

Cognitive load theory and the effects of transient information on the modality effect

Wayne Leahy¹ · John Sweller²

Received: 8 August 2014 / Accepted: 12 October 2015 / Published online: 23 October 2015
© Springer Science+Business Media Dordrecht 2015

Abstract Based on cognitive load theory and the *transient information effect*, this paper investigated the *modality effect* while interpreting a contour map. The length and complexity of auditory and visual text instructions were manipulated. Experiment 1 indicated that longer audio text information within a presentation was inferior to the equivalent longer visual text information demonstrating a reversal of the modality effect due to transient information imposing a heavy working memory load. However, the expected modality effect was not obtained from the equivalent shorter auditory text presentation compared to shorter visual text information. It was hypothesised that the shorter text still contained too much auditory information for working memory to readily process. Experiment 2 further decreased the shorter auditory text information which then resulted in a traditional modality effect including a modality by text length interaction in which shorter, audio-visual information was better than visual only information but longer, audio-visual information was worse than visual only information.

Keywords Educational technology · Cognitive load theory · Modality effect · Transient information effect · Instructional design · Element interactivity

Introduction

The *transient information effect* (see Leahy and Sweller 2011; Wong et al. 2012) occurs when permanent instructions such as in written form are transformed into equivalent transient information such as in spoken form resulting in a decrease in learning. This result

✉ Wayne Leahy
wayne.leahy@mq.edu.au

¹ School of Education, Macquarie University, Sydney, Australia

² School of Education, University of New South Wales, Sydney, Australia

is due to the verbal information not being retained in working memory long enough to be comprehended.

The *modality effect* (see Ginns 2005 for a meta-analysis; Mayer and Moreno 2003; Penney 1975, 1989) arises when audio information replacing written text information, referring to a map, graph, diagram or tabular information results in enhanced processing and learning.

These two effects have contrasting consequences for the presentation of text in spoken form. In the case of the transient information effect, the conversion of written into spoken text can be adverse if the text is long. In the case of the modality effect the same conversion has positive consequences. The conditions under which the two effects are obtainable were investigated in two experiments, testing cognitive load theory to generate the required hypotheses.

Cognitive load theory

Cognitive load theory (Sweller 2011, 2012) is an instructional theory based on our knowledge of human cognitive architecture. It can be summarized by five principles:

1. *Long-term memory and the information store principle.* Humans hold a vast amount of information stored in their long term memory (De Groot 1965).
2. *Schema theory and the borrowing and reorganizing principle.* Most information stored in long-term memory is obtained in schematic form (Bartlett 1932) from other people's long term memories (Sweller 2012; Sweller and Sweller 2006). The schemas (Bartlett 1932; Chi et al. 1982) are borrowed from others by imitating, reading and viewing material, or listening to others.
3. *Problem solving and the randomness as genesis principle.* If information is not available from our own or other's long term memories we must problem solve using a random generate and test for effectiveness process. Novel information is then created.
4. *Working memory and the narrow limits of change principle.* The reorganizing and generation process has the potential to produce unlimited combinations of novel information. Thus, to prevent the production of too many combinations, working memory is limited in both capacity (Miller 1956) and duration (Peterson and Peterson 1959).
5. *Long-term working memory and the environmental organizing and linking principle.* Working memory is only limited when processing novel information. In contrast, it is able to process large amounts of previously organized information transferred from long-term memory (Ericsson and Kintsch 1995).

Based on these principles, the primary purpose of instruction is to construct schemas in working memory to be held in long-term memory. Instructional designs are unlikely to be effective if they fail to result in changes in long-term memory and inefficient if they disregard the limitations of working memory.

Cognitive load theory suggests working memory load can be imposed by extraneous or intrinsic cognitive load (Sweller 2010). Extraneous cognitive load is generated by the instructional format and can be altered. For example, lengthy, complex, spoken information that cannot be adequately processed in working memory may impose an extraneous cognitive load. Intrinsic cognitive load is an inherent component of the information (e.g. the formula for a gradient ratio). It is reliant on the number of elements that, because they interact, must be managed simultaneously in working memory (Ayres 2006). Some elements do not interact with each other and can be learned independently. Learning that circular lines

on a map are termed “contours” involves low element interactivity. There are only two elements that interact; the term “contours” and the physical representations of the lines.

Contrastingly, learning a formula and its associated reference to a map is complex exercise because it is higher in element interactivity. Each word and its associated reference must be understood in relation to the other words the map symbols and indicators. Intrinsic cognitive load for given information presented to learners with a particular knowledge base is fixed and cannot be altered except by changing the information or changing the knowledge levels of learners. A learner with more knowledge in a domain will need to process fewer elements because multiple elements for a novice may be only one element for a more experienced learner.

Cognitive load theory offers a framework that acknowledges these cognitive and information structures, particularly the limitations of working memory, to provide guidelines for instructional design. The theory has been used to develop instructional procedures in a variety of educational fields (Sweller et al. 1998). The modality effect and the transient information effect were derived from cognitive load theory. These effects are outlined next.

The modality effect

Working memory is not a single structure. Research, both contemporary and over a number of decades, suggests that it is composed of multiple channels or processors. There is a visual processor for dealing with images and an auditory processor for dealing with verbal information (Penney 1975, 1989) indicating the two systems process their different forms of information with some degree of independence.

Baddeley (1992) suggested an auditory loop for processing speech and a visual-spatial sketch-pad for processing images. Written text must first be processed visually before being processed as speech by the auditory processor. Under circumstances where two sources of information must be combined in order to be understood, effective working memory capacity appears to be increased if both systems are used (e.g. visual and auditory processors) rather than only one processor (Mayer 2009; Mayer and Moreno 1998, 2003; Mayrath et al. 2011; Moreno and Mayer 1999; Penney 1975: 1989).

The modality effect occurs when audio-visual information results in superior processing to visual only information. Mayer and Moreno (2003) explain the modality effect as reducing essential processing by “off-loading” visual instructional text to the auditory processor. For example, if they refer to each other, a text and diagram cannot be understood in isolation. In order to be understood, attention must be split between an instructional text and diagram and must be mentally integrated. Mentally integrating disparate sources of information requires working memory space that is limited and may be unavailable for processing.

Alternatively, the written text instructions, rather than being presented in visual form can be presented in spoken (auditory) form. If the use of both auditory and visual processors increases working memory capacity, then replacing written text with spoken text should assist in processing.

This modality effect has been repeatedly obtained using a diverse range of material (see Ginns 2005 for a meta-analysis). Research on the modality effect and instructional design has clearly demonstrated that studying instructional materials which employ a dual format consisting of, for example, visual diagrams and auditory (spoken) text may result in more efficient processing when compared to studying a comparable visual only presentation

composed of visual diagrams and visual (written) text (Mayer 2009; Mayer and Moreno 1998, 2003; Mayrath et al. 2011; Moreno and Mayer 1999; Penney 1975: 1989).

The transient information effect

This effect occurs when written text is transformed into transient information such as spoken text with a resultant decrease in learning. It has been demonstrated previously by Leahy and Sweller (2011) and Wong et al. (2012). Brief, low element interactivity information that does not need to be learned because it aligns easily to what has already been stored in long-term memory, may be processed with little effort, even if in transient, auditory form. For example, when we are conducting a conversation or when we are following dialogue during a television/radio program, we have little difficulty in processing auditory information. In contrast, complex, lengthy instructional material may need to be presented in written rather than spoken form simply because of our working memory limits. Written material can be readily re-read so reducing cognitive load compared to transient spoken material that may be difficult or impossible to retrieve.

To demonstrate this point, consider (Grade 8) geography students learning to interpret a contour map. Students may be able to easily maintain and process in working memory the spoken statement, “The distances for this map are 1 cm equals 1000 metres” while looking at the map. They may have substantially more difficulty holding and processing a spoken statement explaining equation notation such as, “ $H1$ (that is Height 1) less $H2$ (Height 2) indicates the highest height less lowest height (of 2 points) and d (distance) equals the horizontal distance between the two points in metres”. If the statement is given in a written rather than spoken format however, learners can follow a different course of action. The main statement can be segmented into two phrases. The first phrase can be looked at while mentally integrating it with the map while ignoring the second. Finally the second phrase can be examined while mentally integrating with the map while ignoring the first.

Note that shorter sentences do not always result in lower element interactivity information nor longer ones necessarily result in higher element interactivity information. Sentence length is confounded with element interactivity but it is element interactivity, not sentence length, that is the single crucial factor. However, sentence length can be used as a proxy for element interactivity.

The present studies

Two experiments used a system based *PowerPoint* delivery where learners were given relatively technical statements accompanying a map using a 2 (short or long verbal statements) \times 2 (audio-visual or visual only presentation) experimental design. We hypothesised that a shorter set of statements was more likely to yield a traditional modality effect (beneficial) while a longer set of statements was more likely to yield no modality effect. We also hypothesised that subjective ratings of task difficulty and efficiency scores would provide evidence that the two effects were due to cognitive load factors. A modality effect should be accompanied by higher cognitive loads rating and lower efficiency scores for visual text and visual map presentations while a modality effect should result in higher ratings and efficiency scores for audio-visual presentations. Measures of cognitive load can reveal important information for cognitive load theory that is not necessarily reflected by traditional performance-based measures. Particularly, the combination of performance and cognitive load measures has been identified to constitute a reliable estimate of the mental efficiency of instructional methods (Paas et al. 2003).

Experiment 1

Experiment 1 tested the modality effect using students learning to read contour maps under conditions that increased and decreased the effect of transient, spoken information by increasing and decreasing the length of statements associated with a map. Changing the length of statements changes extraneous cognitive load. The same information is presented under long and short statement conditions resulting in an identical intrinsic cognitive load. Shorter statements were expected to yield a conventional modality effect because shorter statements should be more easily processed in working memory than longer statements when presented in spoken form. Longer statements were expected to yield no or even a reversal of the modality effect because long, complex statements may be difficult to process in working memory when presented in spoken form. Long, complex statements may be better read than listened to. Thus, in a 2 (modalities) \times 2 (length of texts) experimental design, we hypothesized an interaction. Critically, by independently measuring subjective cognitive load using subjective ratings of difficulty, we hoped to provide evidence that any learning differences were due to cognitive load rather than other factors.

Method

Participants

The participants were 71 male Grade 8 (average age 14 to 15 years) students from a Sydney private school. Six geography classes had been randomly divided into four groups a week before the experiment. All Grade 8 geography classes in the school were unstreamed in ability levels. On the day of the experiment, there were student absentees. Consequently, 14 participants were in the longer audio text instruction group, 21 in the longer visual text instruction group, 16 in the shorter audio text instruction group, and 20 in the shorter visual text instruction group. The experiment was conducted in the second term of the school year and during the first lesson periods of the day.

Materials

The materials used during the learning phase consisted of introductions to contour maps and gradients and worked examples of their interpretation. The secondary school geography curriculum for this age level requires the reading of material similar to the material used in the experiment. The material was taken and adapted from a Grade 8 geography textbook (Morrison 2004). All students were reported by the senior geography teacher to have had very limited experience of reading contour maps and calculating gradients before the experiment.

A series of systems paced *PowerPoint* slides were displayed to each group containing this content in a presentation. The experiment used a 2 (modality) \times 2 (length of texts) design resulting in four instructional groups. Note that the audio speaking pace was at a rate recommended for instructions from a slide presentation of between 120 and 150 words per minute (see Williams 1998).

All four groups were provided with an introduction of a contour map and three worked examples. The first slides introduced a simple contour map. These slides showed various basic components, for example, contour lines, elevation and the scale. Subsequent slides used worked examples to demonstrate how to find a gradient.

Worked Example 1 for the longer visual and audio text groups provided the answer to the question, “What is the gradient between points A and B?” that was written on top of the slide. The steps to solve the question were displayed on this slide within three numbered textboxes. Worked Example 2 provided the answer to a similar question. The third worked example, unlike the first two, allowed a delay time of 190 s to think about a potential answer before the answer was displayed. The question asked in this example was again a similar question requiring the calculation of a gradient between two points followed by the solution steps. Students did not have to give an answer during the 190 s solution time nor write the answer down. They were instructed to just think about what the answer may be. The presentation time of 663 s was identical for all groups.

Modality differences

To establish differences in modality, the written textbox/s information shown to the longer and shorter visual text groups was eliminated and provided in an audio format for the two longer and shorter audio text groups. Questions were also presented in audio format.

The textboxes containing written formulae were retained and given in spoken form to the audio text groups. Note that this was a necessity. The elimination of the written formulae would have been counter-productive as spoken versions only of the formulae are completely unintelligible, again due to transient information effects. Thus, the audio version of the longer visual text version depicted in Fig. 1 was identical except that text boxes 1 and 3 were eliminated and replaced by spoken text. However, textbox 2 containing

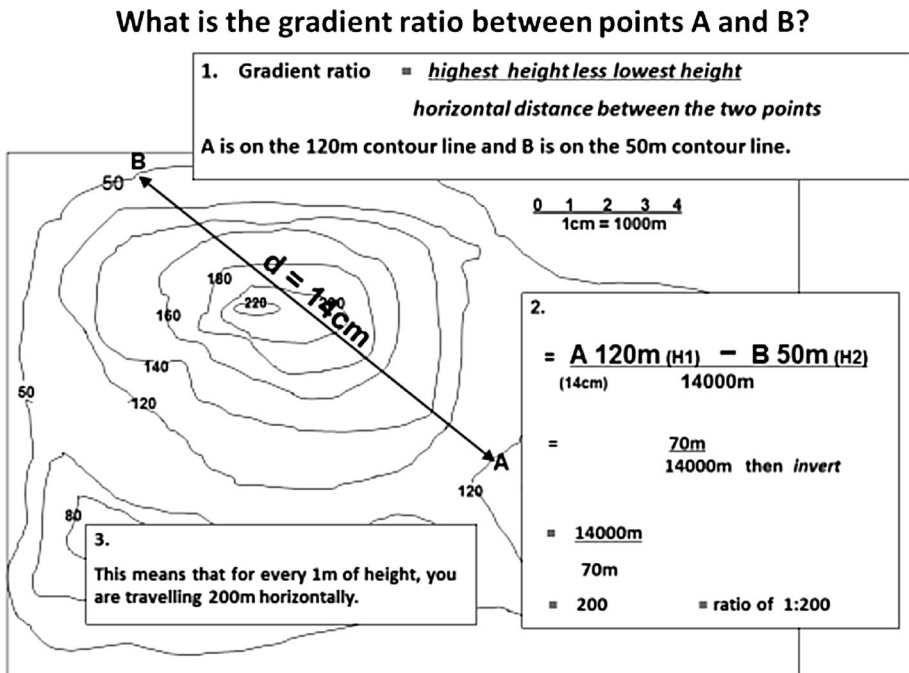


Fig. 1 Slide from Ex 1 Longer visual text group

equations was retained as well as given in spoken form. This design was applicable to all other information containing formulae.

Text length differences

The second variable was concerned with differences in length of text within slides. The longer audio-visual text and longer visual text groups had 9 slides displayed for a total duration of 663 s, timed to automatically progress from 15 to 90 s each, depending on the amount of material in the slide (see Appendix 3). (Prior to this experiment, a small pilot study was completed by the senior geography teacher to determine whether presentation times were appropriate). Presentation times were trialled as audio only and could only be approximated as appropriate for understanding for this age group. This was according to the students' verbal feedback to the senior geography teacher of the school). Equivalent slides for both the audio-visual and visual only groups were of identical total duration. The total word count of the 9 slides was 470 words giving an average of 52.22 words per content slide. Therefore, the maximum reading or listening time allowable for the content in the longer text slides was 1.41 s per word.

In contrast to the longer text groups, the shorter audio-visual text and shorter visual text groups (see Fig. 2) had 29 slides in total, timed to automatically progress from 10 to 38 s each (see Appendix 4). There were more slides needed than for the longer text groups due to the material being segmented into more sections (see Table 1 for an example of segmentation). Using another example illustrating content segmentation of instructional word text, Slide 6 (the full worked example for the longer groups) had 80 words (see Fig. 1) and Slide 17 (a segment part of the same worked example for the shorter groups) contained only 23 words (see Fig. 2).

Equivalent slides for the two shorter audio-visual and visual only text groups were of identical total presentation duration. The total word count was 436 words giving an average

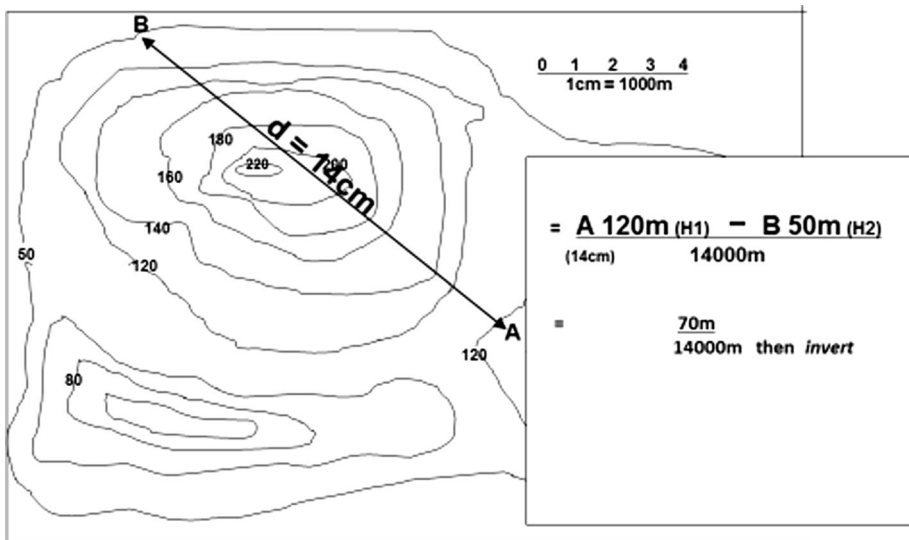


Fig. 2 Slide from Ex 1 Shorter visual text group

Table 1 The modification to group presentations for Experiment 1

Longer text groups	Shorter text groups	
Slide and text		
No. 3	No. 7	No. 8
Gradients can be calculated using contour maps. The gradient is a ratio of each metre of land covered horizontally by each vertical metre rise in the land	Gradients can be calculated using contour maps	The gradient is a ratio of each metre of land covered horizontally by each vertical metre rise in the land
Duration (to nearest whole number)		
24 s	10 s	13 s

of 15.03 words per content slide. The average listening or reading time allowable for the content in the shorter text slides was 1.52 s per word (note: compared to the 1.41 s per word for the longer text groups). The content on the slides was identical between the long and short conditions. They merely were segmented into differing numbers of slides and so differing numbers of words per slide.

The last phase, the test (see Appendix 1) comprised of ten questions designed to tap knowledge of the concepts and procedures contained in the presentation. Subjective cognitive load ratings were combined with each student's test score to provide efficiency scores (see below).

Procedure

The experiment consisted of pre-instruction, instruction, subjective student cognitive load rating and test phase. In the pre-instruction phase, all students were informed from a memorized script presented by the researcher that they were going to be taught how to read a contour map and calculate gradients by being shown worked examples contained in a *PowerPoint* presentation. They were further told that during the entire instruction phase they would have to concentrate carefully by watching the slides. The cognitive load rating scale was explained and students were encouraged to rate as accurately as possible.

After the 663 s presentation phase, the students immediately completed a 7 point cognitive load rating. Paas et al. (2003) assumes participants are able to assess their cognitive processes and to state how much mental effort they exerted (see Appendix 2 for our scale). Although not as objective e.g. a secondary task analysis or a physiological measurement, its validity has been demonstrated. According to Gopher and Braune (1984) participants can estimate with some accuracy their (perceived) mental load.

Lastly, the 20 min test phase then proceeded after the researcher distributed the test sheets. The test required an understanding of contour height indicators. It tapped knowledge from lower to higher element interactivity information. Questions 2–10 required students to process many elements simultaneously to answer correctly in contrast to Questions 1 and 2 which only involved the interaction of a few elements of height indicators. All students completed the test within the 20 min.

Results and discussion

A 2 (length of text) \times 2 (modality) ANOVA was conducted on test scores from the 10 questions (*Cronbach's Alpha* = 0.872). Means and standard deviations (the Hartley *F Max* test for homogeneity of variances was not violated) are provided in Table 2. There was no main effect for modality, $F(1,67) = 2.89$, $MSe = 9.51$, $p = 0.09$, $\eta_p^2 = 0.041$. The main effect for text length was significant, $F(1,67) = 5.34$, $MSe = 9.51$, $p = 0.02$, $\eta_p^2 = 0.079$, with shorter text groups outperforming longer text groups. The interaction was not significant, $F(1,67) = 2.67$, $MSe = 9.51$, $p = 0.10$, $\eta_p^2 = 0.038$. Nevertheless, it might be noted that there was a significant difference between the longer visual and longer audio text groups, $F(1,67) = 5.41$, $MSe = 9.51$, $p = 0.02$, $\eta_p^2 = 0.074$, with the visual group outperforming the audio group indicating the modality effect being reversed.

A 2 (length of text) \times 2 (modality) ANOVA was conducted on the subjective cognitive load ratings out of 7. Means and standard deviations are provided in Table 2. The ratings indicated a significant difference between the longer and shorter text groups $F(1,67) = 5.65$, $MSe = 2.87$, $p = 0.02$, $\eta_p^2 = 0.077$, with longer texts reported to be more difficult for students than shorter texts. There was no main effect for modality $F(1,67) = 1.82$, $MSe = 2.87$, $p = 0.18$, $\eta_p^2 = 0.026$. There was a significant text length by modality interaction, $F(1,67) = 8.12$, $MSe = 2.87$, $p = 0.006$, $\eta_p^2 = 0.108$. Because of the significant interaction, simple effects contrasts were carried out. There was a significant difference between the longer audio-visual text and the longer visual only text groups suggesting that the first group found the task more difficult than the latter, $F(1,67) = 8.57$, $MSe = 2.87$, $p = 0.005$, $\eta_p^2 = 0.121$. There was however no significant difference between the shorter audio-visual text and the shorter visual only text groups $F(1,67) = 1.16$, $MSe = 2.87$, $p = 0.28$, $\eta_p^2 = 0.016$.

Instructional efficiency calculations were conducted on test scores and subjective cognitive load ratings for each student (see Paas et al. 2003) using the equation:

Table 2 Means and standard deviations (in parentheses) of scores

	Test scores (out of 10)	
	Experiment 1	Experiment 2
Audio-visual shorter text group	5.55 (3.56)	7.53 (2.38)
Visual only shorter text group	5.50 (3.20)	5.52 (2.81)
Audio-visual longer text group	2.57 (2.82)	4.08 (2.82)
Visual only longer text group	5.05 (2.63)	5.18 (2.63)
Rating out of 7 (1 = very easy 7 = very difficult)		
Audio-visual shorter text group	3.43 (1.50)	2.10 (1.37)
Visual only shorter text group	4.05 (1.39)	2.68 (1.67)
Audio-visual longer text group	5.57 (1.74)	3.12 (1.78)
Visual only longer text group	3.85 (2.03)	3.40 (2.08)
Efficiency ratings (lower score = lower efficiency)		
Audio-visual shorter text group	0.41 (1.31)	0.72 (1.10)
Visual only shorter text group	0.21 (1.16)	0.01 (1.34)
Audio-visual longer text group	-1.02 (1.15)	-0.51 (1.33)
Visual only longer text group	0.17 (1.25)	-0.36 (1.46)

Likert subjective cognitive load ratings and efficiency in Experiments 1 and 2

$$E = (Z \text{ score} - Z \text{ rating})/\sqrt{2}$$

where E is the efficiency, Z score is the individual score converted to a z score and Z rating is the individual rating associated with the individual's score converted to a Z rating

From this calculation, if a student invests very little mental load and achieves a high test score, efficiency is considered high. If s/he scores low on the test and high on mental load, efficiency is low. A 2 (length of text) \times 2 (modality) ANOVA was then conducted on these data. Means and standard deviations are provided in Table 2.

The results indicated a significant difference between the longer and shorter text groups favoring the shorter text groups, $F(1,67) = 6.31$, $MSe = 1.49$, $p = 0.01$, $\eta_p^2 = 0.08$. There was no main effect for modality, $F(1,67) = 2.90$, $MSe = 1.49$, $p = 0.09$, $\eta_p^2 = 0.00$, but there was a text length by modality interaction, $F(1,67) = 5.70$, $MSe = 1.49$, $p = 0.02$, $\eta_p^2 = 0.07$.

Because of the interaction, simple effects contrasts were carried out. There was a significant difference between the longer visual and longer audio group favoring the longer visual group $F(1,67) = 8.14$, $MSe = 1.49$, $p < 0.01$, $\eta_p^2 = 0.10$. There was no significant difference between the shorter visual and audio groups, $F(1,67) = 0.24$, $MSe = 1.49$, $p = 0.62$, $\eta_p^2 = 0.002$.

These results indicated a possible modality effect reversal when using long, complex textual statements. Cognitive load and efficiency measures further indicated an excessive cognitive load associated with lengthy, complex auditory statements. The transient nature of such statements can be predicted to increase working memory load. Reducing the length of the statements eliminated any suggestion of a modality effect reversal.

The reduced reported cognitive load associated with a reduction in statement length and hence a reduction in difficulties associated with transience was not sufficient to generate a modality effect. Using short statements resulted in only a marginal, non-significant advantage to the audio-visual group. It can be hypothesized that the short spoken statements may not have been sufficiently brief to produce a typical modality effect and that these instructions were still too long and complex to be processed in working memory. As a consequence, Experiment 2 further decreased the textual instructions used in the shorter text groups. The longer text groups' instructional material remained unchanged.

Experiment 2

Experiment 2 was identical to Experiment 1 except that the length of the shorter text groups was reduced further. It was hypothesized that an additional reduction in text length would reduce the difficulty associated with transience and reduce cognitive load for the audio-visual groups sufficiently to permit a full modality effect with superior performance by the audio-visual groups compared to the visual only groups, rather than no difference between groups found in Experiment 1.

Method

Participants

The participants were 100 male Grade 8 students from four classes of a Sydney private school (not the same school as in Experiment 1). The ages ranged from 14 to 15 years. Similar to the participants in Experiment 1, the classes were un-streamed in ability levels

(including mathematics) and, according to the senior geography teacher, had minimal experience of reading contour maps and calculating gradients before the experiment. The experiment was conducted in the fourth term of the school year and during the first lesson periods of the day. The students were randomly assigned from the four Grade 8 geography classes into four groups. Similar to Experiment 1, on the day of the experiment a number of students from each group were unable to participate resulting in uneven groups. Thus, there were 25 participants in the longer audio text instruction group, 22 in the longer visual text instruction group, 28 in the short audio text instruction group, and 25 in the shorter visual text instruction group.

Materials

The instructional materials, subjective cognitive load ratings and test for this experiment were identical except for changes to the slide content presented to the shorter audio text and shorter visual text groups. The longer audio text and longer visual text groups' instructional material remained unchanged from Experiment 1.

The shorter text groups had 29 slides in Experiment 1. For Experiment 2 we again divided the slides into shorter segments resulting in 46 slides. An example of how the information on each slide was reduced can be seen from the change to Slide 22 from the shorter text groups in Experiment 1. This slide was displayed as one slide containing the text " $D 140 m (H1) - E 60 m (H2)/10000 m (10 cm) = 80 m/10000 m$ then invert" on the contour map. In Experiment 2 this slide was segmented into three slides reading: Slide 29 " $D 140 m (H1) - E 60 m (H2)$ "; Slide 30, " $= 80 m$ " and Slide 31, " $80 m/10000 m (10 cm)$ invert". The other slides were segmented in a comparable fashion. Note also, for intelligibility, as in Experiment 1, the written formulae segments on slides were retained for the audio groups as well as being spoken.

As much as possible slide durations were kept equivalent. For example, if a single slide with two sentences from the text content of Experiment 1 was displayed for 20 s, the Experiment 2 text was segmented into two slides each with one sentence, displaying the same textual information shown for a total of 20 s (see Figs. 3 and 4). Because of segmentation and as outlined previously, the total word count for the shorter groups changed to 604 words giving an average of 13.13 words per content slide. The average listening or reading time allowable for the content in the shorter text slides was 1.10 s per word. The total slide time presentations remained unchanged for all groups at 663 s. The same three worked examples (including 190 s for Worked Example 3) were used as in Experiment 1.

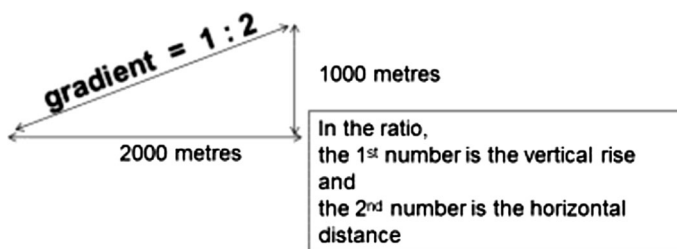
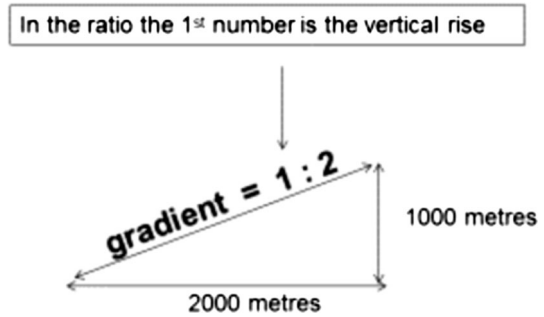


Fig. 3 Slide from Ex 1 Shorter visual text group

Fig. 4 Slide from Ex 2 Shorter visual text group



Results and discussion

A 2 (length of text) \times 2 (modality) ANOVA was conducted on the test scores. Means and standard deviations (the Hartley *F Max* test for homogeneity of variances was not violated) are provided in Table 2. The test scores indicated that there was a significant difference between the longer and shorter text groups, $F(1,96) = 12.63$, $MSe = 7.06$, $p < 0.01$, $\eta_p^2 = 0.120$ but no main effect for modality, $F(1,96) = 0.73$, $MSe = 7.06$, $p = 0.43$, $\eta_p^2 = 0.006$. There was a significant interaction, $F(1,96) = 8.53$, $MSe = 7.06$, $p = 0.006$, $\eta_p^2 = 0.077$. Following the significant interaction, simple effects contrasts were carried out. There was a significant difference between the shorter audio text and the shorter visual text groups favoring the audio text group, $F(1,96) = 7.59$, $MSe = 7.06$, $p = 0.007$, $\eta_p^2 = 0.073$, demonstrating a conventional modality effect. There was no significant difference between the longer visual text and longer audio text groups, $F(1,96) = 2.01$, $MSe = 7.06$, $p = 0.16$, $\eta_p^2 = 0.020$ although means favored the longer visual text group.

A 2 (length of text) \times 2 (modality) ANOVA was conducted on the subjective cognitive load ratings out of 7 using the same scale and questions as Experiment 1. Means and standard deviations are provided in Table 2. There was a significant difference between the shorter and longer text groups, $F(1,96) = 6.31$, $MSe = 2.98$, $p = 0.01$, $\eta_p^2 = 0.004$ indicating a reported lower cognitive load for the shorter text groups. There was no significant difference due to modality, $F(1,96) = 1.54$, $MSe = 2.98$, $p = 0.218$, $\eta_p^2 = 0.005$, nor was there a significant text length by modality interaction, $F(1,96) = 0.168$, $MSe = 2.98$, $p = 0.68$, $\eta_p^2 < 0.001$.

Instructional efficiency calculations were conducted on test scores and subjective cognitive load ratings for each student using the same model and procedure as in Experiment 1. A 2 (length of text) \times 2 (modality) ANOVA was conducted on these data. Means and standard deviations are provided in Table 1. There was a significant difference between the longer and shorter groups, $F(1,96) = 9.39$, $MSe = 1.70$, $p < 0.01$, $\eta_p^2 = 0.09$, but no significant effect due to modality, $F(1,96) = 1.12$, $MSe = 1.70$, $p < 0.29$, $\eta_p^2 = 0.01$, nor a significant interaction, $F(1,96) = 2.66$, $MSe = 1.70$, $p = 0.10$, $\eta_p^2 = 0.02$.

This experiment found a significant interaction on test scores due largely to a modality effect using shorter text lengths and to a lesser extent, due to a non-significant, modality effect reversal using longer text lengths. Subjective ratings of difficulty and efficiency measures indicated that shorter texts were described as easier than longer texts. No significant effects due to modality nor a text length by modality interaction were reported.

General discussion

The two experiments tested several hypotheses. First, we tested the hypothesis that a modality effect reversal or no modality effect would be obtained using long, high element interactivity verbal statements. When presented in spoken form, such statements were hypothesised to impose a heavy extraneous cognitive load. While a significant test length by modality interaction was not obtained, a modality effect reversal was suggested in Experiment 1 with visual only presentations obtaining higher test scores than audio-visual presentations. This result was largely due to the reduced performance of the long, audio-visual group. A similar pattern was obtained in Experiment 2 but the difference between the long, audio-visual group and the long visual only group was not significant.

Second, we tested the hypothesis that shorter, simpler statements would produce a conventional, modality effect with audio-visual statement facilitating learning compared to visual only statements. That result was obtained in Experiment 2 using very short verbal statements. Somewhat longer verbal statements in Experiment 1 did not yield a modality effect, probably because they were too long and complex.

Third, we hypothesized that this interaction between modality of presentation and statement length and complexity was due to cognitive load factors with the transient nature of long, complex, auditory statements increasing working memory load and so reversing the modality effect. Experiment 1 indicated an interaction between modality and statement length on subjective ratings of task difficulty and efficiency scores, largely due to the reported higher cognitive load ratings of the long, audio-visual group compared to the long, visual only group. Experiment 2 indicated lower cognitive load ratings by the audio-visual groups than the visual only groups. This result was expected for short statements but we expected the reverse results for longer statements to match the findings of Experiment 1. That result was not obtained.

While the results in general supported our hypotheses, we obtained no significant difference due to modality in Experiment 1 and a modality effect in Experiment 2. We had a failure of reported subjective ratings to support the performance data of Experiment 2. With respect to the inability to find a modality effect in Experiment 1, we obtained strong evidence from Experiment 2 that the length of the short statements in Experiment 1 prevented the modality effect from occurring. Reducing the length of the statements in Experiment 2 compared to Experiment 1 resulted in a conventional, modality effect.

The failure to obtain a modality effect reversal in Experiment 2 was due to the expected difference between the long audio-visual and long visual only groups not reaching significance. Using the same material in Experiment 1 indicated very poor performance by the long audio-visual group.

There were two reasons the participants of Experiment 2 may have performed relatively better in the long, audio-visual condition than the participants of Experiment 1 and so eliminating a modality effect reversal. First, Experiment 2 was run in Term 4 of a 4 term year while Experiment 1 was run in Term 2. The additional age (about half a year) and experience of the Experiment 2 participants may have allowed them to process the auditory material better than the participants of Experiment 1. In addition, the participants of Experiment 2 came from a school that tended to obtain higher academic results in national tests than the participants of Experiment 1, a fact that again may have assisted them in dealing with the transient information of the long, audio-visual text. A comparison of the test performance (Experiment 1 school mean = 4.66 and Experiment 2 school mean = 5.60) and cognitive load means (Experiment 1 school mean = 4.20 and

Experiment 2 school mean = 2.82) supports this explanation. Thus, it may be that the transient, longer auditory text for the Experiment 2 cohort was not sufficiently long or complex to overwhelm learners' working memory. We know from previous research studies in many areas of cognitive load theory that often an effect may not be obtained because the levels of expertise of the students do not match the materials used (see Kalyuga 2007 for a review).

We had expected that the modality effect using the shorter texts of Experiment 2 would be associated with supporting measures from the subjective ratings. Such support was obtained in Experiment 1 and has been obtained previously (e.g. Paas et al. 2003). While the means in Experiment 2 were in the expected direction, the differences were not significant. At this juncture, we do not have an adequate explanation for the subjective rating scales of Experiment 2 failing to accord with our hypotheses.

We have suggested that (1) the transient information effect can provide a good explanation of both failures to obtain the modality effect and explain the modality effect reversal and (2) the present data and previous work by several authors provide a better explanation of failures to obtain the modality effect than explanations based on learner or machine control of pacing.

A limitation of the studies is that we did not test the effects of manipulating learner and machine control of pacing (see Schmidt-Weigand et al. 2010; Tabbers et al. 2004). In addition, more contemporary measures of cognitive load could be used. Adding the original two longer visual instructions from Experiment 1 to the shorter versions of Experiment 2 making a six group single study may also be valuable. Such experiments would be appropriate for future research.

Conclusions

Audio-visual presentations have the potential to substantially improve instruction. Nevertheless, we should not assume that because instruction can be presented audio-visually it always should be. The current experiments indicate there are some conditions when an audio-visual presentation has negative consequences. There are other conditions when an audio-visual format has considerable benefits. Through the transient information effect, cognitive load theory can be implemented to indicate when an audio-visual format is called for and when it should be avoided.

Our results suggest that when dealing with multiple sources of information that refer to each other and cannot be understood in isolation, presenting the information in an audio-visual format is likely to be beneficial providing that verbal information is relatively simple and short. As verbal information increases in complexity and length, the advantage of audio-visual information decreases and eventually reverses. At that point verbal information should be presented in written form. Clearly, we need to be wary of using instructional technology to transform permanent information into transient information that can result in an overwhelming cognitive load that prevents understanding and learning.

Appendix 1

The 10 questions used in Ex 1 & Ex 2

USE MAP 1 TO ANSWER THE QUESTIONS ON THIS PAGE

Distances are on the line.

Q1 What is the lowest contour on the map?

Q2 What is the highest contour?

What are the gradient ratios for the lines:-

Q3 EF _____ : _____

Q4 AB _____ : _____

Q5 GK _____ : _____

Q6 CD _____ : _____

Q7 LH _____ : _____

Q8 Is the gradient ratio for S to T a. lower b. higher or c. the same as S to N?

USE MAP 2 TO COMPLETE THE FOLLOWING

Q9 Look at the line AB.

The 15cm line commences at A and ends at a lower point B.

The gradient ratio for AB is 1:150

Write down a correct lower height in metres for the contour line at B.

Q10 The line from point C to point D has a gradient ratio of?

Appendix 2

Subjective cognitive load rating scale used in Ex 1 & 2.

Rate how easy/difficult you found this presentation to understand with an X

1	2	3	4	5	6	7
Very easy	Easy	Moderately easy	Not easy nor difficult	Moderately difficult	Difficult	Very difficult

Appendix 3

For the longer audio text and longer visual text groups times in seconds (s) and number of words for each slide viewing time in Experiments 1 & 2 (to nearest whole number).

Slide 1: 60 s (74 wds), Slide 2: 15 s (16 wds), Slide 3: 24 s (26 wds), Slide 4: 34 s (39 wds), Slide 5: 70 s (55 wds), Slide 6: 90 s (75 wds), Slide 7: 90 s (75 wds), Slide 8: 190 s (10 wds—a question) and Slide 9: 90 s (75 wds—the answer).

Appendix 4

For the shorter audio text and shorter visual text groups. Times in seconds (s) and number of words for each slide viewing time- (to nearest whole number) in Experiment 1.

Slide 1: 11 s (10 wds), Slide 2: 12 s (17 wds), Slide 3: 13 s (14 wds), Slide 4: 12 s (19 wds), Slide 5: 12 s (13 wds), Slide 6: 15 s (16 wds), Slide 7: 10 s (7 wds), Slide 8: 13 s (20 wds), Slide 9: 13 s (13 wds), Slide 10: 12 s (18 wds), Slide 11: 10 s (9 wds), Slide 12: 25 s (23 wds), Slide 13: 10 s (21 wds), Slide 14: 35 s (24 wds), Slide 15: 7 s (10 wds—a question), Slide 16: 16 s (15 wds), Slide 17: 35 s (23 wds), Slide 18: 14 s (14 wds), Slide 19: 18 s (11 wds),

Slide 20: 7 s (10 wds—a question), Slide 21: 16 s (15 wds), Slide 22: 35 s (23 wds), Slide 23: 14 s (14 wds), Slide 24: 18 s (11 wds), Slide 25: 190 s (10 wds—a question), Slide 26: 18 s (15 wds), Slide 27: 38 s (23 wds), Slide 28: 16 s (14 wds) and Slide 29: 18 s (11 wds).

References

- Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*(5), 389–400. doi:10.1016/j.learninstruc.2006.09.001.
- Baddeley, A. (1992). Working memory. *Science, 255*, 556–559. doi:10.1126/science.1736359.
- Bartlett, F. C. (1932). *Remembering: a study in experimental and social psycholog*. London: Cambridge University Press.
- Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Stemberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–75). Hillsdale, NJ: Erlbaum.
- De Groot, A. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245. doi:10.1037/0033-295x.102.2.211.
- Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331. doi:10.1016/j.learninstruc.2005.07.001.
- Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26*, 519–532.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509–539. doi:10.1007/s10648-007-9054-3.
- Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*(6), 943–951. doi:10.1002/acp.1787.
- Mayer, R. E. (2009). *Multimedia learning*. New York: Cambridge University Press.
- 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, 20*, 312–320.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 3*(1), 43–53.
- Mayrath, M. C., Nihalani, P. K., & Robinson, D. H. (2011). Varying tutorial modality and interface restriction to maximize transfer in a complex simulation environment. *Journal of Educational Psychology, 103*(2), 257–268. doi:10.1037/a0022369.

- Miller, G. A. (1956). The magical number of seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. doi:[10.1037/h0043158](https://doi.org/10.1037/h0043158).
- Moreno, R. E., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity effects. *Journal of Educational Psychology*, *91*, 358–368. doi:[10.1037/0022-0663.91.2.358](https://doi.org/10.1037/0022-0663.91.2.358).
- Morrison, D. (2004). *Middle years geography*. Sydney: Blake Education.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71.
- Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin*, *82*, 68–84.
- Penney, C. G. (1989). Modality effects and the structure of short term verbal memory. *Memory and Cognition*, *17*, 398–422.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term memory retention of verbal items. *Journal of Experimental Psychology*, *58*, 193–198. doi:[10.1037/h0049234](https://doi.org/10.1037/h0049234).
- Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction*, *20*(2), 100–110. doi:[10.1016/j.learninstruc.2009.02.01](https://doi.org/10.1016/j.learninstruc.2009.02.01).
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, *22*, 123–138. doi:[10.1007/s10648-010-9128-5](https://doi.org/10.1007/s10648-010-9128-5).
- Sweller, J. (2011). Cognitive load theory. In J. Mestre & B. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (Vol. 55, pp. 37–76). Oxford: Academic Press.
- Sweller, J. (2012). Human cognitive architecture: Why some instructional procedures work and others do not. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 1, pp. 295–325). Washington, D.C.: American Psychological Association.
- Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, *4*, 434–458.
- Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.
- Tabbers, H. K., Martens, R. L., & van Merriënboer, J. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, *74*, 71–81. doi:[10.1348/000709904322848824](https://doi.org/10.1348/000709904322848824).
- Williams, J. R. (1998). Guidelines for the use of multimedia in instruction. In proceedings of the human factors and ergonomics society 42nd annual meeting (pp. 1447–1451).
- Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction*, *22*(6), 449–457. doi:[10.1016/j.chb.2008.12.012](https://doi.org/10.1016/j.chb.2008.12.012).