APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **17**: 401–418 (2003)
Published online in Wiley InterScience 26 February 2003
(www.interscience.wiley.com) DOI: 10.1002/acp.877

When Auditory Presentations Should and Should not be a Component of Multimedia Instruction

WAYNE LEAHY*, PAUL CHANDLER and JOHN SWELLER

University of New South Wales, Australia

SUMMARY

Based on cognitive load theory, two experiments investigated the conditions under which audiovisual-based instruction may be an effective or an ineffective instructional technique. Results from Experiment 1 indicated that visual with audio presentations were superior to equivalent visual-only presentations. In this experiment, neither the auditory nor the visual material could be understood in isolation. Both sources of information were interrelated and were essential to render the material intelligible. In contrast, Experiment 2 demonstrated that a non-essential explanatory text, presented aurally with similar written text contained in a diagram, hindered learning. This result was obtained because when compared to a diagram only format, the aural material was unnecessary and therefore created a *redundancy* effect. Differences between groups were stronger when information was high in complexity. It was concluded that the effectiveness of multimedia instruction depends very much on how and when auditory information is used. Copyright © 2003 John Wiley & Sons, Ltd.

With the advent of new technologies, multimedia (e.g. audio/visual) presentations have become increasingly important in the design of instruction (see Mayer *et al.*, 2001 for a recent review). The experiments in this paper were devised to examine some of the conditions under which these presentations may or may not be effective. The hypotheses were generated by cognitive load theory (Sweller, 1988, 1994, 1999; Sweller and Chandler, 1994; Sweller *et al.*, 1998). A focus for this theory, and its major concepts, is provided by the following framework:

(1) Working memory restrictions should be a major concern when designing instruction. These restrictions, in storing and processing information, have been known for some time (Miller, 1956). They apply only to new combinations of elements that must be processed, where an element is a single unit in working memory.
(2) In contrast, the capacity of long-term memory is virtually limitless (Chase and Simon, 1973; Newell and Simon, 1972) and appears to be designed to accommodate an unlimited number of elements organized in the form of hierarchically organized schemas (Bartlett, 1932; Chi *et al.*, 1982; Larkin *et al.*, 1980). Long-term memory deals with old or previously learned combinations of elements. As an example, schemas for text allow us to read and derive meaning from the complex, related squiggles of which text consists.

*Correspondence to: Dr Wayne M. Leahy, School of Education, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: w.leahy@unsw.edu.au

Contract/grant sponsor: Australian Research Council.

(3) Schemas can operate under degrees of controlled or automatic processing (e.g. Kotovsky *et al.*, 1985). Automation allows material to be processed without conscious, or working memory, control. Competent readers process individual letters and even words or phrases automatically.
(4) Schema construction and automation are crucial to learning. Automated schemas are held in long-term memory and reduce working memory load. They enable the processing of large amounts of information that would not be possible if humans had to rely entirely on a limited working memory. Consequently, the aim of instruction should be to facilitate schema construction and automation. If a learner is required to devote mental resources not directly related to the construction and automation of essential schemas, learning may be inhibited.

Many commonly used instructional procedures are inefficient because they are not structured with human cognitive architecture in mind and in particular, ignore the processing limitations of working memory. Often, learners are required to engage in cognitive activities that are unconnected to schema construction and automation. These activities consequently exert a heavy load on limited working memory. By redesigning the material so that unnecessary mental activities are removed or reduced, learning and problem solving may be enhanced.

INTRINSIC AND EXTRANEOUS COGNITIVE LOAD

Cognitive load theory also suggests that there are at least two separate sources of cognitive load: intrinsic and extraneous. Intrinsic cognitive load is determined by the intellectual demands or complexity of the learning material. Instructors have no control of intrinsic cognitive load. In contrast, extraneous cognitive load is determined by instructional design or activities required of the learner and so are under the control of instructors. Both sources of cognitive load affect working memory load. It should be noted that while increases in cognitive load increase learning difficulty, there are other forms of difficulty. These are unrelated to cognitive load because they do not affect working memory load. Sweller (1994) and Sweller and Chandler (1994) indicated that material may be difficult primarily because there may be a large amount of information that needs to be assimilated serially. Since this source is unrelated to either intrinsic or extraneous cognitive load, it is not considered in this paper.

Cognitive load theory proposes that the degree of interactivity between learning elements determines intrinsic cognitive load. An element is a learning item in its simplest form and processed as a single unit in working memory. Some elements cannot be assimilated independently of others. Consider the following example of a two-day temperature graph denoting the degrees Celsius for an hourly time (see Figure 1). The complete reading and interpreting of this type of line graph involves a high intrinsic cognitive load. A high load occurs because several learning elements interact, and need to be processed simultaneously. For example, the graph in Figure 1 consists of a vertical temperature line with eleven points from 16 degrees Celsius to 36 degrees Celsius. There is also a horizontal time line with eight points from 9 am to 4 pm. As well, there are lines indicating time and temperature for a Monday and a Tuesday. To understand a concept such as which day has the lower temperature at 3 pm, the learner has to attend to these various elements simultaneously. Here, the level of element interactivity is high and

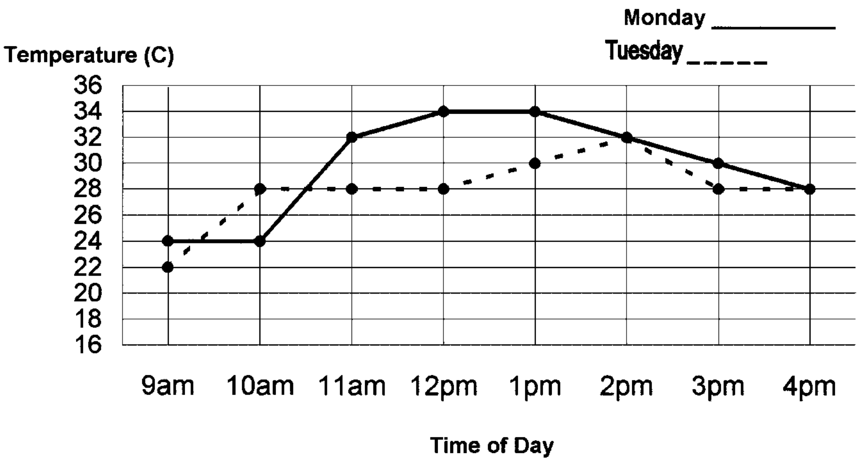

Figure 1. Temperature line graph used in Experiment 1 (modified with text box for Experiment 2)

consequently, intrinsic cognitive load is likely to be high. Intrinsic cognitive load cannot be manipulated by instructional design without affecting intelligibility (see Appendix 1a for an estimation of element interactivity for this example). In contrast, extraneous cognitive load, by definition, is entirely under instructional control. It can be altered by changing the manner in which the information is presented and the activities required of the learner.

In summary, there is a clear distinction between intrinsic and extraneous cognitive load. Both are additive and together contribute to the total cognitive load imposed by a learning task. If this total cognitive load exceeds the capacity of limited working memory resources, learning will be difficult and schema construction and automation inhibited. If learning is to be successful, total cognitive load may need to be reduced. To reduce cognitive load, instructional procedures may need to be modified.

The experience of the learner also needs consideration. One aspect that establishes whether the material to be processed is high in intrinsic cognitive load is the prior knowledge of the learner. Thus, the need to reduce extraneous cognitive load will be reliant on the learner's prior exposure and experience in the learning domain. If the student has well-established schemas for the problem's solution, those schemas will incorporate high element interactivity material within them, reducing intrinsic cognitive load. As a consequence of the overall reduction in cognitive load, presentation of the problem in a way that decreases extraneous cognitive load may make no noticeable difference to understanding. However, when a person is inexperienced in a domain, and therefore many features of the area are likely to be high in element interactivity because the interacting elements are not embedded in schemas, the structure of the material may require careful consideration. This analysis has resonance with studies on aptitude–treatment interactions (e.g. Cronbach and Snow, 1977; Mayer, 1997; Mayer and Gallini, 1990; Mayer *et al.*, 1975; Peterson, 1977; Snow *et al.*, 1980; Snow and Lohman, 1984) in which the structure of material interacts with learner aptitude.

Aptitude–treatment interactions (ATIs) develop when diverse instructional treatments result in differential learning rates depending on student aptitudes (e.g. knowledge, skills, learning styles, personality characteristics, etc.). There are constraints with this type of research with evidence of, non-exclusiveness of abilities for a particular treatment

(Cronbach and Snow, 1977), the altering of abilities as the task progresses (Burns, 1980; Federico, 1980; Snow *et al.*, 1980) and difficulty in generalizing results (Peterson, 1977). There has been a recent renewal of interest in ATI research in connection with the goal of optimizing learning by adapting to a learner's particular traits using thoroughly controlled learning environments in computer-based intelligent tutoring systems (Shute, 1993; Shute and Gluck, 1996).

Over the past 20 years, a number of instructional techniques based on cognitive load theory have demonstrated their superiority over conventional approaches to instructional design (see Sweller, 1999 for a review of instructional techniques). Three related phenomena, namely, the split-attention effect, the modality effect and the redundancy effect, which directly concern the experiments of this paper, are briefly summarized below.

The split-attention effect

Many instructional formats require learners to mentally integrate several interdependent sources of information that are unintelligible in isolation. Mental integration may lead to a heavy cognitive load, exceeding limited working memory. A geometric diagram and its associated statements provide an example. Before the statements can be understood, the diagram and its associated statements must be mentally integrated. A heavy extraneous cognitive load is imposed solely because of the structure of the instructions. Physically integrating the diagram and statements by placing the statements on the diagram negates the need for mental integration and thus reduces extraneous cognitive load. Many experiments have demonstrated the superiority of integrated over conventional worked instructions. For example, prior literature includes Chandler and Sweller (1991, 1992, 1996), Sweller *et al.* (1990); Tarmizi and Sweller (1988) and Ward and Sweller (1990). These experiments demonstrated the split-attention effect in a diverse range of areas such as geometry, numerical control programming, computing, physics, electrical installation testing and human anatomy. The results suggest that reducing or eliminating the split-attention effect may have extensive applications.

The modality effect

Information-processing models of learning have historically highlighted the limits of working memory, but there is evidence (see Penney, 1989) that under certain circumstances, working memory may be increased. Research has suggested a dual-processing model of memory architecture (Baddeley, 1992; Paivio, 1990; Penney, 1989; Tabachneck-Schijf *et al.*, 1997). This model implies that the amount of information that can be processed using both auditory and visual channels may be larger than that of a single channel. Thus, limited working memory may be effectively expanded by using more than one sensory modality, and some instructional materials with dual-mode presentation may be more efficient than equivalent single-modality formats. One important facet of this paper is concerned with an increased capacity of working memory using dual-modality instructional formats and the conditions under which they are most effective.

The above work provides a possible alternative way of dealing with split-attention instructions other than physical integration of related information. If effective working memory can be increased by providing textual content in auditory rather than written form, the cognitive load consequences of split-attention may be ameliorated in the same manner as integrating split-source information. This effect has been labelled the modality effect. Previous research in this domain has been completed by Jeung *et al.* (1997), Kalyuga *et al.* (1999); Mayer (1997); Mayer and Moreno (1998); Mayer (2001); Moreno and Mayer (1999); Mousavi *et al.* (1995) and Tindall-Ford *et al.* (1997). They provided evidence from a comprehensive set of studies in a range of learning areas that suggested the modality effect may have general applications.

The redundancy effect

The split-attention effect occurs when individual sources of information are unintelligible until they have been integrated. For example, the solution to a geometry problem is meaningless until it has been integrated with its diagram. In contrast, if multiple sources of information are all intelligible in isolation such as, for example, textual material that recapitulates the content of a diagram, then only one source of instruction (in this case almost certainly the diagram) should be used. The other source, which is unnecessary or redundant, should be omitted completely from the instructional materials. Much conventional instruction has been based on the assumption that learning is enhanced if the same information is presented in more than one way. However, research evidence to the contrary has been obtained by, for example, Bobis *et al.* (1993); Chandler and Sweller (1991, 1996); Kalyuga *et al.* (1999) and Mayer *et al.* (2001).

RATIONALE OF THE EXPERIMENTS

The experiments of this paper were intended to test the hypothesis that audio with visual (dual modality) instruction can have either a positive or negative effect on learning depending on relations between the visual and auditory components of the instructional material. If both auditory and visual sections are essential for understanding, then an audio with visual presentation should be superior to a visual-only presentation. This hypothesis was tested in Experiment 1. Alternatively, if the auditory and visual components conformed to a redundancy paradigm with the auditory component merely recapitulating and adding no new information to that provided in the visual component, the audio-visual presentation should be inferior to a visual-only presentation. The second hypothesis was tested in Experiment 2. Since both the modality and redundancy effects are cognitive load phenomena, it was further predicted that the effects would be more pronounced using high rather than low element interactivity material. The instructional material used in the experiments involved the interpreting of a simultaneous (two-day line) temperature graph. While the various effects tested in these experiments are all well established, the critical interactions between the effects are not. These interactions provide the major rationale for the experiments. In addition, the effects have not previously been demonstrated with the younger children used in the current experiments.

EXPERIMENT 1

The aim of Experiment 1 was to compare two types of instructional presentations in which the modality of instruction was varied. The first group was a conventional visual-only group where line graphs and related text were presented in a visual mode. The second group instructions were identical to the first group except all text was presented aurally via an audiotape player rather than visually. Neither the line graphs nor the text were intelligible in isolation. The instructions could not be understood without mentally integrating both sources of information, an essential prerequisite to obtaining the modality effect.

In accordance with cognitive load theory (Sweller 1994, 1999; Sweller and Chandler, 1994) and an interpretation of the modality effect, it was hypothesized that the audio/visual group would outperform the visual-only group. This assertion is based on the notion that working memory can be effectively expanded if a mix of sensory modes is used rather than a single mode of information. Thus, while both groups had to split their attention between text (whether it was audio or visual) and diagram, it was expected that the audio/visual group would have effectively more working memory available to relate, coordinate and learn critical map concepts. In other words, this new (to the student) combination of elements (discrete pieces of information needed to interpret the graph) must be processed in a limited working memory. If working memory can be effectively expanded by an audio/visual mode of instructional delivery, that expansion should allow an increase in the number of elements that can be retained and processed simultaneously, thus facilitating schema development and automation.

There was also another consideration. The instructional materials and test questions used could be divided into two sections. One section consisted of instructions/questions tapping knowledge requiring the participant to consider a higher number of interacting elements (a higher intrinsic cognitive load/higher complexity) to achieve a correct concept/answer. The other section of test questions tapped knowledge containing a lower number of interacting elements (a lower intrinsic cognitive load/low complexity). It was predicted that the largest differences between the two instructional groups would be on questions involving high rather than low element interactivity. Thus, it was predicted that the expanded working memory capability of the audio/visual group (compared to the visual only group) would result in a better understanding of the material because of the audio/visual group's more effective development and automation of schema. The consequence should be a greater expertise in the answering of the more complex (high element interactivity) questions.

Method

Participants

The participants were 48 Year Five (in their sixth year of elementary schooling and of age 10 to 11 years) students from a metropolitan primary school in New South Wales. These students had limited experience in interpreting line graphs. They were divided into two groups of 24 participants each.

Materials

There were two sets of instructional material for Experiment 1, visual-only and audio/visual. Both sets of instructional material consisted of a temperature line graph for a Monday and a Tuesday. Figure 1 provides an example of the diagrammatic material used in both sets of instructions. For the visual-only group, the diagram of Figure 1 was accompanied by written instructions (on the same page but below the diagram) which described various aspects of the line graph. The written instructions consisted of ten points which taught students to locate times and temperatures for any two days of the week. From these instructions they were shown how to establish the average rate of change during a set time period. The written instructions also showed students how to compare the average rates of change. The audio/visual instructions did not include the written instructions but instead consisted of an identical 185 second audio tape of the written instructions.

Procedure

The experiment was conducted in two stages, instruction and testing. Each participant was tested individually. Before the instruction stage of the experiment, participants were informed that they were going to be given a short lesson on how to read a type of temperature line graph. The participants were then randomly allocated to one of the two instructional groups. The visual-only group was issued with the graph plus the written instructions. During the instruction stage, those in the visual-only group were told to read the instructions while referring to the graph. They were allowed to commence reading immediately. No time limit was specified and they were directed to inform the instructor when they were finished.

The audio/visual group was given the graph only and listened to the identical text instructions via cassette. The cassette was activated by the instructor and they were allowed to commence looking at the diagram immediately upon the starting of the tape. Their time limit was set entirely by the length of the instructions, which was 185 seconds. The recording of the instructions was played only once.

For the test phase, each group received a graph similar to the one used during the instruction phase and 14 questions relating to the graph (see Appendix 1c). These could be divided into the high and low element interactivity categories. The estimations of element interactivity were derived from a procedure used by Sweller and Chandler (1994). The authors suggested that element interactivity could be estimated by establishing what constitutes a learning element for a particular group of participants and then count the number of interacting elements that are required to understand or perform a task.

Questions tapping high element interactivity knowledge. Several questions tapped knowledge consisting of six or more interacting elements. Questions 1 to 5 involved finding the average rate of change between two differing times. Using the element interactivity estimate measure established by Sweller and Chandler (1994) the number of interacting elements of the knowledge tapped by these questions was estimated to be eight, composed of: (1) Locate correct day, (2) Locate the first time period, (3) Locate the temperature for this time, (4) Locate the second time period, (5) Locate the temperature for this time, (6) Subtract the lowest temperature from the highest temperature, (7) Subtract earlier time from later time (or this can be completed by counting the number of hours) and (8) Divide the answer for (5) by answer for (6).

Questions 9 to 10 consisted of finding the highest rate of change over one hour for each day. Again using the Chandler and Sweller (1994) estimate method, the number of interacting elements of the knowledge tapped by these questions was seven made up of (1) Locate correct day, (2)–(6) Find the steepest slope from the five one-hour periods and, (7) Subtract the lowest from the highest temperature of this slope.

Question 11 required students to recognize a zero average rate of change for a single day. The number of interacting elements of the knowledge tapped by this question was estimated to be six, composed of: (1) Locating the correct day and (2)–(6) Comparing five time periods.

Question 12 involved recognizing the higher average rate of change over a one-hour period between two days. The number of interacting elements of the knowledge tapped by this question was estimated to be eight, composed of: (1) Locating a one-hour period, (2)–(4) Subtracting the temperatures for one day, (5)–(7) Subtracting temperatures for the other day and (8) Choosing the higher figure.

Questions tapping low element interactivity knowledge. Several questions tapped knowledge consisting of three or fewer interacting elements. Question 6 required students to indicate which of two days has the highest rate of temperature change. The number of interacting elements of the knowledge tapped by this question is estimated to be only two, as most of the steps have been completed when answering Q4 and Q5. Questions 7 and 8 were similar and concerned with finding which day had the lowest/highest average rate of change between two time periods. The number of interacting elements of the knowledge tapped by these questions is estimated to be only one step if students looked at the slopes of the lines.

Question 13 required the students to recognize a zero average rate of change by looking at the graph. The element count here was (1) Looking for a straight line.

Question 14 required learners to find periods when the average rate of change was the same. The number of interacting elements of the knowledge tapped by this question is estimated to be three, consisting of: (1) Examine example given in question, (2) Finding identical slopes and (3) Note time period.

Clearly, the measures of element interactivity are likely to be only approximations. However, we did not believe this posed a serious problem for this particular study as there were very distinct differences between the element interactivity counts of the two groups. For instance, the highest count for the low element interactivity questions (three) was considerably lower than the lowest count from the high element interactivity questions (six).

Each correct answer was allocated one mark and no marks were given for an incorrect answer and therefore, the test was marked out of 14. No time limit was specified for the testing phase and students could go back to check their answers.

Results and discussion

The variables under analysis were test scores, broken down into scores from responses to questions tapping knowledge estimated to be higher in element interactivity and scores from responses to questions tapping knowledge estimated to be lower in element interactivity. (Raw scores were converted to percentages as there was an uneven number of questions in each category.)

A 2 (instructional groups) by 2 (levels of element interactivity) ANOVA with repeated measures on the second factor was conducted (see Table 1). There was no significant

Table 1. Comparison of mean percentage test scores correct for the 9 high element questions and 5 low element questions in Experiment 1 (standard deviations in italics)

Group	Level of element interactivity	
	Low	High
Audio/visual	73.33	34.26
	24.79	*28.73*
Visual only	74.17	15.74
	21.65	*20.17*

Table 2. The number of students (out of 24) correct on each of the 14 questions and an estimate of the number of interacting elements needed to answer each question in Experiment 1

Element count	8*	8*	8	8*	8*	2	1	1	7	7	6	8*	1	3
Ques. nos.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Audio/visual group	8	9	4	8	9	22	18	21	9	8	9	10	16	11
Visual-only group	1	2	1	3	3	20	19	23	7	7	7	3	12	15

*Denotes significant difference according to Fisher Exact tests.

difference between the groups, $F(1, 46) = 2.29$, $MSe = 818.47$, $p = 0.137$. There was a significant effect between the two levels of element interactivity, $F(1, 46) = 168.04$, $MSe = 339.46$, $p < 0.0001$, $\omega^2 = 78.50\%$ (indicating the magnitude of the effect), with questions tapping high element interactivity material being harder than those tapping low element interactivity material. The main interest of this experiment was the interaction effects between the groups and the two levels of element interactivity. There was a significant group by element interactivity interaction, $F(1, 46) = 6.62$, $MSe = 339.46$, $p = 0.013$, $\omega^2 = 12.57\%$. Inspection of the means indicates that this effect was caused by a large difference between the high element interactivity means with almost no difference between the low element interactivity means.

Because of the interaction effect, individual *t*-tests were carried out to test for simple effects. The result from the high interactivity test material showed a significant difference between the groups, $t(46) = 2.58$, $p = 0.014$, $\omega^2 = 12.67\%$. In contrast, no significant difference was found using the low interactivity test material, $t(46) = 0.124$, $p = 0.902$, (see Table 1). Fisher Exact tests were also conducted on each question. There was a significant difference between groups in the number of students correctly answering Questions 1, 2, 4, 5 and 12. These all favoured the audio/visual group and tapped knowledge with the higher element count estimate of 8. No other questions yielded significant effects (see Table 2).

It was hypothesized that for the test stage, the audio/visual group would outperform the visual-only group, especially when dealing with complex material high in element interactivity. Results supported this hypothesis. Those results can be explained by assuming that the high element interactivity instructional material (see Appendix 1a for element count estimation) imposed a heavy intrinsic cognitive load that, when added to a high extraneous cognitive load, caused by split-attention between the graph and its written explanation, interfered with learning. In contrast, if the high extraneous cognitive load was reduced by using an audio/visual (dual-mode) presentation, the resultant increase in working memory capacity permitted the information to be more readily assimilated. Thus, Experiment 1 clearly displayed the superiority of audio/visual instructions.

Experiment 2 was designed to test whether audio/visual presentations are always similarly beneficial to learning. As discussed in the introduction to this paper, there are grounds for assuming that under some conditions, audio/visual presentation may have negative rather than positive effects. That possibility was tested in Experiment 2.

EXPERIMENT 2

In Experiment 1, it was shown that instructional material presented in an audio/visual format was superior to a conventional, visual-only format under conditions of high element interactivity but not low element interactivity. The line graphs used in Experiment 1 were not self-contained, that is, self-explanatory as diagrams. Additional text-based instructions were essential for the instructional material to be understood. Therefore in isolation, the graphs were insufficient. For this type of instructional content, an audio with visual (dual-mode) presentation was an advantage when compared to a visual-only presentation, thus demonstrating a modality effect.

However, some conditions could result in dual-mode presentations being ineffective. If, for example, in contrast to Experiment 1, a diagram is self-contained in the sense that it is intelligible without further information, the addition of auditory information (replicating as closely as possible the diagram's written text) may be redundant. Processing the auditory information may impose an additional, extraneous working memory load that reduces, rather than facilitates, learning because the auditory component just recapitulates the textual explanation associated within the diagram. Based on cognitive load theory, overloading an already limited working memory cannot facilitate effective learning (schema acquisition and automation). Thus, the hypothesis that audio instructions plus a self-explanatory diagram is an inefficient format of instructional delivery was tested in Experiment 2.

The instructional material was similar to that used in Experiment 1. A temperature line graph (diagram) of the two-day type was modified to make it self-explanatory/self-contained by having a small section of text embedded into the diagram. Thus it was a diagram with words placed near the corresponding portion of the diagram. This figure is identical to that used in Experiment 1 except that text within a box was included. The text replicated as closely as possible the information/concept contained on the audio tape. Thus, a words with diagram only (words/diagram-only) group and a words with diagram plus audio instructions (words/diagram/audio) group were compared.

It was hypothesized that the students in the words/diagram-only group would benefit more with this presentation format than the words/diagram/audio group. The rationale was that the second group had to continuously attend to the potentially redundant audio explanations when the diagram itself was already self-explanatory. Attending to the redundant auditory explanations simultaneously with inspection of the diagram could impose an additional cognitive load thus decreasing the performance and efficiency of the instructional presentation.

This experiment was similar to Experiment 1 in relation to the element interactivity required to correctly answer the test questions. As in Experiment 1, the instructional materials and questions were of the type that tapped high or low element interactivity knowledge and could be divided and classified as such. For the test questions, it was predicted that the difference between groups would be higher on those questions tapping high rather than low element interactivity material.

Method

Participants

The participants were 30 Year Five students (in their sixth year of elementary schooling and of age 10 to 11 years) from a metropolitan primary school in New South Wales, Australia. The participants had limited experience with reading line graphs. They were divided into two groups of 15 participants each.

Materials

The materials consisted of two line graphs. The first graph was a single-line graph for temperatures during a Monday (similar to Figure 1 but with only a single line denoting one day). The second graph was a dual (simultaneous) line-graph for temperatures during a Monday and a Tuesday. This graph was similar to Figure 1 but modified to be self-explanatory for the purpose of the experiment. The modifications consisted of a text box with the key concept written within it. The text and diagrams were presented in an integrated format to eliminate split-attention. Additional material consisted of an 85 second pre-instruction audio tape to be used with the single-line graph. This material was in the form of ten points. Those points taught students to locate a time and temperature for a particular day. There was also an audio tape of 135 seconds replicating as closely as possible the contained information supplied in the dual-line graph. It was also in a similar format to the pre-instruction tape. It consisted of 12 points designed to teach students to locate times and temperatures for the two days and establish differences in temperatures (see Appendix 2).

Procedure

The students were randomly divided into two groups, a words/diagram-only group and a words/diagram/audio group. The subjects were tested individually. The experiment was conducted in three stages: a pre-instruction stage, an instruction stage and a test stage. During stage one, all participants listened to the 85 second pre-instruction tape on reading a single-line graph while viewing the single-line graph. This tape was played twice. The purpose of this process was to ensure that all students were familiar with a basic concept involved with reading temperature line graphs. This procedure was not used in the previous experiment because Experiment 1 was conducted later in the school year. It was established that the students used in Experiment 1 had at least some previous instruction on graph formats. In contrast, students in Experiment 2 had not received prior equivalent instruction.

Stage two contained new information. During stage two, the words/diagram-only group was instructed to study the instructional material of the dual-line graph and to indicate when they were confident they understood it. Their time in seconds was recorded. The words/diagram/audio group was instructed to study the same diagram while attending to the audio instructions simultaneously. Their instruction time was limited to the tape duration of 135 seconds and the tape was played once only.

During stage three, the testing stage, each student was issued with a dual-line graph similar to the second graph used in the instructions of stage two but without the integrated text. A total of 14 questions was given. Each question was given equal weighting and students were allocated 1 mark for each correct answer. No marks were given for an incorrect answer. Where a question required two responses, and only one response was given, no half marks were allocated. Therefore the test was marked out of a total of 14. These questions could be divided into two categories. Questions were divided into those tapping high and low element interactivity knowledge. High element interactivity questions had element interactivity counts of 7 or 8 while low element interactivity questions had counts of 1, 2 or 3. (Element count estimations for the test questions were derived from the same procedure used in Experiment 1.) No time limit was specified for the testing phase and the students could go back to clarify their answers.

Questions tapping high element interactivity knowledge. Questions 3, 4, 6, 7 and 8 required the finding of temperature differences between the two days at a given time. The number of interacting elements tapped by these questions was seven: (1) Locate one day, (2) Locate time, (3) Locate temperature, (4)–(6). Follow same steps for other day and (7) Subtract lowest from highest temperature.

Questions 13 and 14 involved finding two time periods for (1) the highest and (2) the lowest temperature differences between Monday and Tuesday. The number of interacting elements tapped by the two questions was 8. These were (1) Study graph for first time period with highest/lowest temperature difference or line slope between the days, (2) Locate temperatures for both days, (3) Subtract lowest from highest temperature, (4) Note time, (5) Study graph for second time period with highest/lowest temperature difference or line slope between the days, (6) Locate temperatures for both days, (7) Subtract lowest from highest temperature and (8) Note time.

Questions tapping low element interactivity knowledge. Questions 1 and 2 involved finding temperatures for a single day at a given time. The number of interacting elements tapped by the two questions was 3. These were (1) Locate correct day, (2) Locate time and (3) Locate temperature.

Question 5 required students to find which day had the lowest temperature. The number of interacting elements tapped by this question was one, the selection of the correct day from the answer to the preceding question that required students to find temperature differences between two days. Question 9 required students to find the day with the lowest temperature at a fixed time. The number of interacting elements tapped by this question was two, and could be completed by comparing the two days. There were three interacting elements tapped by Questions 10 and 11 which required students to find the temperature differences between two days at a given time. The elements were (1) Locate highest temperature, (2) Locate lowest temperature, (3) Subtract. Question 12 required students to find the day with the lowest temperature at a fixed time. The number of interacting elements tapped by this question was two. This question could be completed by comparing the two days.

Results and discussion

Instruction time between the two groups favoured the words/diagram-only group. The words/diagram-only group mean was only 62.3 seconds compared to the audio tape length of 135 seconds used by the words/diagram/audio group. The variable under analysis was overall test scores (see Table 3). A 2 (instructional groups) by 2 (levels of element interactivity) ANOVA with repeated measures on the last factor indicated a significant presentation effect favouring the words/diagram only group, $F(1, 28) = 9.18$, $MSe = 222.41$, $p = 0.005$, $\omega^2 = 24.68\%$. There was a significant difference between the two

Table 3. Comparison of mean percentage test scores correct for the 7 high element questions and 7 low element questions from Experiment 2 (standard deviations in italics)

Group	Level of element interactivity	
	Low	High
Words/diagram only	47.14	37.14
	5.26	*12.13*
Words/diagram/audio	40.48	20.48
	9.98	*17.47*

Table 4. The number of students (out of 15) correct on each of the 14 questions from groups and an estimate of the number of interacting elements needed to be considered for each question in Experiment 2

Element count	3	3	7*	7	1	7	7	7*	2	3	3	2	8*	8*
Ques. nos.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Words/diagram	15	15	14	11	14	10	13	15	15	14	13	13	7	8
Words/diagram/audio	14	14	9	8	12	6	9	7	14	10	12	10	2	1

*Denotes significant difference according to Fisher Exact tests.

levels of interactivity, $F(1, 28) = 50.06$, $MSe = 67.42$, $p < 0.0001$, $\omega^2 = 64.14\%$. There was also a significant interaction, $F(1, 28) = 5.56$, $MSe = 67.42$, $p = 0.026$, $\omega^2 = 19.81\%$. Inspection of the means indicates that this interaction was caused by a larger difference between the high than the low element interactivity questions.

Simple effects analyses indicated there was a significant difference between the high element interactivity questions favouring the words/diagram-only group, $t(28) = 3.03$, $p = 0.006$, $\omega^2 = 24.75\%$. There was also a significant difference between the low element interactivity questions, $t(28) = 2.29$, $p = 0.030$, $\omega^2 = 15.75\%$. While both simple effects analyses were significant, the significant Group $\times$ Levels of Element Interactivity effect and the difference in effect sizes indicates that the difference between groups on the high element interactivity questions was larger than the difference between groups on the low element interactivity questions.

Fisher Exact tests on the number of students who responded correctly were also completed for each test question (see Table 4). The results indicated a significant difference in favour of the words/diagram-only group for Questions 3, 8, 13 and 14. These questions tapped knowledge with higher element interactivity count estimations. It also may be noted that the number of questions correct for the words/diagram-only group compared to the words/diagram/audio group was higher on every one of the 14 questions. There were no other significant effects.

These results clearly demonstrate the superiority of the words/diagram-only format over the words/diagram/audio (an audio/visual) format, when the audio component is not essential for understanding and therefore redundant. The first group, despite spending less than half the time studying the instructions than the second group, performed at a superior level demonstrating a clear redundancy effect.

GENERAL DISCUSSION

The two experiments conducted for this paper were generated by some of the hypotheses that flow from cognitive load theory. The theory proposes that instructional consequences originate from the limitations of human cognitive capacity. We can only process a very small amount of new information at one time in working memory. It follows that information presented to learners should be designed in such a way as to reduce any avoidable load upon working memory. Three instructional techniques derived from these suggestions are: the split-attention effect, which occurs when learners must split their attention between disparate sources of essential information and mentally integrate that information to derive meaning; the modality effect which ameliorates the consequences of split-attention by presenting written text in auditory form instead; and the redundancy effect which occurs when learners must process information that is redundant rather than essential. (While all these effects are based on assumptions concerning human cognitive architecture, including a limited working memory and a large long-term memory holding large numbers of automated schemas, the testing of these assumptions was beyond the scope of this paper.)

Experiment 1 tested the modality effect by using instructional presentations for graph reading. Neither the graph nor the associated instructions were intelligible in isolation. Both were required for the instructional material to be rendered intelligible. Two instructional groups were used: a visual-only group (graph and written instructions presented in split-attention format) and an audio/visual group (graph and identical taped instructions). Results from this experiment confirmed the advantage of dual-mode presentations to overcome the problems associated with attending to multiple sources of information (the audio/visual group outperformed the visual-only group). This effect occurred only when dealing with high element interactivity information.

As an alternative to the modality effect of Experiment 1, if auditory explanations are used concurrently with, for example, a diagram, which contains sufficient information to be understood alone, the dual-mode duplication of information is redundant and may hinder learning. This design might increase the risk of working memory overload and have a detrimental effect on learning. Removal of redundant sources of information might be beneficial for learning. Experiment 2 tested this premise. The instructional modes compared were (1) written text near a diagram format and (2) the same instructions with audio instructions added. The results demonstrated that the inclusion of audio text presented simultaneously with a self-explanatory diagram (as in the case of the words/diagram/audio group) was not an efficient format for learning. The words/diagram-only group outperformed the words/diagram/audio group with the effect being larger for high element interactivity material compared to low element interactivity material. These results occurred under a condition where the words/diagram/audio group had significantly more instruction time than the words/diagram-only group. The contrasting results of Experiments 1 and 2 using similar materials provides the major finding of this paper. Furthermore, the fact that the findings are theoretically coherent and in accord with other findings suggests that they are stable and replicable. Obtaining significant effects using small sample sizes is difficult and only possible with very large effects.

From an instructional design perspective, the results of this paper highlight the dangers of an uncritical acceptance of multimedia (audio with visual) presentations. There is now considerable evidence that clearly demonstrates that the inclusion of auditory instructions within multimedia instruction can be an effective educational tool (Jeung *et al.*, 1997; Kalyuga *et al.*, 1999; Mayer, 1997; Mayer and Moreno, 1998; Moreno and Mayer, 1999; Mousavi *et al.*, 1995; Tindall-Ford *et al.*, 1997). However, this paper has also shown that depending on conditions, the inclusion of auditory instructions as part of multimedia presentations can have strong negative learning outcomes.

In summary, determining which approach to use should not be haphazard. Nonetheless, while not simple, rules to guide instructional designers are available. To understand those rules requires an understanding of human cognitive structures and the instructional design principles that flow from them. A failure to understand cognitive architecture and the instructional design principles that emerge can result in procedures that are random in their effectiveness.

ACKNOWLEDGEMENTS

The work reported in this paper was supported by the Australian Research Council. The authors wish to thank the following people for their time and valuable assistance; the Principal, Sr Veronica Powell and the staff and students of Holy Cross Primary School Woollahra, John O'Grady, Sue Lawrence and the students of St Therese Primary School Mascot.

REFERENCES

Baddeley A. 1992. Working memory. *Science* **255**: 556–559.

Bartlett FC. 1932. *Remembering: A Study in Experimental and Social Psychology.* Cambridge University Press: London.

Bobis J, Sweller J, Cooper M. 1993. Cognitive load effects in a primary school geometry task. *Learning and Instruction* **3**: 1–21.

Burns RB. 1980. Relation of aptitudes to learning at different points in time during instruction. *Journal of Educational Psychology* **72**(6): 785–795.

Chandler P, Sweller J. 1991. Cognitive load theory and the format of instruction. *Cognition and Instruction* **8**: 293–332.

Chandler P, Sweller J. 1992. The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology* **62**: 233–246.

Chandler P, Sweller J. 1996. Cognitive load while learning to use a computer program. *Applied Cognitive Psychology* **10**: 1–20.

Chase WG, Simon HA. 1973. Perception in chess. *Cognitive Psychology* **4**: 55–81.

Chi M, Glaser R, Rees E. 1982. Expertise in problem solving. In *Advances in the Psychology of Human Intelligence*, Sternberg R (ed.). Erlbaum: Hillsdale, NJ; 7–75.

Cronbach LJ, Snow RE. 1977. *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*. Irvington: New York.

Federico P. 1983. Changes in the cognitive components of achievement as students proceed through computer-managed instruction. *Journal of Computer-Based Instruction* **9**(4): 156–168.

Fisher RA. 1970. *Statistical Methods for Research Workers*. Hafner Publishing Company: New York.

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 multi-media instruction. *Applied Cognitive Psychology* **13**: 351–371.

Kotovsky K, Hayes JR, Simon HA. 1985. Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology* **17**: 248–294.

Larkin J, McDermott J, Simon D, Simon H. 1980. Models of competence in solving physics problems. *Cognitive Science* **4**: 317–348.

Mayer RE. 1997. Multimedia learning: are we asking the right questions? *Educational Psychologist* **32**: 1–19.

Mayer RE. 2001. *Multimedia Learning*. Cambridge University Press: New York.

Mayer R, Gallini J. 1990. When is an illustration worth a thousand words? *Journal of Educational Psychology* **82**: 715–726.

Mayer RE, Heiser J, Lonn S. 2001. Cognitive constraints on multi-media learning: when presenting more material results in less understanding. *Journal of Educational Psychology* **93**(1): 187–198.

Mayer RE, Moreno R. 1998. A split-attention effect in multimedia learning: evidence for dual-processing systems in working memory. *Journal of Educational Psychology* **20**: 312–320.

Mayer R, Stiehl C, Greeno J. 1975. Acquisition of understanding and skill in relation to subjects preparation and meaningfulness of instruction. *Journal of Educational Psychology* **67**: 331–350.

Miller GA. 1956. The magical number of seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* **63**: 81–97.

Moreno RE, Mayer RE. 1999. Cognitive principles of multimedia learning: the role of modality and contiguity effects. *Journal of Educational Psychology* **91**: 358–368.

Mousavi SY, Low R, Sweller J. 1995. Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology* **87**: 319–334.
Newell A, Simon H. 1972. *Human Problem Solving*. Prentice Hall: Englewood Cliffs, NJ.
Paivio A. 1990. *Mental Representations: A Dual Coding Approach*. Oxford University Press: New York.
Penney CG. 1989. Modality effects and the structure of short term verbal memory. *Memory and Cognition* **17**: 398–442.
Peterson PL. 1977. Interactive effects of student anxiety, achievement orientation, and teacher behaviour on student achievement and aptitude. *Journal of Educational Psychology* **69**(6): 779–793.
Shute VJ. 1993. Aptitude–treatment interactions and cognitive skill diagnosis. In *Cognitive Approaches to Automated Instruction*, Regian WJ, Shute VJ (eds). Erlbaum: Hillsdale, NJ; 35–67.
Shute VJ, Gluck KA. 1996. Individual differences in patterns of spontaneous online tool use. *The Journal of the Learning Sciences* **5**: 329–355.
Snow R, Federico PA, Montague WE (eds). 1980. *Aptitude, Learning and Instruction: Cognitive Process Analysis*. Erlbaum: Hillsdale, NJ.
Snow R, Lohman D. 1984. Towards a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology* **76**: 347–376.
Sweller J. 1988. Cognitive load during problem solving: effects on learning. *Cognitive Science* **12**: 257–285.
Sweller J. 1994. Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction* **4**: 295–312.
Sweller J. 1999. *Instructional Design in Technical Areas*. ACER Press: Victoria, Australia.
Sweller J, Chandler P. 1994. Why some material is difficult to learn. *Cognition and Instruction* **12**: 185–233.
Sweller J, Chandler P, Tierney P, Cooper M. 1990. Cognitive Load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General* **11**: 176–192.
Sweller J, van Merrienboer J, Paas F. 1998. Cognitive architecture and instructional design *Educational Psychology Review* **10**: 251–296.
Tabachneck-Schijf H, Leonardo A, Simon H. 1997. CaMeRA: a computational model of multiple representations. *Cognitive Science* **21**: 305–350.
Tarmizi R, Sweller J. 1988. Guidance during mathematical problem solving. *Journal of Educational Psychology* **80**: 424–436.
Tindall-Ford S, Chandler P, Sweller J. 1997. When two sensory modes are better than one. *Journal of Experimental Psychology: Applied* **3**(4): 257–287.
Ward M, Sweller J. 1990. Structuring effective worked examples. *Cognition and Instruction* **7**(1): 1–39.

APPENDIX 1A: ESTIMATE OF ELEMENTS THAT MUST BE CONSIDERED SIMULTANEOUSLY FOR THE TEMPERATURE DIFFERENCES EXAMPLE

Finding temperature differences for set times during Monday and Tuesday equals to total of seven elements e.g. 1] Locate one day. 2] Locate time. 3] Locate temperature. 4–6] Follow same steps for other day. 7] Subtract lowest from highest temperature.

APPENDIX 1B: INSTRUCTIONAL MATERIAL BOTH WRITTEN AND AUDIO (TAPE LENGTH 185 SECONDS) USED IN EXPERIMENT 1

1. We are going to learn how to calculate the average rate of change in time and temperature for two different days.
2. When we are calculating the average rate of change, both time and temperature are changing (neither time or temperature are fixed).

3. The rate of change can be calculated by the following formula:

$$\text{average rate of change} = \frac{\text{change in temperature}}{\text{change in time}}$$

4. An example may help you to understand. Let us first select the time period, say between 10 am and 12 pm. Now we wish to calculate change in temperature during this time period for Monday.
5. The change in temperature is worked out by finding the temperature for the later of the two times and subtracting it from the temperature at the earlier time (Change in temperature = temperature at later time – temperature at earlier time).
6. So at 12.00 pm on Monday the temperature is 34C while at 10.00 am it is 24C. The change in temperature is therefore 34 – 24 = 10 degrees.
7. We now divide the change in temperature by the change in time. The change in time is the time difference between 10 am and 12 pm, which is 2 hours.
8. To complete the final calculation of average rate of change we simply divide change in temperature by the change in time, which is 10 divided by 2 which equals 5 degrees per hour. Five degrees per hour is the average rate of change for Monday 10 am to 12 pm. The higher the number the higher the rate of change.
9. Compare this with the average rate of change for the same time period on Tuesday. At 12 pm on Tuesday it is 32C and at 10 am it is 26C. Therefore, the change in temperature is 32 – 26 = 6 degrees.
10. Therefore, the average rate of change is 6 (change in temperature) divided by 2 (change in time) which equals 3 degrees per hour. So the rate of change for 10 am to 12 pm is greater on Monday (5 degrees per hour) than on Tuesday (3 degrees per hour).

APPENDIX 1C: TEST QUESTIONS USED IN EXPERIMENT 1

Q1 What is the average rate of change between 11 am and 1 pm on Monday?
Q2 What is the average rate of change between 11 am and 1 pm on Tuesday?
Q3 What is the average rate of change between 9 am and 2 pm on Tuesday?
Q4 What is the average rate of change for 9 am to 11 am on Monday?
Q5 What is the average rate of change for the same time period (9 am to 11 am) on Tuesday?
Q6 Which of these days has the highest average rate of change?
Q7 Which day had the lower average rate of change between 10 am and 12 pm?
Q8 Which day had the higher average rate of change between 11 am and 1 pm?
Q9 By just looking at the line graph, find the highest average rate of change over a one hour period on Monday.
Q10 By just looking at the line graph, find the highest average rate of change over a one hour period on Tuesday.
Q11 On what time on Tuesday was the average rate of change zero (o) degrees per hour?
Q12 How do you think we can recognize the higher average rate of change over a one hour period between two days?
Q13 How do you think we can recognize a zero (0) average rate of change, just by looking at the graph?
Q14 Between 10 am and 11 am on both Monday and Tuesday, the average rate of change was the same, 2 degrees per hour. In what other one hour period did this occur?

APPENDIX 2A: AUDIO INSTRUCTIONS FOR GRAPH 2 FOR WORDS/DIAGRAM/AUDIO GROUP USED IN EXPERIMENT 2 (TAPE LENGTH 135 SECONDS)

Finding temperature differences for set times during Monday and Tuesday.

1. To find the temperature differences for a set time of day, we need to follow a certain procedure and apply a general rule.
2. First we choose a set time of day, let us say 10 am.
3. Now we look straight up from 10 am to the first of the two black points.
4. This is the lower of the two points at 10 am.
5. Since this point lies on the blue line, we know the lower temperature occurs on Monday.
6. Now if we go left to the Temperature line, we can see that at this low point on Monday at 10 am the temperature was 24 degrees C.
7. Then we continue further up the 10 am line until we find the second of the two black points.
8. This is the higher of the two points at 10 am. Since this point lies on the red line, we know that the higher temperature occurs on the Tuesday.
9. Once again, follow the line left from this point to the Temperature line. This higher temperature is 28 degrees C.
10. The difference in temperature for a set time of day is the higher temperature minus the lower temperature.
11. Therefore in our example at 10 am, the highest temperature on Tuesday (red line) was 28 degrees C. The lowest temperature on Monday (blue line) was 24C.
12. So at 10 am, the temperature difference is 28C minus 24C which equals 4C.