Original article

doi: 10.1111/jcal.12058

Journal of Computer Assisted Learning

Effects of text modality in multimedia presentations on written and oral performance

G.S.E. van den Broek, E. Segers & L. Verhoeven

Behavioural Science Institute, Radboud University Nijmegen, The Netherlands

Abstract

A common assumption in multimedia design is that audio-visual materials with pictures and spoken narrations lead to better learning outcomes than visual-only materials with pictures and on-screen text. The present study questions the generalizability of this modality effect. We explored how modality effects change over time, taking into account study strategies during learning and the modality of the final performance measure. Eighty-four university students ($M_{age} = 21.4$) studied learner-paced visual-only or audio-visual multimedia presentations and answered written and oral retention and transfer questions immediately after learning and after 1 day. There was no performance difference between the audio-visual and the visual-only groups immediately after learning, but after 1 day, the visual-only group had significantly higher scores on three of four outcome measures. This reversed modality effect was independent of test modality, but both groups scored higher on written than on oral questions. While both groups spent on average 33 min studying, the visual-only group went through the materials at a faster pace and repeated more slides. In sum, results contradict common multimedia design recommendations and instead suggest that learner-paced presentations should include on-screen text. Benefits of on-screen text could be due in part to the successful use of reading strategies.

Keywords

long-term outcomes, modality effect, multimedia design, oral versus written assessment, study behaviours in learner-paced presentations, transfer of learning.

A prominent conclusion from previous multimedia research has been that audio-visual presentations that combine pictures with spoken narrations produce better outcomes than visual-only presentations with pictures and written text (Ginns, 2005; Low & Sweller, 2005). This *modality effect* is traditionally explained with two assumptions about the architecture of human working memory: First, the visuospatial load hypothesis that learners make better use of their cognitive capacities when studying audio-visual materials that involve auditory and visual working memory subsystems than when studying visual-only materials that initially load only on visual subsystems (Mayer, 2009; term 'visuospatial load' introduced by Rummer, Schweppe, Fürstenberg, Seufert, & Brünken, 2010). Second, the complementary split-attention hypothesis

Accepted: 8 December 2013
Correspondence: Gesa van den Broek, Behavioural Science Institute, Radboud University Nijmegen, PO Box 9104, 6500 HE Nijmegen, The Netherlands. Email: g.vandenbroek@pwo.ru.nl
The materials used in this study were made by the researchers based on a CD-ROM published by Bayer Schering Pharma AG/Dr. Specht Film & Medien GmbH. The publishers of these original materials were not involved in the research.

that learners can simultaneously listen to spoken narrations and look at pictures, but can only look at either written text or the accompanying pictures at one time. For instance, eye-tracking studies have shown that learners spend more time studying visualizations in audio-visual than in visual-only presentations, where they start reading before alternating between text and visualizations (Schmidt-Weigand, Kohnert, & Glowalla, 2010a, 2010b). Because of these alternations, learners could miss part of the visual-only information, especially under time pressure, so that the integration of pictorial and verbal information could be hampered (e.g., Tabbers, Martens, & Van Merriënboer, 2004; Ayers & Sweller, 2005).

However, the generalizability of modality effects has been questioned as recent studies have drawn attention to its potential boundary factors (e.g., Stiller, Freitag, Zinnbauer, & Freitag, 2009; Lindow *et al*., 2011; Schüler, Scheiter, & Schmidt-Weigand, 2011; Tabbers & van der Spoel, 2011; Crooks, Cheon, Inan, Ari, & Flores, 2012; Reinwein, 2012). As a case in point, evidence for modality effects is stronger for system-paced presentations than for learner-paced presentations (Ginns, 2005; Stiller *et al*., 2009; Tabbers, 2002), possibly because split attention is not an issue when learners can freely determine how much time they spend studying (e.g., Tabbers, Martens, & van Merriënboer, 2001).

A further argument against modality effects is that learners can use beneficial reading strategies with printed text when given enough time (Leahy & Sweller, 2011; Schmidt-Weigand *et al*., 2010a; Tabbers *et al*., 2004). For example, learners can scan printed text for a specific piece of information that they want to selectively re-study and can flexibly adjust the reading speed to the difficulty of the information (e.g., Crooks *et al*., 2012). The same strategies are not available with auditory narrations, because these are necessarily presented at a fixed pace and in a certain order. Furthermore, auditory information is fleeting or transient, in the sense that it disappears after presentation and must therefore be maintained in working memory to be integrated with subsequently presented information (Kalyuga, 2011, 2012; Kalyuga, Chandler, & Sweller, 1999). This transiency is especially problematic when verbal information is long (Leahy & Sweller, 2011), a phenomenon that was recently termed *transient information effect* (Wong, Leahy, Marcus, & Sweller, 2012). Thus, there might be a trade-off between, on the one hand, advantages of audio-visual materials in terms of increased working memory resources and reduced split of visual attention (which predict classic modality effects) and, on the other hand, disadvantages in terms of reduced opportunities to study strategically following a self-chosen pace and order and increased working memory load due to the transiency of auditory information (which predict *reversed* modality effects). This makes it difficult to determine what the overall effect is of text modality in learner-paced presentations, especially on long-term outcomes.

Most research on multimedia design has focused on performance measures immediately or shortly after practice (Fletcher & Tobias, 2005). However, sometimes materials that make the initial acquisition of information more effortful lead to *better* long-term outcomes even if they reduce immediate outcomes, a phenomenon that has been termed *desirable difficulties* (e.g., Bjork, 1994). This is yet another reason to be sceptical about modality effects, because effort needed to integrate pictorial and verbal information during the study of visual-only materials could be desirable and lead to improved long-term outcomes. To evaluate this prediction, the long-term effects of audio-visual and visual-only materials must be measured.

The only three studies that we could find on long-term outcomes of text modality were conducted with children and suggest that after a delay, performance is indeed better after working with visual-only presentations than after working with audio-visual presentations (Segers, Verhoeven, & Hulstijn-Hendrikse, 2008; She & Chen, 2009; Witteman & Segers, 2010). Segers *et al*. (2008) found classic modality effects immediately after learning, but after 1 week, effects disappeared for retention measures (i.e., the amount of information that children could reproduce), and even reversed for transfer measures (i.e., the degree to which children were able to apply the acquired knowledge to solve new problems), so that long-term transfer performance was better for visual-only than for audio-visual materials. Witteman and Segers (2010) replicated this reversed modality effect for transfer measures after a delay of just 1 day. She and Chen (2009) studied long-term modality effects in high school students and found a reversed modality effect with retention questions 5 weeks after working with interactive learner-paced simulations. In sum, it is by no means clear whether modality effects – if found immediately after learning – persist over time and first studies with children suggest that reversed modality effects may be visible after a delay, depending on the performance measure that is used. We therefore further investigated delayed modality effects on different performance measures in the present paper.

The present study

The first aim of this study was to test if reversed modality effects on delayed tests are a developmental phenomenon or if they can be replicated with adult learners. For this purpose, adults studied learner-paced visual-only or audio-visual multimedia presentations and answered test questions immediately after learning and after 1 day. Because of the ongoing controversy about modality effects immediately after learning (Lindow *et al*., 2011), we did not formulate a directed hypothesis about the immediate outcomes, but we did expect reversed modality effects on the delayed test, based on available long-term data. We reasoned that if reversed modality effects are due to strategic reading of written information, the effects should be present for adults even more than for children, because children are less able to monitor their comprehension and read strategically than adults (Perfetti, Landi, & Oakhill, 2005). Thus, we formulated as first hypothesis that visual-only materials should produce better outcomes 1 day after learning than audio-visual materials. This prediction challenges the classic modality design principle that pictures should best be combined with spoken narrations (e.g., Mayer, 2009). Instead, we build on the argumentation that characteristics of auditory information limit the use of study strategies, such as adjusting the reading speed to the difficulty of the materials and engaging in selective rereading, and should therefore lead to worse retention. Note that while a delay of 1 day is relatively short compared with other learning studies, Witteman and Segers (2010) also found reversed modality effects after a delay of 1 day. Importantly, this delay sufficed to ensure that participants left the learning environment for a substantial amount of time and had a night of sleep before the delayed test, so that some offline memory consolidation processes during sleep could take place (e.g., Stickgold, 2005).

Second, to shed more light on processes going on during studying of on-screen text and spoken narrations, we investigated how learners worked with the two different presentations by measuring individual study times. Based on the split-attention assumption, we expected overall longer study times and longer viewing times per slide in the visual-only condition than in the audio-visual condition, due to the necessary switch between reading text and looking at the pictures (Schmidt-Weigand *et al*., 2010a, 2010b). In addition, we tested correlations between measures of study behaviour and performance to see whether and how study behaviours are related to modality effects.

Third, we investigated if long-term reversed modality effects are restricted to written questions, or whether they can also be found with oral questions. In previous studies, Segers and colleagues (Segers *et al*., 2008; Witteman & Segers, 2010) explained reversed modality effects after a delay by arguing that while on-screen text may limit visual working memory resources available for the processing of accompanying pictures (cf. Mayer, 2009), at the same time it may improve retention because it activates both auditory and visual traces in memory whereas spoken narrations leave only auditory traces. This idea is similar to theories that reading has positive effects on memory because it activates both orthography and phonology, whereas listening only activates phonology (Nelson, Balass, & Perfetti, 2005). Such double sensory processing could provide participants in the visual-only condition with more retrieval cues that facilitate later recall, in particular in response to written questions. It is questionable, though, whether the same benefits would be visible on auditory questions that do not offer visual cues. Therefore, we manipulated the test modality in the present study and included auditory questions.

An additional argument to evaluate modality effects with both written and auditory test questions is that the degree of overlap between the situation in which information is learned and the situation in which the same information is recalled affects memory performance – a principle known as transfer-appropriate processing (Roediger, Gallo, & Geraci, 2002). For example, performance is better for words that are presented in the same modality at study and test (e.g., spoken versus printed) than for words that are studied and tested in different modalities (Mulligan & Osborn, 2009).

However, previous multimedia learning studies on modality effects only included written tests. This could have caused a negative effect on performance in the audio-visual condition because the match of written questions with audio-visual materials is lower than with visual-only materials. In order to see if this modality match contributed to reversed modality effects in previous experiments, we experimentally manipulated the test modality in the present study. We hypothesized that long-term outcomes of audio-visual presentations could be more comparable to visual-only presentations when tested with auditory questions than when tested with written questions due to the greater overlap between study and test conditions.

In sum, the present study investigated adult learners' performance on oral and written retention and transfer questions immediately and 1 day after studying audio-visual or visual-only presentations. This was done to establish whether long-term reversed modality effects found with children generalize to adult learners, are related to study behaviours during learning, and depend on the modality of the performance measure.

Method

Sample and design

Participants

Eighty-four undergraduate university students (83% female, $M_{age} = 21.4$, $SD = 2.7$) participated in the experiment for course credits or a monetary compensation. The data of two additional participants were lost due to technical problems during the recording of spoken answers. Students did not have relevant prior knowledge about the topic of the materials. Their native language was Dutch or German; they had learned English for 6.8 years ($SD = 1.7$) on average and all reported experience with studying information in English. To increase motivation, the students were informed at the beginning of the experiment that they could request feedback on their test results and could win a bonus of 10 euro for good performance. The students were randomly assigned to either the visual-only group (V, $n = 41$) or the audio-visual group (AV, $n = 43$). The resulting groups did not differ from each other in terms of reported age, gender, native language, English skills or prior education.

Design

The study has a 2 (between-subject factor study modality: audio-visual or visual-only materials) × 2 (within-subject factor test modality: written questions or oral questions) × 2 (within-subject factor time: immediate, 1 day after learning) design with transfer and retention performance as two separate dependent variables.

Instruments and measurements

Study materials and text modality

The study materials were based on multimedia materials with animations and spoken narrations about magnetic resonance imaging physics (Schild & Specht, 2007), from which we made a learner-paced presentation of 128 slides each with an (edited) still frame from the animation and text of one to three sentences. In the visual-only condition, on-screen text was displayed under the picture. In the audio-visual condition, participants listened to spoken narrations, which started automatically when a slide was opened and could be repeated unlimitedly by clicking on a speaker symbol under the picture. The total length of the text was 2695 words; the total length of all narrations was 20 min 16 s. All verbal information was presented in English. The narrations were taken from the original materials which were spoken by a professional male native speaker at a rate of approximately 133 words per minute.

The materials were informally piloted with six participants without prior knowledge to see how long studying would take approximately. The pilot participants all evaluated the materials as informative and understandable. There were no ceiling or floor effects on retention and transfer tests. Pictures were crucial for understanding because the materials covered spatial concepts that were difficult to understand based on verbal information only (e.g., spatial alignment of protons in a magnetic field). The 84 participants of the full experiment rated the statement 'The pictures helped me understand the materials' with an average of 4.6 on a 5-point scale (1 = *not at all*, 5 = *very much*) after learning.

Study behaviours

The presentation was learner controlled: Participants pressed arrow keys on the keyboard in order to move back and forth through the slides at a self-chosen speed, with two numbers in the corner of the screen showing the number of the current slide and the total number of slides. Without the participants' knowledge, the timing of every key press and every mouse click on the sound replay button was recorded. From these data, we derived three measures: (1) the total time spent studying; (2) the total number of slides that the students looked at; and (3) the average time spent studying per slide. Measures (2) and (3) were calculated only for slides that were displayed for at least 1 s, to exclude the slides that the participants opened while quickly paging through the presentation. Participants had to move through all slides before the experiment proceeded to the next phase, the immediate test.

Performance measures: oral and written retention and transfer questions

Participants completed one test immediately after studying and a second test on the next day. Both tests contained a block of 14 written and a block of 14 oral open questions. The presentation order of these blocks was counterbalanced across participants. Written questions were displayed on the screen and participants entered their answers into a field under the question. Oral questions were presented with headphones and could be repeated by clicking on a speaker symbol. A female native speaker recorded the oral questions at a rate of approximately 121 words per minute. Participants gave their answer by speaking into a microphone. Recording started automatically at the beginning of each question and continued until participants proceeded to the next question. Participants had unlimited time to answer each question, but it was not possible to change a previous answer once a new question had started. The button to go to the next question became active after 20 s to keep the participants from giving up too quickly. The participants could see the number of current questions and the total number of questions during the test.

Example questions and scoring

We used 40 *retention* questions to measure the quantity of learning and 16 *transfer* questions to measure the quality of learning. Retention questions could be answered by reproducing information from one or more slides; transfer questions required the participants to apply their knowledge to a new context. For example, one retention question was: 'What is spin?' (answer based on the materials: 'the rotating movement of protons around their axis'). One transfer question was: 'What do you need to build an MR scanner?' (answer: 'a strong magnetic field, a gradient in the magnetic field, radio frequency pulses and something to pick up signals'). Because students often gave incomplete answers in which some but not all aspects of the correct answer were mentioned, we used a quotation scheme to assign points for parts of the correct answer. For example, the reply 'a strong magnetic field and radio frequency pulses' to the cited transfer question was scored with 2 points out of 4. Because the number of aspects scored with the quotation scheme differed somewhat between questions, relative scores were calculated per item. These item scores were subsequently averaged per participant for further analyses. Thus, the following scores that we report represent the average proportion correct per item.

Experimental control

The presentation order of the oral and written question blocks was counterbalanced across participants. To ensure that the content of all blocks was equally difficult, we randomly assigned 10 (of the 40) retention questions and 4 (of the 16) transfer questions to each of the four testing blocks (written/oral questions × day 1/day 2) for each participant from the audio-visual group, and presented the same set of questions to one participant from the visual-only group. The order of the questions within blocks was random, with the exception that there were no transfer questions among the first three questions to let participants start with the comparably easier retention questions.

Inter-rater reliability and internal consistency

All open-ended questions were scored by the first author. To control the reliability of scoring, a second independent rater who was familiar with the learning materials scored a random selection of 24 tests. The intra-class correlations were 0.89 for the retention scores and 0.85 for the transfer scores, which indicates high inter-rater reliability (Cicchetti, 1994).

Commonly used estimates of internal consistency, like Cronbach's α, could not be used for the present study because due to the random assignment of questions to test blocks, the questions were organized into different scales for every second participant. Therefore, we instead computed the correlations between the average scores of all items on a test with the single item scores on the same test (e.g., the correlation between each participant's score on question 1 and the same participant's average score on the test in which question 1 was presented). This revealed one retention question that did not significantly correlate with the average score. An inspection of participants' answers suggested that the formulation of this question was confusing and, therefore, the question was excluded from analysis. Retention scores used for further analysis were calculated based on the remaining 39 questions. All 16 transfer questions were retained.

Procedure

The first session with study phase (audio-visual or visual-only materials) and first test (oral and written retention and transfer questions) took about 1 h and 15 min. At the beginning of the session, participants filled in a questionnaire on demographic information and prior knowledge. Before studying, participants were informed that it took on average approximately 30–40 min to go through the materials once (an estimation based on pilot tests), but that they should revise the materials as often as they wanted and take as much time to study as they wanted. Participants were tested as soon as they decided to finish studying and there was no time limit for the test. Afterwards, they filled in a questionnaire about the materials and their motivation during learning. The second session 1 day later took about 35 min and consisted of an unannounced second test and a questionnaire in which participants were asked whether they had reviewed related materials in the delay between the two sessions. The participants were told that this information would have no consequence for their chance to win the bonus, but none of them reported having reviewed related materials.

Results

Table 1 summarizes the mean scores and standard deviations for performance at day 1 and day 2. Two repeated measures analyses of variance were conducted for retention and transfer scores, respectively, with study modality (audio-visual or visual-only) as between-subject factor and test modality (written or oral questions) and time (immediately after learning, after 1 day) as within-subject factors.

Retention scores

The analysis of retention scores showed no effect of study modality, $F(1, 82) = 2.90$, 8, $p = .093$, $\eta_p^2 = 0.034$, or time, $F(1, 82) = 3.68$, $p = .059$, $\eta_p^2 = 0.059$; but a large significant effect of test modality due to higher scores on written than on oral questions, $F(1, 82) = 16.91$, $p < .001$, $\eta_p^2 = 0.171$; and a significant interaction between study modality and time, $F(1, 82) = 8.73$, $p = .004$, $\eta_p^2 = 0.096$. The interactions between study modality and test modality, $F(1, 82) < 1$, test modality and time, $F(1, 82) = 2.75$, $p = .10$, $\eta_p^2 = 0.03$, and the three-way interaction of study modality, test modality, and time, $F(1, 82) = 1.73$, $p = .19$, $\eta_p^2 = 0.02$, were not significant. To follow up on the significant interaction between study modality and time, we performed pairwise comparisons using separate one-way *F*-tests. These revealed that performance did not differ between the visual-only group and the

Table 1. Written and Oral Retention and Transfer Performance Immediately and One Day After Learning

	Immediate test		Test after 1 day	
Performance measure	Audio-visual condition *M* (*SD*)	Visual-only condition *M* (*SD*)	Audio-visual condition *M* (*SD*)	Visual-only condition *M* (*SD*)
Written retention	0.55 (0.19)	0.58 (0.18)	0.48 (0.20)	0.56 (0.19)
Oral retention	0.49 (0.20)	0.49 (0.19)	0.43 (0.20)	0.54 (0.21)
Average retention	0.52 (0.18)	0.54 (0.16)	0.45 (0.19)	0.55 (0.18)
Written transfer	0.55 (0.24)	0.54 (0.21)	0.50 (0.23)	0.52 (0.22)
Oral transfer	0.49 (0.22)	0.41 (0.21)	0.38 (0.21)	0.50 (0.21)
Average transfer	0.52 (0.19)	0.47 (0.16)	0.44 (0.17)	0.51 (0.17)

Note. The 41 participants in the visual-only condition studied pictures and on-screen text and the 43 participants in the audio-visual condition studied pictures and spoken narrations. The table lists relative item scores (calculated as proportion correct per item, averaged per participant).

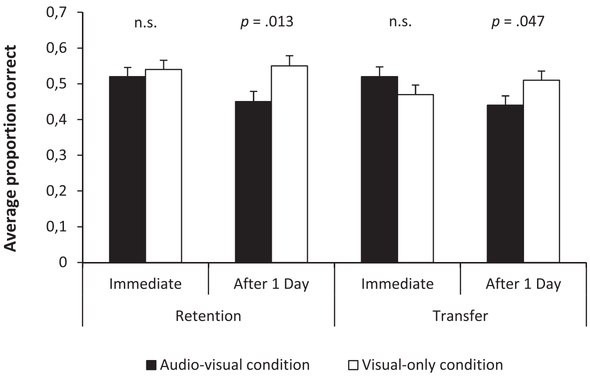

Figure 1 Average Retention and Transfer Scores Immediately and One Day After Studying for the Audio-Visual and the Visual-Only Groups. Error Bars Represent Standard Errors

audio-visual group immediately after learning, $F(1, 82) < 1$, but performance was higher in the visual-only group than in the audio-visual group after 1 day, $F(1, 82) = 6.39$, $p = .013$, $\eta_p^2 = 0.07$, $d = 0.56$, see Figure 1.

Transfer scores

The main effects of study modality and time on transfer scores were not significant, both $F(1, 82) < 1$. However, the main effect of test modality was again significant due to higher scores on written questions than on oral questions, $F(1, 82) = 17.27$, $p < .001$, $\eta_p^2 = 0.17$. Again, there was also an interaction between time and study modality, $F(1, 82) = 9.95$, $p = .002$, $\eta_p^2 = 0.11$, because performance was significantly better for the visual-only group than for the audio-visual group after 1 day, $F(1, 82) = 4.11$, $p = .046$, $\eta_p^2 = 0.05$, $d = 0.45$, but not immediately after learning, $F(1, 82) = 1.45$, $p = .233$, $\eta_p^2 = 0.02$, $d = 0.27$. The interactions between study modality and test modality, and between test modality and time, were not significant, both $F(1, 82) < 1$. However, the three-way interaction among study modality, test modality and time was significant, $F(1, 82) = 4.30$, $p = .041$, $\eta_p^2 = 0.05$. This was due to the fact that the advantage of visual-only materials over audio-visual materials was significant after 1 day for the oral transfer questions, $F(1, 82) = 7.81$, $p = .006$, $\eta_p^2 = 0.09$, $d = 0.62$, but did not reach significance for the written transfer questions, $F(1, 82) = 3.05$, $p = .084$, $\eta_p^2 = 0.04$, $d = 0.39$.

Study behaviours

Mean study times in the two experimental conditions were similar; participants in both groups spent on average 33 min studying (visual-only: $M = 33.0$, $SD = 13.5$; audio-visual: $M = 33.1$, $SD = 7.6$), $F(1, 82) < 1$. In spite of this similarity in total study times, participants in the visual-only group looked at the slides for 8.9 s on average ($SD = 3.1$), whereas participants in the audio-visual group spent 11.4 s on each slide ($SD = 1.6$). This difference was significant, $F(1, 82) = 21.03$, $p < .001$, $\eta_p^2 = 0.20$, $d = 0.51$. At the same time, participants in the visual-only group switched between the 128 slides of the presentation on average 61 times more often than participants in the audio-visual group, $F(1, 82) = 19.91$, $p < .001$, $\eta_p^2 = 0.20$, $d = 1.01$, looking at more slides in total (visual-only: $M = 245.9$, $SD = 79.2$; audio-visual: $M = 184.4$, $SD = 40.0$). Note that not only average measures of study behaviours differed between the two conditions, but also variations in study behaviours: All standard deviations were almost twice as large in the visual-only group as in the audio-visual group.[1] Clearly, there was more inter-individual variation in the visual-only group than in the audio-visual group.

Correlations between study behaviours and test scores

Table 2 contains the correlations between measures of study behaviour and performance outcomes. For the visual-only group, there were no significant correlations. For the audio-visual group, several correlations between total study time and different outcomes at day 1 or day 2 were significant at $\alpha = .05$, but these results must be interpreted with caution due to the high chance for type I errors, given that we tested 24 correlations in total. Nevertheless, it appears as if there is a relation between the time learners spent studying audio-visual materials and performance outcomes, but no such relation for visual-only materials.

Discussion

The purpose of this paper was to investigate how the effects of text modality in learner-paced multimedia presentations change over time, how they relate to study behaviours during learning, and how they are

Table 2. Correlations Between Study Behaviours and Retention and Transfer Performance

Performance measure	Session	Total study time		Mean study time per slide (>1 s)		Number of slides studied (>1 s)	
		AV	V	AV	V	AV	V
Retention questions	Day 1	.31*	.09	.27	.20	.18	–10
	Day 2	.31*	.10	.10	.08	.29	.07
Transfer questions	Day 1	.38*	–.00	.31*	.09	.24	–.05
	Day 2	.24	.04	.07	–.01	.18	.10

Note. The table displays correlations between three measures of study behaviours and retention and transfer scores at day 1 and day 2, averaged across written and oral questions. Correlations were calculated separately for the two experimental conditions (AV = audio-visual; V = visual-only).
*$p < .05$, two tailed. The asterisks indicate significance uncorrected for repeated testing.

influenced by the modality of the final performance measure. Participants studied extensive learner-paced presentations that contained pictures with either on-screen text (visual-only group) or spoken narrations (audio-visual group) and answered oral and written test questions immediately and 1 day after learning. This experiment led to three major results. First, there was no modality effect immediately after learning, but the visual-only group performed significantly better than the audio-visual group 1 day after learning on three of four performance measures. This reversed modality effect contradicts the common multimedia design principles that audio-visual materials are superior to visual-only materials (e.g., Mayer, 2009). Instead, after a delay of just one night, learners seem to benefit more from on-screen text. Second, while total study times were the same for audio-visual and visual-only materials, learners went through the visual-only materials at a faster pace and repeated more slides. This suggests that strategic rereading of the materials may have contributed to reversed modality effects. Third, reversed modality effects generalized across test modalities but oral questions appeared to be more difficult to answer than written questions.

Long-term effects of text modality

Visual-only presentations led to better long-term outcomes than audio-visual presentations, as we had predicted in our first hypothesis based on earlier experiments with children (Segers *et al*., 2008; She & Chen, 2009; Witteman & Segers, 2010). Moreover, whereas two of the earlier studies found long-term reversed modality effects only with transfer questions, we found effects on both retention and transfer questions. This could be due to stronger advantages of on-screen text for the more experienced adult readers in this study, or to the more complex materials and test questions that we used.

The reported reversed modality effects have a medium effect size, and are relevant for the literature because they contradict the idea that audio-visual materials are superior to visual-only materials (e.g., Mayer, 2009). Previous estimates of modality effects reported in a frequently cited meta-analysis by Ginns (2005) have recently been questioned on the grounds that statistical corrections for publication bias may not have been applied sufficiently. After such corrections, estimates of modality effect sizes are markedly smaller (Lindow *et al*., 2011). The present results further add to this discussion by documenting a *reversed* modality effect.

A possible explanation for the reversed modality effects that we found is that auditory narrations could negatively affect working memory processing when learners must maintain complex verbal information online in order to integrate it with subsequently presented information (e.g., Kalyuga, 2011). This can lead to detrimental working memory overload if the presented information is long and complex (Wong *et al*., 2012); the consequence may be a better integration and deeper processing of on-screen text than spoken narrations, which could make the information more resistant to forgetting (e.g., Lockhart & Craik, 1990; Bjork, 1994). Second, the transiency of auditory information could also influence which study behaviours learners choose and in turn influence how information is processed and integrated.

Study behaviours

In our second hypothesis, we predicted that the visual-only group would need longer study times than the audio-visual group based on the split-attention idea that participants first read text and then look at pictures (as reported in eye-tracking studies like Schmidt-Weigand *et al*., 2010a), whereas they can look at pictures and simultaneously listen to narrations. However, the results surprisingly indicated that learners spent on average the same time, about half an hour, studying in both experimental groups, and the visual-only group even went through the slides at a faster pace and switched more often between slides than the audio-visual group. Like this, participants in the visual-only group studied on average around 30% more slides than the audio-visual group in the same total study time. These results contradict the idea that learners compensate for split attention with longer study times in visual-only conditions (Tabbers *et al*., 2004). However, results are in line with several earlier studies in which total study times were the same for visual-only and audio-visual presentations (e.g., Tabbers, 2002, Exp.2.2, Exp.4.1; Schmidt-Weigand, 2006; Harskamp, Mayer, & Suhre, 2007).

The recorded differences in study pace suggest that learners choose different study strategies when working with audio-visual and visual-only materials. Possibly, learners in the visual-only group repeated slides to (re)read difficult or important parts of the information. Such skimming through materials may have helped them to better connect and integrate the presented information and thereby improved retention over time (e.g., McNamara & Kintsch, 1996). In contrast, auditory narrations do not allow learners to quickly find and selectively review parts of the text and instead force learners to adapt to the pace of the narrations (Kalyuga, 2011, 2012; Leahy & Sweller, 2011). Our results support this idea by showing a higher variation of study behaviours in the visual-only than in the audio-visual condition. Apparently, learners who studied on-screen text showed more differences in reading strategies whereas learners who studied spoken narrations followed a more similar pace, which may have been influenced by the timing of the narrations. This outcome may also tentatively explain why longer total study times in the audio-visual but not in the visual-only condition correlated with later performance, if it reflects a stronger relation between study times and motivation in the audio-visual condition. Learners in the audio-visual condition had to adapt to the pace of the presented narrations and could not speed them up or slow them down. Variations in study times are therefore likely influenced by learners' choices to invest extra time to repeat narrations or slides instead of following the pace of the narrations. In contrast, variations in study times in the visual-only condition could also be influenced by individual differences in reading speed.

The interpretation that modality effects were driven by study behaviours must be treated with some caution, however, because although we found differences in the *average* number of slides studied and the *average* speed of studying between the audio-visual and the visual-only groups as well as differences in performance between the groups, there was no correlation between the study behaviours and performance across participants. A possible explanation for this is that positive effects of study behaviours on performance were concealed by negative effects of performance on study behaviours. For example, more repetitions may have improved performance, but at the same time, trouble understanding the materials may have led students to do more repetitions. It is not possible to statistically disentangle these two effects. Therefore, we can only state that, in the present study, differences in the amount of repetitions of slides exist on the group level and could possibly explain reversed modality effects, but we cannot back up this interpretation with results on the level of individual participants.

Test modality

With respect to our third hypothesis, we did not find a matching effect for study modality and test modality in the form of an advantage for the visual-only group on written questions, or for the audio-visual group on oral questions. If anything, evidence for reversed (study) modality effects at the delayed test was stronger for oral questions than for written questions. Hence, it is not likely that long-term benefits of visual-only presentations in previous studies were due to the modality of the test questions. Instead, reversed modality effects generalized across test modalities, which suggests that direct effects of text modality are limited to early stages of perceptual and working memory processing, after which information is integrated into a mental model that does not retain initial text modality (cf. Mayer, 2009). This supports the idea that differences between audio-visual and visual-only materials are due to differences in semantic integration processes after initial sensory processing, and are therefore independent of the modality of later performance measures.

Intriguingly, the oral questions were in general more difficult to answer than the written questions, even though their content was experimentally controlled to be the same. This large effect could be due in part to the fact that participants had more experience with written than with oral tests. Furthermore, the oral answers were recorded continuously, with no option to later delete parts of them, whereas the written answers could be edited in a text field. Although the participants were encouraged to correct their oral answers if necessary, they rarely made use of this option.

There is little research that directly compares performance on written and oral tasks. Recently, it was reported that college students achieve higher scores on oral examinations than on equivalent written tests (Huxham, Campbell, & Westwood, 2012). However, these students had several weeks to prepare for the announced written or oral examinations. Results by Huxham *et al*. (2012) could therefore be influenced by the amount of preparation that students invested, which could be higher for oral examinations than for written examinations, for example, if oral examinations are perceived as more intimidating or difficult. In contrast, our tests were unannounced and computerized, a situation in which oral tests proved more difficult to answer than equivalent written tests. As we combined printed questions with written answers and auditory questions with oral answers, future research must establish to what extent the difficulty was due to the modality of the questions (printed or spoken questions), to the response mode (written or oral answers) or the combination of both.

Some limitations of the present study should be taken into account when interpreting the results. First, materials were presented in English, a foreign language for the participating students. The students regularly used English study materials and neither the visual-only nor the audio-visual group indicated trouble understanding the language of the materials on a questionnaire after learning. This makes it unlikely that modality effects were influenced by the language of the materials, but the language may have overall increased the difficulty of the task. Second, participants received one test immediately after learning and a different test after 1 day. It is likely that the immediate test had a practice effect on later performance (cf. the literature on testing effects, Karpicke & Roediger, 2008). However, both groups underwent the same tests and different questions were used at the two testing moments to reduce carry-over effects. Moreover, performance on the immediate test was similar in the audio-visual and the visual-only groups, which makes it unlikely that the two groups experienced different testing effects that may have influenced modality effects. Nevertheless, an interesting question for future research is whether testing effects and modality effects interact, if possible taking into account the modality of both study and test materials. Another question for future research is how modality effects change over longer periods of time. We only measured performance outcomes 1 day after learning but in order to model forgetting over time, it would be interesting to include more measurements with varying delays. Finally, it would be relevant to investigate the effect of text modality in other multimedia formats, such as interactive simulations or materials that combine auditory narrations with on-screen keywords.

Conclusion

The learner-paced multimedia presentations used in this study produced better results with on-screen text than with spoken narrations after a one-night delay. This reversed modality effect was of a medium size and it was documented with both oral and written questions, adding to the growing literature on boundary conditions of classic modality effects (e.g., Schüler *et al*., 2011; Tabbers & van der Spoel, 2011; Reinwein, 2012). Study behaviours could offer an explanation of reversed modality effects as on-screen text led to more frequent rereading, whereas study behaviours in the audio-visual condition varied less – possibly because learners were more limited by the fixed pace of the auditory narrations. Overall, the present results demonstrate that in order to evaluate modality effects, it is necessary to take into account how active learners respond to multimedia input and how text modality limits or enables beneficial study behaviours. Moreover, we make a case for the fact that it is not sufficient to evaluate multimedia design decisions immediately after learning because effects can change over time.

Acknowledgements

The authors thank Paul Konstantin Gerke for technical support and Caressa Janssen for assistance with test scoring. This research was supported by a Huygens scholarship from the Dutch Ministry of Education, Culture and Science and by a grant from the Netherlands Organization for Scientific Research (NWO 056-33-014). No conflicts of interest exist.

Notes

[1]Because this violates the statistical assumption of homogeneity of variance, we repeated the analyses reported in this section with Welch *t*-tests, which account for unequal sample variances. All results remained significant with $p < .001$.

References

Ayers, P. L., & Sweller, J. T. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 135–146). Cambridge: Cambridge University Press.

Bjork, R. A. (1994). Memory and meta-memory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi:10.1037/1040–3590.6.4.284

Crooks, S. M., Cheon, J., Inan, F., Ari, F., & Flores, R. (2012). Modality and cueing in multimedia learning: Examining cognitive and perceptual explanations for the modality effect. *Computers in Human Behavior*, *28*(3), 1063–1071. doi:10.1016/j.chb.2012.01.010

Fletcher, J., & Tobias, S. (2005). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 117–134). New York, N.Y.: Cambridge University Press.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*(4), 313–331. doi:10.1016/j.learninstruc.2005.07.001

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction*, *17*(5), 465–477. doi:10.1016/j.learninstruc.2007.09.010

Huxham, M., Campbell, F., & Westwood, J. (2012). Oral versus written assessments: A test of student performance and attitudes. *Assessment and Evaluation in Higher Education*, *37*(1), 125–136. doi:10.1080/02602938.2010.515012

Kalyuga, S. (2011). Effects of information transiency in multimedia learning. *Procedia – Social and Behavioral Sciences*, *30*(0), 307–311. doi:10.1016/j.sbspro.2011.10.061

Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review*, *7*(2), 145–159. doi:10.1016/j.edurev.2011.12.002

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*(4), 351–371. doi:10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2–6

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. doi:10.1126/science.1152408

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

Lindow, S., Fuchs, H. M., Fürstenberg, A., Kleber, J., Schweppe, J., & Rummer, R. (2011). On the robustness of the modality effect: Attempting to replicate a basic finding. *Zeitschrift für Pädagogische Psychologie*, *25*(4), 231–243. doi:10.1024/1010-0652/a000049

Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *44*(1), 87–112. doi:10.1037/h0084237

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147–158). New York, N.Y.: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, N.Y.: Cambridge University Press.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247–288. doi:10.1080/01638539609544975

Mulligan, N. W., & Osborn, K. (2009). The modality-match effect in recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(2), 564–571. doi:10.1037/a0014524

Nelson, J. R., Balass, M., & Perfetti, C. A. (2005). Differences between written and spoken input in learning new words. *Written Language and Literacy*, *8*(2), 25–44. doi:10.1075/wll.8.2.04nel

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford: Basil Blackwell.

Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research*, *41*(1), 1–32. doi:10.1007/s10936-011-9180-4

Roediger, H. L., Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: the impetus from the levels-of-processing framework. *Memory (Hove, England)*, *10*(5–6), 319–332. doi:10.1080/09658210224000144

Rummer, R., Schweppe, J., Fürstenberg, A., Seufert, T., & Brünken, R. (2010). Working memory interference during processing texts and pictures: implications for the explanation of the modality effect. *Applied Cognitive Psychology*, *24*(2), 164–176. doi:10.1002/acp.1546

Schild, H. H. (Writer), & Specht, C. (Director) (2007). MRI made easy [CD-ROM]. Germany: Bayer Schering Pharma AG/Dr. Specht Film & Medien GmbH.

Schmidt-Weigand, F. (2006). *Dynamic visualizations in multimedia learning: The influence of verbal explanations on visual attention, cognitive load and learning outcome* (Unpublished doctoral dissertation, Justus-Liebig-Universitaet Giessen). Retrieved from http://geb.uni-giessen.de/geb/volltexte/2006/2699/

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010a). 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.2009.02.011

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010b). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology*, *24*(2), 226–237. doi:10.1002/acp.1554

Schüler, A., Scheiter, K., & Schmidt-Weigand, F. (2011). Boundary conditions and constraints of the modality effect. *Zeitschrift für Pädagogische Psychologie*, *25*(4), 211–220. doi:10.1024/1010-0652/a000046

Segers, E., Verhoeven, L., & Hulstijn-Hendrikse, N. (2008). Cognitive processes in children's multimedia text learning. *Applied Cognitive Psychology*, *22*(3), 375–387. doi:10.1002/acp.1413

She, H.-C., & Chen, Y.-Z. (2009). The impact of multimedia effect on science learning: Evidence from eye movements. *Computers and Education*, *53*(4), 1297–1307. doi:10.1016/j.compedu.2009.06.012

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278. doi:10.1038/nature04286

Stiller, K. D., Freitag, A., Zinnbauer, P., & Freitag, C. (2009). How pacing of multimedia instructions can influence modality effects: A case of superiority of visual texts. *Australasian Journal of Educational Technology*, *25*(2), 184–203.

Tabbers, H. K. (2002). *The modality of text in multimedia instructions: Refining the design guidelines* (Unpublished doctoral dissertation, Open University of the Netherlands). Retrieved from http://openjuridischehogeschool.com/Docs/Onderzoek/Promoties/2002/Doctoral_dissertation_Huib_Tabbers_%20webversion.pdf

Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2001). *The modality effect in multimedia instructions*. Paper presented at the Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society, Mahwah, NJ.

Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *The British Journal of Educational Psychology*, *74*(1), 71–81. doi:10.1348/000709904322848824

Tabbers, H. K., & van der Spoel, W. (2011). Where did the modality principle in multimedia learning go? A double replication failure that questions both theory and practical use. *Zeitschrift für Pädagogische Psychologie*, *25*(4), 221–230. doi:10.1024/1010-0652/a000047

Witteman, M. J., & Segers, E. (2010). The modality effect tested in children in a user-paced multimedia environment. *Journal of Computer Assisted Learning*, *26*(2), 132–142. doi:10.1111/j.1365-2729.2009.00335.x

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.learninstruc.2012.05.004