 358-368.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319-334.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 351-371.

Paas, F. G. W. C., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419-430.

Park, I., & Hannafin, M. J. (1994). Empirically-based guidelines for the design of interactive multimedia. *Educational Technology, Research & Development, 41*, 66-85.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257-287.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training.* Englewood Cliffs, NJ: Educational Technology Publications.