Integrating written text and graphics as a desirable difficulty in long-term multimedia learning

Judith Schweppe*, Ralf Rummer
University of Erfurt, Department of Psychology, P.O. Box 900 221, 99105 Erfurt, Germany

ABSTRACT

Many principles for the design of multimedia learning materials share the recommendation to facilitate processing. One prominent example is the modality principle, according to which pictures should be presented with auditory rather than visual texts. Research on desirable difficulties, however, indicates that unlike short-term learning — long-term learning benefits when processing is more demanding and therefore more effortful. In a classroom experiment (Experiment 1) and in a laboratory study (Experiment 2), we tested whether the modality principle serves long-term learning. In a multimedia presentation on the formation of lightning, we varied the text modality (oral vs. written) and the delay between learning and test (retention and transfer performance tested immediately after instruction vs. one week later). In the immediate tests, there was either an auditory advantage (Experiment 1) or no difference (Experiment 2). However, when learning was tested after a delay, the combined processing of written text and animations led to better transfer performance than oral text and animations in both experiments. This suggests that written text presentation serves as a desirable difficulty that supports long-term learning.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The primary goal of research on multimedia learning is to provide recommendations for presenting learning materials. Several principles have been put forward, most of which were derived from or explained in terms of the Cognitive Theory of Multimedia Learning (CTML, Mayer, 2009, 2014) and the Cognitive Load Theory (CLT, Paas & Sweller, 2014; Plass, Moreno, & Brünken, 2010; Sweller, Ayres, & Kalyuga, 2011). Both theories focus on working memory, which is conceptualised as a limited cognitive resource, and assume that processing and learning fail whenever more capacity is needed than is available. The common core of most principles is thus to facilitate processing by reducing cognitive demands on working memory.

One influential recommendation based on this idea is to present texts in a spoken format when they accompany pictures, in order to address both modalities and therefore demand different subsystems of working memory (auditory and visual; for overviews see Ginns, 2005; Low & Sweller, 2014). According to this modality principle, both texts and pictures (or animations) must, at least initially, be processed in the same (visuo-spatial) working memory subsystem when texts are presented in a written format. In contrast, spoken texts and pictures (or animations) are processed in different working memory subsystems. The distribution of information among different working memory subsystems is assumed to help avoid cognitive overload (e.g., Mayer & Moreno, 1998; Moreno & Mayer, 1999). The explanation that oral and written text are stored and processed in different working memory subsystems has been criticized recently, since it is at odds with Baddeley’s (e.g., 1986) working memory model and with research on working memory (Rummer, Schweppe, Fürstenberg, Scheiter, & Zindler, 2011; Rummer, Schweppe, Fürstenberg, Seufert, & Brünken, 2010; Rummer, Schweppe, Scheiter, & Gerjets, 2008; Schüler, Scheiter, Rummer, & Gerjets, 2012; Schüler, Scheiter, & van Genuchten, 2011; Tabbers, 2002). In addition (or alternatively), it is assumed that written texts and pictures/animations cannot be focussed on simultaneously and must therefore be processed sequentially. This creates a split-attention situation, which forces learners to effortfully integrate the text and picture information (Ayres & Sweller, 2014). In line with the latter explanation, the modality effect, i.e. the finding that learners perform better with a combination of spoken text and pictures than with written...
text and pictures, is observed less consistently with the sequential presentation of texts and pictures (e.g., Baggett & Ehrenfeucht, 1983; Tiene, 2000; but see Moreno & Mayer, 1999). A further boundary condition that applies to both explanations is the finding that the modality effect disappears (or even reverses) when learners are free to determine the pace at which they process the learning materials (e.g., Ginns, 2005; Tabbers, Martens, & van Merriënboer, 2004).

1.1. Desirable difficulties

We aim to investigate multimedia learning from a perspective that hinges on the apparent paradox of fostering learning by impeding processing rather than facilitating it and that focuses on long-term learning in particular. Several studies have demonstrated that cognitively more demanding learning conditions improve long-term learning, although this often comes at the expense of processing speed or initial learning (for overviews, see Pashler, Rohrer, Cepeda, & Carpenter, 2007; Roediger & Karpicke, 2006).

This suggests that one cannot reliably conclude that good initial performance (as observable in an immediate test) translates into good long-term learning. One explanation for this phenomenon, which has been coined desirable difficulties (Bjork, 1994), is that conditions that facilitate processing induce a shallow processing mode that is particularly harmful for the establishment of stable long-term memory representations. In contrast, learning with more difficult material gives rise to processing difficulties, which consequently leads to deeper processing and representations that are less susceptible to forgetting (Bjork, 2013; Craik & Lockhart, 1972).

There is one example concerning the presentation of materials in multimedia research that can be regarded as a desirable difficulty: De Crock, van Merriënboer, and Paas (1998) presented learners with practice problems in a random schedule with problems that varied with respect to the procedures necessary for performing each task compared to a blocked schedule. The random schedule increased the complexity and the cognitive demands but resulted in better transfer performance.

While in tests administered immediately after instruction short-lived advantages of the easier conditions often outweigh the beneficial effects of difficult encoding, those latter effects are particularly strong when it comes to long-term tests of learning (Pashler et al., 2007; Roediger & Karpicke, 2006). Given that several recommendations for the design of multimedia learning materials emphasize the facilitation of processing, this may imply that what is beneficial for short-term multimedia learning sometimes hinders long-term learning. In research on multimedia learning, the literature is quite clear that an increase in processing demands usually impedes rather than aids learning. Yet maybe the negative effects that cognitive load has on initial understanding only mask its positive effects when learning outcomes are restricted to a condition of immediate testing. Effects attributed to desirable difficulties are stronger and often only present when performance is tested after a delay, that is, after forgetting had occurred (e.g., Roediger & Karpicke, 2006). If this is the case, conditions that are beneficial in immediate tests of multimedia learning might have adverse effects in the long term and the respective principles may need to be revised for long-term learning.

One condition for which this could be the case is the processing of written versus oral text in combination with pictures. A situation in which learning materials must be processed sequentially—which is the case with visual texts plus pictures/animations—forces the learner to switch between text and picture. It is thus necessary to retain the pictorial information while reading the text and the verbal information while processing the picture (Rummer et al., 2008, 2010). This causes learners to effortlessly integrate the information sources—a process that learners are likely to engage in when they are motivated and that may underly the higher cognitive load associated with written compared to oral texts (e.g., Ayres & Sweller, 2014). Such a demanding integration process may be harmful for initial processing, as evident by the immediate modality effect, but may have the advantage of leading to deeper processing in terms of the levels-of-processing hypothesis (Craik & Lockhart, 1972), and therefore be particularly important for long-term memory. By presenting written text along with pictures or animations, learners might be “tricked” into deeper processing, which, in turn, results in more stable long-term memory representations. Of course, such positive long-term benefits of difficult learning conditions can only occur in cases in which the difficulty impedes processing but does not entirely preclude comprehension.

1.2. Long-term tests of multimedia learning

In surveying the literature on multimedia learning, we found only few experiments in which learning was measured after a noticeable delay. With respect to the modality effect that we investigate in our study, we found four such studies (Segers, Verhoeven, & Hulstijn-Hendrikse, 2008; She & Chen, 2009; Van den Broek, Segers, & Verhoeven, 2014; Witteman & Segers, 2010). Segers et al. (2008) observed an inverse modality effect (i.e., an advantage for written text) in a transfer test one week after the learning phase. Witteman and Segers (2010) observed a written text advantage for a transfer test after one day and no modality differences one week later. She and Chen (2009) observed a written advantage for retention after a five-week delay between encoding and test, and Van den Broek et al. (2014) after a delay as short as one day. These findings are important first steps towards a test of the long-term consequences of reading versus listening. Remarkably, none of the delayed tests revealed a modality effect in terms of an advantage for oral texts—there was either no modality difference or a written advantage. However, these inverse modality effects do not suffice to question the applicability of the modality principle to long-term learning since in all studies the presentation of the materials was learner-paced. As indicated above, learner-paced presentation often results in visual advantages even with immediate tests and is therefore a boundary condition of the modality effect (e.g., Ginns, 2005; Tabbers et al., 2004). Consequently, further research is imperative to determine whether reading is advantageous even with system-paced presentation.

1.3. Objectives of the present study

We aim to investigate the modality effect after a delay, but under conditions in which it is otherwise highly likely to occur. That is, in contrast to previous studies (e.g., Segers et al., 2008), we will test the modality principle with experimenter-paced learning. We will compare participants’ performance on retention and transfer tests after multimedia lessons with auditory versus visual text (based on the lightning materials by Mayer & Moreno, 1998; Moreno & Mayer, 1999) when these tests are administered either immediately or one week after the instruction. According to the modality principle, there should be an advantage of auditory over visual texts even with a delayed test. According to the desirable difficulties approach, however, visual text presentation should lead to superior performance on the delayed test because learners were forced to process the texts more effortfully and more deeply, leading to more stable memory representations that are less susceptible to forgetting.
2. Experiment 1

2.1. Method

2.1.1. Design and participants

The experiment followed a 2 (text modality: auditory vs. visual) x 2 (test delay: immediate test vs. test after one week) between-subjects design. Performance on a retention test and a transfer test served as dependent variables. 48 undergraduates (mean age: 21.6; native language: German) from different majors at the University of Erfurt participated in the experiment and received course credit for participation. Each of the four groups consisted of 12 participants.

2.1.2. Materials

We used Moreno and Mayer’s (1999) materials comprised of a knowledge checklist to assess prior meteorology knowledge, a computerized multimedia presentation of a lesson on lightning formation, and retention and transfer questions to measure learning outcomes. We refrained from using the visual-verbal matching test originally used as a third dependent variable, learning outcomes. We refrained from using the visual-verbal computerized multimedia presentation of a lesson on lightning knowledge checklist to assess prior meteorology knowledge, a ggle- or double-sentences as in Moreno and Mayer (1999).

To implement this, the learning text was presented piece-by-piece using the same partitioning into 16 sentences, two double-sentence texts) and corresponding short animation. In the visual condition, the text was successively presented underneath the animation with the same words, timing and speed of the auditory condition. A previously recorded auditory version of the learning material. In the auditory condition, the verbal information was presented via loudspeakers. Prior to presentation of the learning material. In the auditory condition, a previously recorded auditory version of the learning text was checked by a native English speaker.

2.1.2.1. Meteorology knowledge checklist. To assess participants’ prior meteorology knowledge, we used Moreno and Mayer’s (1999) 7-item knowledge checklist and 5-item self rating scale as a paper-pencil questionnaire. We computed experience scores following Moreno and Mayer’s scoring procedure.

2.1.2.2. Learning material. The original learning material used by Moreno and Mayer (1999, simultaneous presentation) describing the process of lightning formation in 16 brief texts (14 single sentences, two double-sentence texts) and corresponding short animations was presented continuously (total duration: 192 s). In the auditory condition, a previously recorded auditory version of the learning text was presented via loudspeakers along with the animation. In the visual condition, the text was successively presented underneath the animation with the same words, timing and speed of the auditory condition. To implement this, the learning text was presented piece-by-piece using the same partitioning into 16 single- or double-sentences as in Moreno and Mayer (1999).

2.1.2.3. Dependent variables transfer and retention tests. To measure learning performance, we used Moreno and Mayer’s (1999) retention and transfer tests as paper-pencil questionnaires. In the retention test, participants were asked to provide an explanation of the lightning formation process in a single open question. The transfer test consisted of four open questions asking for an explanation of one specific transfer problem each.

2.1.3. Procedure

Participants were tested in classroom settings. All participating classes were undergraduate foreign language courses with students of different majors. One class each was randomly assigned to the four experimental conditions. First, participants completed the meteorology knowledge checklist at their own rates, followed by the completely computer-paced presentation of the learning materials. Each class collectively watched a centrally projected presentation of the learning material. In the auditory condition, the verbal information was presented via loudspeakers. Prior to presentation, participants were not told when they would be tested. After viewing the presentation, participants in the immediate testing condition completed the retention and transfer tests. Each question was printed on a separate sheet of paper and completed one after the other. First, participants worked for 5 min on the retention question, followed by the four transfer questions with a processing time of 3 min each. For participants in the delayed testing condition, the procedure differed in one respect: instead of being tested immediately after viewing the learning materials, the two delayed groups were tested using the same tests and procedure as the immediate groups one week after the initial learning session. They were informed of this after they had viewed the presentation.

2.1.4. Scoring

A retention score was determined for each participant by counting the number of correctly recalled idea units reported by Moreno and Mayer (1999), regardless of wording. A maximum score of 19 was possible. For the analysis of the transfer test, the number of acceptable answers was counted for each participant in line with the open ended scoring procedure by Moreno and Mayer (1999). A transfer score was computed for each participant by summing the counts for the four questions (with no maximum score).

In accordance with the scoring procedure of Moreno and Mayer (1999), a portion of all questionnaires were rated by two independent scorers (double ratings for 31% of the questions). All scorers were blind to the participants’ experimental conditions. We reached a reasonably high inter-rater reliability for the retention (r = .95) and transfer tests (Question 1: r = .77, Question 2: r = .88, Question 3: r = .81, Question 4: r = .82).

2.2. Results

Means and 95% confidence intervals are reported in Table 1. Results are analyzed in a three-step manner. First, ANOVAs were conducted to test whether prior knowledge differed across the four experimental conditions. Second, to test the hypotheses, MANOVAs with the two dependent variables (retention and transfer scores) were conducted, followed by the two separate ANOVAs for each dependent variable. Partial eta-squared ($\eta^2_p$) is reported as a measure of effect size. For eta-square, effect sizes of .01, .06, and .14 correspond to small, medium, and large effects respectively (Cohen, 1988).

2.2.1. Prior knowledge

There were no significant differences between the four experimental conditions regarding prior knowledge (all Fs < 1.7).

2.2.2. Learning outcomes

The MANOVA revealed a significant main effect for delay in that immediate performance was better than delayed performance (Wilks’ $\lambda = .79$; $F(2,43) = 5.64$; $p = .007$; $\eta^2_p = .208$). While there was no main effect for text modality (Wilks’ $\lambda = .96$; $F < 1$), the interaction between modality and delay reached significance (Wilks’ $\lambda = .80$; $F(2,43) = 5.32$; $p = .009$; $\eta^2_p = .198$).

A 2 x 2 between-subjects ANOVA on the retention data revealed no significant effects. Participants in the auditory conditions outperformed those in the visual conditions numerically only when retention was tested immediately after the learning session. However, this difference did not reach statistical significance as a main effect ($F(1,44) = 1.45$; $p = .24$; $\eta^2_p = .032$) and only approached significance in a planned comparison contrasting the two immediately tested groups ($F(1,44) = 3.45$; $p = .07$; $\eta^2_p = .144$). Planned comparisons for the delayed conditions revealed no significant modality influence ($F < 1$). The main effect for delay ($F(1,44) = 1.45$; $p = .23$; $\eta^2_p = .032$) and the two-way interaction $F(1,44) = 2.03$; $p = .16$; $\eta^2_p = .044$) did not reach significance either.
In the case of the transfer test, there was no main effect for modality \((F < 1)\) but a significant main effect for delay \((F(1,44) = 6.97; p = .011; \eta^2 = .137)\). Most importantly, however, there was a significant two-way interaction between modality and delay \((F(1,44) = 10.71; p = .002; \eta^2 = .196)\). While immediate transfer performance was better following auditory text presentation as opposed to visual presentation (planned comparison: \(F(1,44) = 5.02; p = .03; \eta^2 = .171)\), delayed transfer performance was superior for participants in the visual presentation conditions (planned comparison: \(F(1,44) = 5.71; p = .021; \eta^2 = .225)\).

2.3. Discussion

Our findings reveal different learning outcomes depending on the delay of testing: when performance was tested immediately after presentation (as is usually the case in experiments on multimedia learning), participants in the auditory group significantly outperformed those in the visual group on the transfer test and had a tendency to do so on the retention test. However, when participants were tested one week after viewing the learning materials, the modality difference vanished entirely for the retention test and even reversed for the transfer test, which is, according to Gini (2005) and Mayer (2009), the most sensitive measure of learning outcomes. These results cast doubt on the assumption that auditory presentation of texts, as compared to visual presentation, improves long-term learning.

A potential caveat of our procedure, however, is the classroom setting. Even though the setting improves ecological validity, it is possible that certain features of the learning sessions or the tested classes that are unrelated to our manipulations contributed to the group differences. In addition, the sample was rather small and only the classes but not the individual participants were randomly assigned to the conditions. It is therefore warranted to replicate these findings in a more rigorously controlled laboratory setting.

3. Experiment 2

3.1. Method

3.1.1. Design and participants

The experiment followed a two-factorial design, in which text modality (auditory vs. visual) and test delay (immediate test vs. test after one week) were varied between participants. Performance on a retention test and a transfer test served as dependent variables. 124 undergraduates (mean age: 21.0; native language: German) from different majors at the University of Erfurt participated in the experiment for course credit or a small honorarium and were randomly assigned to the four conditions. We assigned 30 participants to each of the immediate conditions and 32 to the delayed conditions to compensate for potential dropout in the delayed tests. Six participants did not return for the second session. Thus, the final sample included 27 participants in the visual delayed and 31 participants in the auditory delayed group.

3.1.2. Materials and procedure

Participants were tested in individual sessions. Each participant was invited for two experimental sessions separated by one week. Prior to the experiment, participants were informed that they would take part in a series of experiments so that participants (in the delayed conditions) could not guess beforehand that their performance would not be tested before the second session. The procedure was similar to the first experiment. Participants first completed the meteorology knowledge checklist at their own rates, followed by the completely computer-paced presentation of the learning materials. Finally, participants in the immediate conditions answered the same retention and transfer questions with the same timing as in Experiment 1. Participants in the delayed conditions were instead informed that their learning performance would be tested in the second session one week later. In the second session, they answered the retention and transfer questions in the same procedure as the participants with an immediate test. Unlike in Experiment 1, both question presentation and the mode of responding were computer-based.

3.1.3. Scoring

The retention and transfer scores were determined analogously to Experiment 1. All questionnaires were rated by two independent scorers, who were blind to the participants’ experimental condition. We reached a reasonably high inter-rater reliability for the retention \((r = .89)\) and transfer tests (Question 1: \(r = .94, Question 2: r = .99, Question 3: r = .96, Question 4: r = .95)\).

3.2. Results

We analyzed the data in the same manner as in Experiment 1 and will report the results analogously. Means and 95\% confidence intervals are reported in Table 2.

3.2.1. Prior knowledge

An ANOVA revealed no significant differences between the four experimental conditions regarding prior knowledge \((all F < 1.3)\).

3.2.2. Learning outcomes

The MANOVA revealed significant main effects for text modality with an advantage for written over spoken text \((Wilks’ \lambda = .92; F(2,114) = 4.85; p < .01; \eta^2 = .078)\) and delay in that immediate performance was better than delayed performance \((Wilks’ \lambda = .82; F(2,114) = 12.18; p < .001; \eta^2 = .176)\). The interaction between modality and delay did not reach significance \((Wilks’ \lambda = .99; F < 1; \eta^2 = .06; \eta^2 = .008)\).

In a 2 x 2 between-subjects ANOVA on the retention scores, the main effects for text modality \((F(1,115) = 7.51; p = .007; \eta^2 = .061)\) and delay \((F(1,115) = 23.94; p < .001; \eta^2 = .172)\) also reached significance. Participants who read the text recalled more contents \((6.95, Cl = 6.28, 7.62)\) than participants who listened to the text \((5.66, Cl = 5.02, 6.30)\), and recall performance declined substantially from the immediate test \((7.45, Cl = 6.80, 8.10)\) to the retention test after one week \((5.16, Cl = 4.49, 5.82)\). However, there was no
interaction between text modality and test delay \((F < 1)\). We additionally ran planned comparisons to investigate the modality effect in the standard condition, i.e. with an immediate test, and in the delayed condition separately. While there was no significant difference between the auditory and the visual condition in the immediate test \((F(1,115) = 2.49; p = .12)\), the visual group performed significantly better than the auditory group on the delayed retention test \((F(1,115) = 5.24; p = .024)\).

For transfer scores there was also a significant main effect of modality, \(F(1,115) = 6.45; p = .012; \) \(\eta^2_p = .053\). Transfer performance was better with written than with oral text \((4.66, CI = 4.17, 5.15 \text{ vs. } 3.79, CI = 3.32, 4.26, \text{ respectively})\). However, neither the main effect of delay \((F(1,115) = 1.90; p = .17; \text{ and } \eta^2_p = .016)\) nor the interaction reached significance \((F < 1)\). Additional planned comparisons revealed a non-significant difference between the visual and the auditory condition on the immediate test \((F(1,115) = 1.27; p = .26)\), but a significant written advantage when the transfer test was administered one week after the learning session \((F(1,115) = 5.99; p = .016)\).

3.3. Discussion

In this experiment with higher statistical power and more experimental control, we found inverse modality effects for both transfer and retention. Moreover, we did not obtain significant interactions with delay, which indicates that in this experiment, the written advantage occurred irrespective of the delay between learning and test. This finding is surprising given the many observations of modality effects in immediate tests and it also contrasts with our first experiment. It is implausible that the non-occurrence of an immediate modality effect is due to a lack of power, since the current experiment included a lot more participants than our first experiment, in which an immediate modality effect was observed. Moreover, an analysis of the delayed groups revealed a significant inverse modality effect. In recent years, there is, however, an increasing number of studies that did not obtain a modality effect on an immediate test, even in conditions in which it is predicted based on the CMTL and the CLT (e.g., the studies assembled in a special issue on null-effects of modality edited by Schüler, Scheiter, & Schmidt-Weigand, 2011). This suggests that the modality effect is not as robust as previously assumed (see also the distribution of effect sizes and their standard errors in the funnel plot reported by Lindow et al. (2011), p. 241). The absence of a modality effect in the groups with immediate tests is also reflected in the fact that the interaction between text modality and test delay did not reach significance, in contrast to Experiment 1.

However, the findings corroborate those of Experiment 1 in that long-term learning was consistently better when participants had seen an animation with written text compared to an animation combined with spoken text. We can therefore conclude that the inverse modality effect that was found in the classroom setting was not simply due to other differences between the classes and the testing situations. This indicates that the modality effect does not expand to long-term learning.

4. General discussion

We tested the assumption that when learning performance is tested after a delay of one week, performance is better when pictorial information is combined with written rather than oral text. This contradicts typical recommendations for the design of multimedia learning materials. The conditions that were applied here meet the known boundary conditions for the modality effect (e.g., Givens, 2005): written but not auditory text creates a split-attention situation due to simultaneous presentation; the pacing of the materials is experimenter-determined; and texts are short. Our results cast doubt on the assumption that the auditory as compared to visual presentation of texts improves long-term learning. Instead, they provide evidence for the assumption that combining pictures or animations with written text serves as a desirable difficulty that leads to superior long-term learning and memory as compared to spoken text.

In recent years, evidence has accumulated that not even the immediate modality effect is as general as previously assumed (e.g., Gyselinck-Janet, & Dubois, 2008; Schmidt-Weigand, 2011; Schüler, Scheiter, & Gerjets, 2011; Schüler et al., 2011; Tabbers et al., 2004; Tabbers & van der Spoel, 2011). The findings of the few studies that have tested the influence of text modality on long-term multimedia learning are surprisingly consistent in that auditory text presentation never led to better performance in delayed tests of learning; instead, it was either equivalent or inferior to visual text presentation. The evidence on long-term learning that exists so far thus suggests that the durability of the modality effect is a further reason for caution and that the recommendation to use oral texts should be restricted to situations in which immediate comprehension is essential (in addition to further boundary conditions). In case it is more important to remember what has been comprehended initially, written text presentation should be preferred. Finding out whether knowledge has been successfully consolidated, however, is of major importance to the study of learning and should be one focus of studies on multimedia learning as well as of theoretical considerations.

Compared to the large body of data on the immediate modality effect, the current study (combined with the few other studies that have included delayed tests) provides only a first step toward understanding the long-term influences of text modality. The learning materials and the settings do not yet allow for generalization. Also, the mechanisms that contribute to the delayed written advantage need to be investigated further. Theoretical accounts of desirable difficulties are still far from comprehensive. The broadest explanation relates to the levels-of-processing approach, according to which the durability of memory traces is a positive function of processing depth (Craik & Lockhart, 1972). For the processing of written texts and graphics the need to effortfully integrate the sequentially processed verbal and pictorial information is usually interpreted as hindering learning, but it may in fact induce the deep processing that helps long-term learning. Research on text processing also suggests that (particularly long and/or difficult) texts are processed more deeply in the visual than in the auditory
modality (Green, 1981; Kürschner, Schnottz, & Eid, 2006; Sanders, 1973). Oral texts are transient and their comprehension is necessarily externally paced (see also Wong, Leahy, Marcus, & Sweller, 2012). In contrast, written text is usually available for re-inspection and reading speed is usually self-determined. Consequently, written text presentation allows individuals to devote more time to the difficult parts of the text and re-read them as well as to read easier parts of the text faster—both of which is not possible with spoken text. These options should be particularly beneficial when the learner can freely determine processing speed (as was the case in the studies by Segers and colleagues), but can even be helpful when pacing is externally determined. These benefits of reading might therefore be a reason why learner-vs.-system-pacing is a boundary condition of the modality effect with immediate learning (e.g., Taibbers et al., 2004) and text length is another (Rummer et al., 2011; Schüler, Scheiter, & Gerjets, 2013; Wong et al., 2012). Consequently, the mechanisms that may be responsible for the long-term written advantage may also apply to modality influences on immediate tests.

Our data suggest that combining pictorial information and written text is beneficial for long-term learning, contrary to what has been suggested in the modality principle of multimedia learning. It is, however, an open question whether such a reversal of the standard recommendation also applies to other principles of multimedia learning. A couple of studies suggest that this is not the case for the multimedia effect (e.g., Peck, 1974; Scheiter, Schüler, Gerjets, Huk, & Hesse, 2014; Schweppe, Eitel, & Rummer, 2015). However, since only very few studies so far have included delayed tests, further research is necessary to determine the long-term implications of other influential design principles.

With respect to the desirable difficulties idea, it is important to determine and predict when and which kinds of difficulties are desirable (see also McDaniel & Butler, 2011; Schnottz & Kürschner, 2007). It is also still important to design learning materials that avoid cognitive overload. In analogy to the “garbage in, garbage out” principle: if the learner was unable to comprehend the materials in the first place (due to overly complex content or an overly demanding design), there is nothing (sensible) that can be remembered. However, once overload is avoided, it does not appear to be advisable to reduce cognitive load per se. Here, the theories of multimedia learning need to be revised. These theories see learning mainly as a function of comprehension (for a similar argument see Schweppe et al., 2015; Schweppe & Rummer, 2014). Therefore, it is only natural that working memory as well as how it can be used to its full capacity are the focus. However, based on the current findings, we suggest that the focus should be broadened to long-term memory and processes of consolidation (or forgetting, respectively) and retrieval (see also Schweppe & Rummer, 2014, 2016).

References
Schweppe, J., & Rummer, R. (2016). Attention and long-term memory in system and...


