Contribution

On the Robustness of the Modality Effect: Attempting to Replicate a Basic Finding1

Stefanie Lindow1, Heather M. Fuchs1, Anne Fürstenberg1, Janet Kleber2, Judith Schweppe1 and Ralf Rummer1

1Department of Psychology, University of Erfurt, 2Department of Psychology, University of Vienna

Abstract. Two experiments were conducted to replicate the modality effect and to test two opposing theoretical explanations: From the visuo-spatial load hypothesis (as implied by the Cognitive Theory of Multimedia Learning), we derived the prediction that the modality effect depends on the need to simultaneously store and process text and pictorial information in the same working memory subsystem, which implies that the modality effect is not influenced by text length. However, based on the concurrent auditory recency hypothesis, we expected the modality effect to occur with very short texts only, irrespective of whether a picture is present. Adopting a German translation of the learning materials used by Moreno and Mayer (1999), we failed to detect a modality effect for sequential text-picture-presentation in Experiment 1. In Experiment 2, a modality effect was expected for simultaneous presentation of the same texts and pictures, yet again remained absent. We conclude with the suggestion that the apparent robustness of the modality effect reported in the literature might be explained – at least in part – by publication bias.

Keywords: modality effect, multimedia learning, publication bias

In numerous studies investigating learning with texts and pictures, a learning advantage for auditory over visual text presentation has been reported (cf. Ginns, 2005, for a meta-analysis; the introduction to this issue by Schüler, Scheiter & Schmidt-Weigand, 2011, for an overview). This so-called modality effect is attributed to two simultaneously yet independently operating principles: one captured by the split-attention assumption, which attributes the effect to limitations of the perceptual system and the resulting cognitive consequences; and the other by the visuo-spatial load hypothesis, which pinpoints the effect in working memory (e.g. Mayer, 2009; the term, however, was introduced by Rummer, Schweppe, Fürstenberg, Seufert & Brünken, 2010). The split-attention assumption ac-

1 Dieser Beitrag wurde unter der geschäftsführenden Herausgeberschaft von Jens Möller angenommen.
counts for the occurrence of the modality effect in the case of simultaneous text-picture-presentation when a picture and visual text cannot be perceived conjointly, whereas a picture and auditory text can (cf. the introduction to this issue by Schüler, Scheiter & Schmidt-Weigand, 2011, for a detailed description). The visuo-spatial load hypothesis also predicts a modality effect when texts and pictures are presented sequentially: i.e. because pictures and visual texts are (at least initially) processed in the same subsystem of working memory (i.e. visuo-spatial system). This is assumed to go along with an overload of the visuo-spatial subsystem when visually presented texts and (visual) pictures must be stored and processed. In contrast, this overload does not occur when pictures are accompanied by auditory texts, because auditory text draws upon the capacity of a different working memory subsystem.

With reference to the original conception of Baddeley’s (e.g. 2000) working memory model, this explanation has recently been called into question (e.g. Gyselinck, Jamet & Dubois, 2008; Rümmer, Fürstenberg & Schweppe, 2008; Rümmer et al., 2010). According to Baddeley’s model, visually presented text directly enters the verbal subsystem through subvocal articulation by the rehearsal component of the phonological loop. Thus, independent of presentation modality, both visually and auditory presented texts are stored and processed in the phonological loop – a subsystem dedicated not only to auditory-verbal input but rather to verbal input in general regardless of modality. According to Baddeley’s (2000) conception of working memory, no modality-specific overload in the visuo-spatial subsystem is predicted when learning with texts and pictures. This is not in line with the explanation of the modality effect given by the visuo-spatial load hypothesis. In summary, the visuo-spatial load hypothesis is challenged by the working memory model on which it is based.

Going along with this uncertainty regarding the theoretical basis of the modality effect, empirical investigations also raise doubts concerning its generality and robustness. A recently conducted meta-analysis (Ginns, 2005) confirms the large modality effect reported by Moreno and Mayer (1999) for simultaneous (Experiment 1; cf. also e.g. Mayer & Moreno, 1998) and sequential text-picture-presentation (Experiment 2; cf. also Mousavi, Low & Sweller, 1995). However, other studies yield contradictory results (e.g. studies not included in the analysis: Tiene, 2002; Baggett & Ehrenfeucht, 1983; cf. the introduction to this issue for an overview). The meta-analysis also highlights potential boundary conditions of the effect. For example, Ginns (2005) found a significant moderator effect for system vs. self-paced presentations. While the effect of the former was large ($d = .93$), that for the latter was small and negative ($d = -.14$). Although one must interpret the results of this meta-analysis with caution due to the small sample sizes in the moderator groups as well as for the reasons discussed below, the meta-analysis provides evidence that whenever the impairments in the learning process caused by split-attention can be compensated for (e.g. by sufficient time to process the learning material), no modality effect occurs (cf. Tabbers, 2002).

If, however, the learning advantage of auditory over visual text presentation is indeed merely due to split-attention, no modality effect should occur in the case of sequential text-picture-presentation. Among the few studies using this presentation format, there is evidence both for (Moreno & Mayer, 1999; Mousavi et al., 1995) and against (Tiene, 2002; Baggett & Ehrenfeucht, 1983) a modality effect. Rümmer and colleagues argue that in the sequential presentation studies reporting a learning advantage for auditory over visual text presentation, a modality effect in terms of the visuo-spatial load hypothesis did not occur but rather an auditory recency effect (e.g. Rümmer, Schweppe, Fürstenberg, Scheiter & Zindler, 2011). This effect, which stems from basic memory research, refers to the memory advantage of auditory over visual text presentation when memorizing lists of letters, numbers or words – but only applies to the last and second to last items (e.g. Crowder & Morton, 1969; Penney, 1989). Additionally, this advantage occurs regardless of whether a picture must be processed together with the list (cf. Penney, 1989 for an overview of explanations of the effect). Recently, these findings have been generalized to sentence recall (cf. Rümmer et al., 2011, for an overview): for sentences, too, the last words are remembered better for auditory as compared to visual text presentation, which enhances – relative to visual presentation – the recall of the entire sentence.

The studies reporting an occurrence of the modality effect for sequential text-picture-presentation used only very brief, typically single-sentence texts to accompany pictures. Better memory of those learning texts (caused by the auditory recency effect) could lead to a more adequate integration of sentence and picture content compared to visual text presentation. This should, in turn, enhance understanding of the whole learning material. In the case of longer texts, as used in the studies failing to report a modality effect with sequential text-picture-presentation, the auditory recency effect is limited to the last sentence and thus no substantial learning advantage for the text as a whole should occur (i.e. no modality effect is expected). This auditory recency hypothesis (cf. Rümmer et al., 2010), therefore, attributes the modality effect with sequential text-picture-presentation to the occurrence of an auditory recency effect. In this way, the effect should be restricted to very brief texts (or to the last sentence of longer texts); and its occurrence should not depend on the presence of pictures.

---

2 Logie, Della Sala, Wynn and Baddeley (2000) published a study demonstrating that the visuo-spatial working memory system contributes to the retention of visually presented verbal stimuli under some circumstances. However, Fürstenberg, Rümmer and Schweppe (subm.) show that such a contribution is restricted to situations in which the phonological loop is blocked by an articulatory suppression task.
These different dependency patterns of the modality effect on certain characteristics of the learning materials – namely, text length and picture presence – derived from the visuo-spatial load and auditory recency hypotheses have been tested against each other previously. Rummer and colleagues presented participants with learning materials containing pictures accompanied by either audio or visual learning texts comprised of five sentences each (Rummer et al., 2008; Rummer et al., 2011). In contrast to earlier studies measuring learning success as a whole, their experimental material allowed for testing retention of each of the five text positions separately. The results clearly support the predictions of the auditory recency hypothesis: A significant advantage for auditory over visual text presentation is restricted to the last sentence of each text. For the second and third sentence positions, a descriptive advantage for visual text presentation was even observed. Moreover, these results were replicated when learning texts were presented without pictures (Rummer et al., 2011). These results suggest that the occurrence of the learning advantage of auditory over visual text presentation is influenced by text length but independent of the presence of pictures. These findings, therefore, support the auditory recency hypothesis and raise doubts regarding the existence of the modality effect in terms of the visuo-spatial load hypothesis.

Note that the generalizability of these findings to multimedia research may be limited due to the content of the learning material used. Contrary to the object of multimedia research, the study from Rummer and colleagues (2011) did not examine the learning of cause and effect systems that require the extraction and integration of information from different sources. However, other studies investigating the learning process (e.g., eye-movements) with the original learning material from Mayer and colleagues (Mayer & Moreno, 1998; Moreno & Mayer, 1999) also yield inconsistent results (e.g., replication of the modality effect: Schmidt-Weigand, Kohnert & Glowalla, 2010b; no replication: Schmidt-Weigand, 2011; Schmidt-Weigand, Kohnert & Glowalla, 2010a; Tabbers & van der Spoel, 2011).

Taken together, the results and empirical uncertainty surrounding the modality effect that, in sum, questions its robustness and generality. Factors moderating the occurrence of the modality effect still remain unclear (e.g., text length and picture presence as derived from the visuo-spatial load and auditory recency hypotheses). Building on these shortcomings, the aim of our research activities was twofold. On the one hand, we aimed to complement the field of research on the modality effect with additional data by conducting replications of the original studies on the effect (cf. the introduction to this issue for a discussion on the relevance of replication e.g., for meta-analyses). To do so, we adopted the multimedia learning material of Moreno and Mayer (1999) – a set of materials for which the authors observed strong modality effects with simultaneous (replicated in Experiment 2) as well as sequential text-picture-presentation (replicated in Experiment 1).

On the other hand, we aimed to clarify the inconsistent results of research on the modality effect reported thus far. In this sense, Experiment 1 was conducted not only as a replication but also with the goal of investigating whether the learning advantage of auditory over visual text presentation reported for sequential text-picture-presentation is due to a modality effect in terms of the visuo-spatial load hypothesis or merely caused by an auditory recency effect. Specifically, we tested whether the originally reported modality effect disappears when we alter text length for each animation (single vs. multiple sentences). The original material consisted of brief, almost exclusively single-sentence texts for each accompanying (animated) picture. We combined three to four of these originally brief learning texts, as well as their brief animations, in one multiple-sentence text with correspondingly long animation. Additionally, we investigated whether the reported modality effect depends on the presence of pictorial information in the learning material (animation present vs. absent). Thus, Experiment 1 allows us to directly test the opposing theoretical explanations of the modality effect against each other: Following the auditory recency hypothesis, the modality effect should occur for single-sentence texts (i.e., the original material) and disappear when texts are presented as blocks of multiple sentences. This pattern should occur regardless of whether animations are included in the learning material. Note that in the condition without animations, the memory advantage of auditory text presentation does not support the integration of text with the subsequently presented pictorial information. Still, a modality effect is expected for single-sentence texts because each learning text is followed by a break (while the animation is presented in the corresponding condition) in which the processing of the learning text should benefit from this auditory memory advantage. On the contrary, according to the visuo-spatial load hypothesis, it is predicted that, regardless of text length, the modality effect will occur in all experimental groups with animations and be absent in experimental groups without animations.

Table 1 summarizes these predictions. Note that the first row in Table 1 represents the replication conditions of the original study from Moreno and Mayer (1999) for which both opposing hypotheses would predict the occurrence of the modality effect.

<table>
<thead>
<tr>
<th>Text length</th>
<th>Picture presence</th>
<th>Visuo-spatial load hypothesis</th>
<th>Auditory recency hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single sentences</td>
<td>Present</td>
<td>ME</td>
<td>ME</td>
</tr>
<tr>
<td>Absent</td>
<td>no ME</td>
<td>ME</td>
<td></td>
</tr>
<tr>
<td>Multiple sentences</td>
<td>Present</td>
<td>ME</td>
<td>no ME</td>
</tr>
<tr>
<td>Absent</td>
<td>no ME</td>
<td>no ME</td>
<td></td>
</tr>
</tbody>
</table>
the modality effect. The remaining rows refer to our modifications of the original presentation procedure.

1 Experiment 1

Experiment 1 aimed at replicating the modality effect with sequential text-picture-presentation reported by Moreno and Mayer (1999; Experiment 2) and additionally examined the cognitive mechanisms behind the effect’s occurrence (i.e., we tested the visuo-spatial load hypothesis against the auditory recency hypothesis). We restricted our replication to the sequential conditions in which the learning text precedes the animation because no order effect was reported in the original study.

1.1 Method

1.1.1 Participants and design

215 undergraduates (178 female; mean age: 22.2; native language: German) from different majors at the University of Erfurt participated in the experiment lasting approximately 45 min. They received either course credit or payment for participation.

The experiment followed a 2 (text modality: auditory vs. visual) x 2 (animation presence: present vs. absent) x 2 (text length: single or multiple sentences) between-subjects design.

1.1.2 Materials

We used Moreno and Mayer’s (1999) original study materials comprised of a knowledge checklist to assess prior meteorology knowledge, a computerized multimedia presentation of a lesson on lightning formation, and a retention and transfer questionnaire to measure learning success (we refrained from using the visual-verbal matching test originally used as a third dependent variable because the measure had been shown to be of little use due to ceiling effects; Moreno & Mayer, 1999). All verbal materials were translated into German by a native German speaker and counterchecked by a native English speaker.

1.1.2.1 Meteorology knowledge checklist

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

1.1.2.2 Learning material

We adopted the original learning material used by Moreno and Mayer (1999) describing the process of lightning formation in 16 brief texts (14 single sentences, two double-sentence texts, word count in original material: 286; word count in translation: 275) and corresponding short animations as follows. We assembled eight different learning material sets using Microsoft Power Point. As depicted in Figure 1, each set started with the presentation of the first piece of learning text. In the visual text condition, the learning text was presented in the centre of the screen, while in the auditory text condition, a fixation cross was shown while the text was read to the participant at a normal speech rate with 261.5 syllables per minute by a female speaker via headphones (following Moreno & Mayer, 1999, the on-screen text was displayed for the duration of the spoken text).

Directly following the presentation of the learning text, the corresponding animation was shown to participants in the experimental conditions with animations present. Thus, we implemented the sequential text-picture-presentation format used by Moreno and Mayer (1999; Experiment 2). In the experimental conditions without animations, participants saw a fixation cross in the centre of the screen for the duration of the omitted animation. The animation or fixation cross (depending on experimental condition) was then followed by the next part of learning text, and so on, resulting in a text-animation sequence lasting 312 seconds in total for each of the experimental groups.

For our experimental conditions with single sentences, this sequence consisted of the 16 short text-plus-animation pairs as presented in the original study from Moreno and Mayer. For conditions with multiple sentences, we combined three to four of these short texts to build 5 text blocks consistent in terms of content (presentation time corresponded to the single sentence condition).³

1.1.2.3 Dependent variables transfer and retention tests

To measure learning success, 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 one open question. The transfer test consisted of four open questions asking for an explanation of one specific transfer problem each.

³ Following the numbering of Moreno and Mayer (1999), the original learning texts were blocked for the multiple sentence conditions of the current study as follows: text block 1 = texts 1 to 3, block 2 = 4 to 7, block 3 = 8 to 10, block 4 = 11 to 13, block 5 = 14 to 16.
### 1.1.3 Procedure

According to the procedure described by Moreno and Mayer (1999), participants were tested in small groups of 1 to 10 persons per session in individual cubicles. First, participants completed the meteorology knowledge checklist at their own rates, followed by the entirely computer-paced presentation of the learning materials. At the beginning, participants were informed that they would view a presentation on how lightning works and then answer a series of questions. After viewing the presentation, participants completed the retention and transfer tests. Each question was printed on a separate sheet of paper and completed one after the other by the participants. Participants first worked on the retention questionnaire for 5 minutes, followed by the four transfer questions with a processing time of 3 minutes each. At the end of each session, participants were debriefed regarding the aim of the study and compensated for participation.

### 1.1.4 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.

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 interrater reliability for the retention (\(r = .96\)) and transfer tests (\(r = .80\)).

### 1.2 Results

Results are reported separately for transfer and retention scores. In addition to the analysis of the modality effect in the replication conditions (i.e. visual vs. auditory single sentence texts with animations present) via independent samples t-tests, we report analyses of variance (ANOVAs) with text length, animation presence, and text modality as between-subjects factors: Following the visuo-spatial

---

<table>
<thead>
<tr>
<th>Text length</th>
<th>Picture presence</th>
<th>Text modality visual</th>
<th>Text modality auditory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single sentences</td>
<td>Present</td>
<td><img src="image1" alt="Visual Example" /></td>
<td><img src="image2" alt="Auditory Example" /></td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td><img src="image3" alt="Visual Example" /></td>
<td><img src="image4" alt="Auditory Example" /></td>
</tr>
<tr>
<td>Multiple sentences</td>
<td>Present</td>
<td><img src="image5" alt="Visual Example" /></td>
<td><img src="image6" alt="Auditory Example" /></td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td><img src="image7" alt="Visual Example" /></td>
<td><img src="image8" alt="Auditory Example" /></td>
</tr>
</tbody>
</table>

*Figure 1. Learning material sets used in the 8 experimental groups with text length (single vs. multiple sentences), animation presence (present vs. absent), and text modality (visual vs. auditory) manipulated.*
load hypothesis, an interaction of the factors text modality and animation presence is expected (i.e. a modality effect only when animations are present), whereas no interaction of the factors text modality and text length should occur (i.e. a modality effect irrespective of text length). According to the auditory recency hypothesis, an interaction of the factors text modality and text length (i.e. a modality effect only with brief texts) and no interaction of the factors text modality and animation presence (i.e. a modality effect irrespective of animation presence) are expected.

1.2.1 Sample size and data reduction due to prior knowledge

An a-priori power analysis was conducted to ensure sufficient power for replicating the modality effect in the replication conditions: We used the effect sizes reported by Moreno and Mayer for the modality effect (1999; Experiment 2; $d_{\text{retention}} = 0.94, d_{\text{transfer}} = 1.09$) as expected population effect sizes, set $\alpha = .05$ and aimed at a power of $1-\beta = .90$. For the smallest expected effect in the retention test, this resulted in a necessary sample size of 21 participants in each experimental group. Moreno and Mayer (1999) included only low experience participants in their analyses, referring to previous research demonstrating that the modality effect is stronger for low compared to high experience learners. Following their procedure, we considered only the data of the 167 participants classified as having low experience in meteorology, as assessed by a knowledge checklist, in our analyses (cell distribution cf. Table 2). To accompany this main analysis of the reduced data-set, we used the data of all 215 participants to additionally assess the influence of prior knowledge on retention and transfer test performance.

### Table 2

<table>
<thead>
<tr>
<th>Animation</th>
<th>Text Modality</th>
<th>Text Length</th>
<th>Text Modality</th>
<th>Text Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Auditory</td>
<td>Single</td>
<td>7.67 (2.80) [21]</td>
<td>3.67 (1.39) [21]</td>
</tr>
<tr>
<td>Absent</td>
<td>Visual</td>
<td>Single</td>
<td>8.76 (2.63) [21]</td>
<td>4.43 (1.89) [21]</td>
</tr>
<tr>
<td>Present</td>
<td>Auditory</td>
<td>Multiple</td>
<td>7.86 (2.69) [21]</td>
<td>3.52 (1.36) [21]</td>
</tr>
<tr>
<td>Absent</td>
<td>Visual</td>
<td>Multiple</td>
<td>8.86 (3.62) [22]</td>
<td>4.14 (1.70) [22]</td>
</tr>
</tbody>
</table>

### 1.2.2 Retention test

As can be seen in Figure 2, the mean difference between auditory and visual text presentation corresponds to a reversed modality effect in the two replication conditions, $t(40) = 1.31, p = .20, g = 0.404$. According to this finding for the replication conditions, the visual groups outperformed the auditory groups in all further conditions. This resulted in a small main effect of text modality, $F(1, 159) = 7.19, p = .01, \eta^2 = .043$. The main effect for animation presence revealed a significant advantage for the multimedia conditions ($M = 8.29, SD = 2.96, n = 85$) over conditions without animations ($M = 6.96, SD = 3.13, n = 82$), $F(1, 159) = 7.96, p = .01, \eta^2 = .048$. Neither the main effect of text length, $F(1, 159) = 1.66, p = .20, \eta^2 = .010$, nor the interaction effects were close to the significance threshold (animation presence x text length: $F(1, 159) = 2.57, p = .11, \eta^2 = .016$; all others $F < 1, \eta^2 \leq .001$).

### 1.2.3 Transfer test

A look at the means of the transfer test scores shown in Figure 2 reveals that no modality effect was found in the two replication conditions. We even found slightly higher transfer test scores for visual ($M = 4.43, SD = 1.89, n = 21$) as compared to auditory text ($M = 3.67, SD = 1.39, n = 21$).
presentation, $t(40) = 1.49, p = .14, g = 0.460$. As depicted in Table 2, this reversed pattern was found in all experimental conditions and constitutes a small main effect for modality, $F(1, 159) = 3.96, p = .05, \eta^2 = .024$. The main effect for animation presence reveals a significant advantage for the multimedia conditions ($M = 3.94, SD = 1.61, n = 85$) over conditions without animations ($M = 3.24, SD = 1.88, n = 82$), $F(1, 159) = 6.54, p = .01, \eta^2 = .039$. In brief, neither the main effect for text length nor any of the interactions reached significance, all $F < 1; \eta^2 \leq .002$.

### 1.2.4 Influence of prior knowledge

We found a small correlation between participants’ prior knowledge and retention test scores ($r = .183, p = .01$) and no correlation with their transfer test scores ($r = .033, p = .63$). As indicated by independent samples $t$-Tests, high experience participants (transfer: $M = 3.85, SD = 1.64, n = 48$; retention: $M = 8.63, SD = 3.39, n = 48$) and low experience participants (transfer: $M = 3.60, SD = 1.78, n = 167$; retention: $M = 7.64, SD = 3.11, n = 167$) significantly differ in retention, $t(213) = 1.89, p = .03, g = 0.310$, but not transfer test performance, $t(213) = 0.89, p = .19, g = 0.146$.

Despite the significant influence of prior knowledge on retention test performance, we found comparable results based on repetitions of the above reported analyses with ANCOVAs using prior knowledge as covariates. There was again a non-significant reversed pattern of the modality effect in the replication conditions (auditory: $M = 8.04, SD = 2.94, n = 28$; visual: $M = 9.21, SD = 3.05, n = 28$), $F(1, 53) = 2.62, p = .11, \eta^2 = .047$. The ANCOVA including all experimental conditions resembled the respective ANOVA for the low prior knowledge participants (main effect modality: $F(1, 205) = 9.57, p < .01, \eta^2 = .045$; main effect animation presence: $F(1, 205) = 7.80, p < .01, \eta^2 = .037$; main effect text length: $F(1, 205) = 1.58, p = .21, \eta^2 = .008$; all interactions $F < 1, \eta^2 < .002$). In sum, we find the descriptive pattern of a reversed modality effect in the replication conditions and a small, significant reversed modality effect over all experimental groups.

### 1.3 Discussion

Contrary to the expectations derived from the visuo-spatial load and auditory recency hypotheses, we were unable to replicate the modality effect reported by Moreno and Mayer (1999, Experiment 2) with sequential text-picture-presentation. The modality effect was found neither in the replication nor in our additional experimental conditions. Rather, we found a very small, significant advantage for
visual text modality. These results contradict the predictions derived from the visuo-spatial load and auditory re- cency hypotheses, in that both expect a modality effect for sequential text-picture-presentation in certain (divergent) conditions. To supplement and further investigate our non- effects on the modality effect, we intended to replicate the original finding of a modality effect with simultaneous text-picture-presentation in Experiment 2.

2 Experiment 2

Experiment 2 aimed at replicating the modality effect reported by Moreno and Mayer (1999; Experiment 1) with simultaneous text-picture-presentation (cf. also Mayer & Moreno, 1998). According to Moreno and Mayer (1999), a strong modality effect is expected for the simultaneous presentation format, because both a text-modality specific overload in working memory (visuo-spatial load hypothesis) and split-attention should jointly cause a learning advantage for auditory over visual text presentation. We restricted our replication to the condition in which the visual text is presented physically far from the animation (as opposed to text that is integrated into the animation) because a larger auditory advantage is expected under these circumstances (spatial-contiguity effect, e.g. Moreno & Mayer, 1999).

2.1 Method

2.1.1 Participants and design

Fifty undergraduates (35 female; mean age: 22.8; native language: German) from different majors at the University of Erfurt participated in the experiment lasting approximately 40 min. They received either course credit or pay- ment for participation.

Text modality (auditory vs. visual) was manipulated between subjects.

2.1.2 Material, procedure and scoring

Experiment 2 followed the same general and scoring procedures described in Experiment 1, except that texts and animations were presented simultaneously instead of sequentially. Thus, participants viewed the same animations as presented in Experiment 1, but this time continuously without the intermissions necessary with sequential presentation (total duration of the continuous animation: 192 s). In the auditory text presentation condition, the learning text was read to participants via headphones along with the animation. In the visual text presentation condition, the text was successively presented below the animation with the same words, timing and speed as in 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; the partitioning corresponds to the single sentence conditions of Experiment 1).

2.2 Results

Results are reported separately for the two dependent variables – retention (maximum score: 19) and transfer (no maximum score). We report descriptive statistics as well as independent samples t-tests comparing the auditory and visual text presentation conditions.

As described in Experiment 1, only the data of participants classified as having low experience in meteorology were used in analyses (N = 42 with equal cell distribution). An a-priori power analysis was performed to determine appropriate sample size: We used the effect sizes reported by Moreno and Mayer (1999; Experiment 1; d\text{retention} = 1.00, d\text{transfer} = 1.06) as expected population effect sizes, set α = .05 and aimed at a power of 1-β = .90. For the smallest expected effect in the retention test, this resulted in a necessary total sample size of N = 36. We complement this main analysis of the reduced data-set by separately analysing the influence of prior knowledge on transfer and retention test performance with the data of all 50 participants.

2.2.1 Retention test

Group means of retention test scores do not differ significantly between the conditions with visual (M = 8.57, SD = 3.23, n = 21) and auditory text presentation (M = 8.33, SD = 2.71, n = 21), t(40) = 0.26, p = .80, g = 0.080.

2.2.2 Transfer test

We also found no modality effect regarding the group means of transfer test scores. Instead, we even observed slightly better performance for the visual (M = 4.33, SD = 2.06, n = 21) as compared to auditory text presentation condition (M = 3.81, SD = 1.94, n = 21). However, this difference was far from significant, t(40) = 0.85, p = .40, g = 0.260.

2.2.3 Influence of prior knowledge

No correlations between participants’ experience scores and performance in retention (r = .023, p = .87) and transfer tests were found (r = .183, p = .20). Corresponding to this result, the 95% confidence intervals of means (CI) greatly overlap for low and high experience learners (retention: M\text{low} = 8.45, SD\text{low} = 2.95; 95% CI\text{low} [7.53, 9.37] vs. M\text{high} = 9.38, SD\text{high} = 3.07; 95% CI\text{high} [6.81, 11.94];
Studies were smaller (one to five participants). Nevertheless, we were equipped with a computer and headphones. Thus, regardless of group size, all participants were studying the learning material by themselves.

Regarding the participants (i.e. university students in all studies), it seems noteworthy that the percentage of high-experience learners that had to be excluded from analyses was more than twice as high in the replications (Experiment 1: 22%; Experiment 2: 18%) as in the original studies (Moreno & Mayer, 1999, Experiment 1: 8%, Experiment 2: 7%). This could be an indication of different levels of prior knowledge on meteorology between the student populations. We cannot rule out the possibility that the larger dropout of participants in our studies might have caused the samples to differ in other learner characteristics as well (e.g. general knowledge). In addition, our finding that prior knowledge correlates only weakly, if at all, with test performance possibly suggests that the questionnaire used was not an appropriate measure of prior knowledge for our participants. This, however, cannot explain the reversed modality effects found in the analyses of all participants regardless of prior knowledge, as the modality effect is expected to differ between high and low knowledge learners only in strength but not in direction (cf. Moreno & Mayer, 1999). To entirely rule out the discussed possible influences of prior knowledge on our results, our studies should be repeated with either a revised version of the prior knowledge measure or thematically different learning material (e.g. the functional principle of breaks: Mayer & Moreno, 1998).

In terms of the learning material used, it should be noted that a translation of the original material into German was necessary due to our participant sample. It is, indeed, conceivable that language (i.e. sentence structure, for example, in terms of the importance of the last word of a sentence for comprehension) is a boundary condition of the modality effect. Note, for example, that the other studies reported in this special issue that also fail to replicate the modality effect use translated learning material (German: Schmidt-Weigand, 2011; Schüler, Scheiter & Gerjets, 2011; Dutch: Tabbers & van der Spoel, 2011; but see Schmidt-Weigand et al., 2010b for a replication of the effect with German material). Studies comparing the effect of language on the occurrence of the modality effect could shed light on this potential moderating factor.

With regard to the learning material used, it is further noteworthy that, although the content of the learning material was congruent between the studies in question (i.e. we presented participants all 16 original texts; and participants viewed the steps of the animation exactly as reported by Moreno and Mayer, 1999), slight differences in duration did occur. Moreno and Mayer (1999) report an animation duration of 180 sec. The duration of text presentation is not explicitly reported but can be inferred to be less, learning conditions and, thus the intended audience (Moreno & Mayer, 1999: students who independently learn cause and effect systems) seem quite comparable: participants were seated in separate cubicles, each equipped with a computer and headphones. Thus, regardless of group size, all participants were studying the learning material by themselves.

Regarding the participants (i.e. university students in all studies), it seems noteworthy that the percentage of high-experience learners that had to be excluded from analyses was more than twice as high in the replications (Experiment 1: 22%; Experiment 2: 18%) as in the original studies (Moreno & Mayer, 1999, Experiment 1: 8%, Experiment 2: 7%). This could be an indication of different levels of prior knowledge on meteorology between the student populations. We cannot rule out the possibility that the larger dropout of participants in our studies might have caused the samples to differ in other learner characteristics as well (e.g. general knowledge). In addition, our finding that prior knowledge correlates only weakly, if at all, with test performance possibly suggests that the questionnaire used was not an appropriate measure of prior knowledge for our participants. This, however, cannot explain the reversed modality effects found in the analyses of all participants regardless of prior knowledge, as the modality effect is expected to differ between high and low knowledge learners only in strength but not in direction (cf. Moreno & Mayer, 1999). To entirely rule out the discussed possible influences of prior knowledge on our results, our studies should be repeated with either a revised version of the prior knowledge measure or thematically different learning material (e.g. the functional principle of breaks: Mayer & Moreno, 1998).

In terms of the learning material used, it should be noted that a translation of the original material into German was necessary due to our participant sample. It is, indeed, conceivable that language (i.e. sentence structure, for example, in terms of the importance of the last word of a sentence for comprehension) is a boundary condition of the modality effect. Note, for example, that the other studies reported in this special issue that also fail to replicate the modality effect use translated learning material (German: Schmidt-Weigand, 2011; Schüler, Scheiter & Gerjets, 2011; Dutch: Tabbers & van der Spoel, 2011; but see Schmidt-Weigand et al., 2010b for a replication of the effect with German material). Studies comparing the effect of language on the occurrence of the modality effect could shed light on this potential moderating factor.

With regard to the learning material used, it is further noteworthy that, although the content of the learning material was congruent between the studies in question (i.e. we presented participants all 16 original texts; and participants viewed the steps of the animation exactly as reported by Moreno and Mayer, 1999), slight differences in duration did occur. Moreno and Mayer (1999) report an animation duration of 180 sec. The duration of text presentation is not explicitly reported but can be inferred to be

---

S. Lindow et al.: On the Robustness of the Modality Effect 239

Z. Pädagog. Psychol. 25 (4) © 2011 Verlag Hans Huber, Hogrefe AG, Bern
120 sec. Since text presentation is shorter than the animation, parts of the animation (i.e. one third) are presented without text in the simultaneous condition. In our study, these values differ slightly: Our text presentation also lasts 120 sec (with comparable word count; Moreno & Mayer, 1999: 286; replication: 275); however, our animation is 12 sec longer than the original animation (i.e. 192 sec). Thus, in our simultaneous study (Experiment 2), a slightly longer portion of the animation was presented without concurrent text presentation. This could explain a possible reduced influence of split-attention in Experiment 2 as compared to the original study, but does not account for the absence of the modality effect based on visuo-spatial load in Experiment 1 and 2 (because a simultaneous presentation of text and animation is not necessary in this case).

Taken together, we cannot rule out the possibility that differences between our replication and the original studies may have caused the replication to fail due to unknown boundary conditions not considered when conducting the studies. However, it seems unlikely that these potential differences have caused large differences in the understanding and handling of the learning material between the replications and original studies, as a comparison of the absolute learning scores achieved by our participants to those in the original studies shows that overall test performance is quite comparable (cf. means and confidence intervals displayed in figure 2).

A thorough discussion of power issues can be found in the introduction to this special issue (Schüler, Scheiter & Schmidt-Weigand, 2011). We do not consider a further discussion with regard to our studies useful at this point, as our replication did not fail due to non-significant results but rather due to the direction of differences in means between the visual and auditory groups. Moreover, even a joint analysis of the two studies provided no evidence for an advantage of auditory text presentation.5

3.2 What can we learn from the results? – A critical look at the current state of research

Contrary to our initial aim, our research does not help to clarify the theoretical discussion on the cognitive mechanisms responsible for the modality effect. The failed replication of the effect in Experiment 1 with sequential text-picture-presentation as well as the finding of a reversed modality effect over all experimental groups contrast the visuo-spatial load and auditory recency hypotheses. The results of Experiment 2 contradict the visuo-spatial load hypothesis and split-attention assumption, which both predicted the occurrence of a modality effect with simultaneous text-picture-presentation. Overall, our results do not yield supportive evidence for the assumption of a general learning advantage of auditory over visual text presentation when learning with multimedia materials.

In addition to contradicting any explanation of the modality effect (i.e. visuo-spatial load hypothesis, auditory recency hypothesis, split-attention assumption), our results are inconsistent with a huge body of empirical research. Summarizing this research in a meta-analysis, Ginn’s (2005) reports a modality effect (cf. the introduction to this special issue for a summary of reviews on the effect with consistent results). Although meta-analyses aim to estimate population effect magnitudes on a more reliable basis than can be achieved with individual studies alone, they are, as argued in the introduction to this special issue, highly susceptible to error due to systematic exclusion of non-significant results. Indeed, there is growing evidence that psychological research is widely affected by this so-called publication bias (cf. Cooper, DeNeve & Charlton, 1997; Renkewitz, Fuchs & Fiedler, in press), which leads to an overestimation of the true magnitude of the effect in question. Together with the fact that our results, which raise doubts concerning the robustness and strength of the modality effect reported thus far, are in line with the findings of other replication studies (e.g. Schmidt-Weigand, 2011; Schmidt-Weigand et al., 2010a; Tabbers & von der Spoel, 2011), we see an urgent need to regard the summarization of earlier research on the modality effect by Ginn’s meta-analysis with caution. Thus, we tested the original meta-analytic dataset for the presence of publication bias.

The possibility of publication bias in the present meta-analysis is suggested by several aspects of the study selection that should be evaluated critically.6 Almost half of the studies included in the analysis (46 %) were published in a single journal, the Journal of Educational Psychology. This may be problematic, as there is evidence that journals may favour studies for publication that produce outcomes consistent with theories popularly published by that particular journal (cf. Palmer, 2000). Furthermore, in the present meta-analysis a preferential selection of significant results occurred. In most studies on the modality effect, multiple outcome measures (e.g. transfer, retention and matching

---

4 The simultaneous presentation of text and animation lasts 180 sec. For the sequential text-picture presentation format (duration: 300 sec), the presentation of text takes additional time compared to the animation presentation. Thus, it can be assumed that the difference of 120 sec between the duration of the two formats corresponds to the text presentation.

5 We combined Experiment 2 with the replication conditions of Experiment 1 and compared the auditory and visual text presentation conditions with independent samples t-tests – retention test: \( M_{\text{auditory}} = 8.00, SD_{\text{auditory}} = 2.74, n_{\text{auditory}} = 42, M_{\text{visual}} = 8.67, SD_{\text{visual}} = 2.91, n_{\text{visual}} = 42, t(82) = 1.08, p = .28, g = 0.236 \); transfer test: \( M_{\text{auditory}} = 3.74, SD_{\text{auditory}} = 1.67, n_{\text{auditory}} = 42, M_{\text{visual}} = 4.38, SD_{\text{visual}} = 1.95, n_{\text{visual}} = 42, t(82) = 1.62, p = .11, g = 0.354 \).

6 The following discussion refers only to the analysis of 39 between-design effect sizes. The four within-design effects were not considered, as this sample was not adequately large to assess publication bias.
test scores) were reported. Because effects included in meta-analyses must be independent, certain effects must be selected over others. In the case of the present meta-analysis, the results of the transfer test were preferred over other outcome measures, but only when significant. A non-significant transfer effect was only included when all other measures were also non-significant. Otherwise, the next best measure was preferred. This procedure led to the preference of four significant effects stemming from other measures over non-significant transfer effects – i.e. a biased selection in favour of significant results. Additionally, approximately 77% of effects included in the present meta-analysis were taken from peer-reviewed studies, only a small number of unpublished studies are included. The inclusion of unpublished studies is a central method to overcome publication bias (Rosenthal, 1994).

Given these considerations regarding the procedure of study selection, we tested the original meta-analytic dataset for the presence of publication bias by compiling a funnel plot (cf. Light & Pillemer, 1984), using the well-established procedure trim-and-fill analysis (Duval & Tweedie, 2000), and calculating relevant subgroup meta-analyses.

The funnel plot (cf. Figure 3) shows evidence of publication bias. The trim-and-fill analysis reinforces our interpretation of the funnel plot, indicating that 13 effects smaller than the estimated true population effect must be added to the meta-analytic sample in order to achieve a nonbiased sample. Adding these studies reduces the mean population effect estimate to $\bar{g} = 0.388$, 95% CI [0.157, 0.620] – a decrease in magnitude of approximately 47%. The moderator analyses also indicated the presence of bias. The true effect estimate based on unpublished studies was significantly smaller ($\bar{g} = 0.239$, 95% CI [–0.179, 0.656]) than that based on published studies ($\bar{g} = 0.888$, 95% CI [0.668, 1.108]). The direction of the difference between true effect estimates based on studies published in the Journal of Educational Psychology ($\bar{g} = 0.987$, 95% CI [0.772, 1.203]) and those published in other journals ($\bar{g} = 0.738$, 95% CI [0.311, 1.164]), although non-significant, is also consistent with the presence of bias.

The analyses of bias in combination with the other critical issues on the selection of studies in the meta-analysis discussed above raise concern regarding the reliability of its findings and indicate that the true magnitude of the modality effect was likely overestimated. The two studies reported in this paper go along with this critical interpretation of the state of research. Although we conducted an exact replication of the original studies demonstrating a strong modality effect (Moreno & Mayer, 1999), we are unable to report supportive evidence. In line with the argumentation raised in the introduction to this special issue, we argue that our results, despite the failure to replicate a well-documented effect, are indeed meaningful.

In conclusion, the results and discussion reported in this paper raise doubt whether the magnitude and generalisability of the modality effect is as great as currently as-
sumed and emphasise the necessity of identifying boundary conditions. In addition to deducing these boundary conditions from theory, we emphasise the necessity for researchers to take every finding on the modality effect seriously. By systematically considering both studies observing the effect and those failing to do so, potential boundary conditions can be identified (cf. e.g. Schüler, Scheiter & Gerjets, 2011: performance measures and text length) and directly manipulated in later research. The language of the learning material (e.g. English vs. German) is one possible starting point for further investigation that arises from our two experiments as well as other failed replication studies using a translation of the original learning materials (e.g. German: Schmidt-Weigand, 2011; Schmidt-Weigand et al., 2010a; Dutch: Tabbers & van der Spoel, 2011; but see Schmidt-Weigand et al., 2010b for a replication of the effect with German material). Only when we know the boundaries of the effect can we understand the underlying cognitive mechanisms and derive appropriate and differentiated guidelines for designing learning materials.

Acknowledgements

This research is supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) (Ru891/4-1). We would like to thank Nicole Briese, Janine Buchholz, Birgit Berkenhagen, Julia Müller, and Manja Bauer for their help in preparing the experiments, as well as Kristina Euler, Edvina Otto, Clemens Möckel, René Schlegelmilch, Carla Greving and Marie-Kristin Sommer for data collection. In addition, we thank Huib Tabbers for providing a Dutch version of the Moreno-Mayer learning materials.

References

Schüler, A., Scheiter, K. & Gerjets, P. (2011). Does the modality effect in multimedia learning appear only with text containing


Ralf Rummer
Department of Psychology
University of Erfurt
P.O. Box 900 221
DE-99105 Erfurt
E-mail: ralf.rummer@uni-erfurt.de