Applied Cognitive Psychology, Appl. Cognit. Psychol. **28**: 488–501 (2014)
Published online 7 February 2014 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.3004

Disfluency Meets Cognitive Load in Multimedia Learning: Does Harder-to-Read Mean Better-to-Understand?

ALEXANDER EITEL[1]*,†, TIM KÜHL[2,†], KATHARINA SCHEITER[1] and PETER GERJETS[1]

[1]*Knowledge Media Research Center, Tuebingen, Germany*
[2]*University of Mannheim, Psychology of Education, Mannheim, Germany*

Summary: In a series of four experiments, we examined the impact of disfluency in multimedia learning by testing contrasting predictions derived from disfluency theory and cognitive load theory against each other. Would a less legible text be beneficial to learning when accompanied by pictures, and what would be the role of less legible pictures? Students (N = 308) learned with text and pictures that were either easy-to-read (i.e., fluent) or harder-to-read (i.e., disfluent) about how a toilet flush works (Experiments 1–3) and about how lightning develops (Experiment 4). In line with disfluency theory, a disfluent text led to better performance in the transfer test and to more invested mental effort in Experiment 1. However, these beneficial effects could not be replicated in Experiments 2, 3, and 4, leaving open questions regarding the stability and generalizability of the disfluency effect, and thus raising concerns regarding its impact for educational practice. Copyright © 2014 John Wiley & Sons, Ltd.

INTRODUCTION

Learning with multimedia (i.e., text and pictures) has gained a lot of attention in recent years (cf. Mayer, 2009; Sweller, Ayres, & Kalyuga, 2011). To learn successfully with a multimedia instruction, students are usually required to actively select, organize, and integrate information from the instruction (cf. Mayer, 2009). These cognitive processes are carried out in working memory. Working memory resources, however, are limited (cf. Baddeley, 1992). In consequence, according to cognitive load theory (CLT; Sweller et al., 2011; Sweller, van Merriënboer, & Paas, 1998), instructional material should be designed in a way that unnecessary demands on working memory are avoided. According to the traditional conception of CLT, there are three sources that impose a load on working memory when learning with multimedia instruction (e.g., Sweller et al., 1998). First, a learner's working memory is loaded by the intrinsic characteristics of the instructional material (intrinsic cognitive load, ICL). ICL is conceptualized as the load on working memory that depends on the element interactivity (i.e., inherent complexity) of the instructional material, as well as the learner's level of expertise (or prior knowledge). Second, learners may experience a load that is caused entirely by the format of instruction (extraneous cognitive load, ECL). ECL refers to a working memory load that is caused by a poorly designed instruction, for instance, by presenting text and pictures in a format that requires learners to split attention (Ayres & Sweller, 2005). Because working memory resources are limited but crucial to understanding (multimedia) instruction, reducing ECL is considered being beneficial to learning (e.g., Sweller et al., 2011). Third, there is a desirable load on working memory, which is called germane cognitive load (GCL). GCL reflects cognitive processes that directly contribute to learning and a deeper comprehension of an instructional message (via schema construction and automation). This load may be triggered by increasing the complexity of the instructional procedure; for instance, by increasing variability when learning with worked-examples (e.g., Paas & van Merriënboer, 1994; Sweller et al., 1998). Thus, following the concept of GCL, introducing difficulties in the learning phase can sometimes be beneficial to performance, specifically as long as the amount of ICL and ECL does not overburden the limited capacity of working memory so that GCL can still be invested (e.g., De Croock, van Merriënboer, & Paas, 1998; Paas & van Merriënboer, 1994).

In a similar vein, according to research on 'desirable difficulties', intentionally introducing difficulties in the learning process is considered to be beneficial to learning by triggering a deeper processing of the learning contents (cf. Bjork, 2013; Craik & Lockhart, 1972), reflected by a higher amount of invested mental effort (Salomon, 1984). This has been confirmed in a number of empirical studies (e.g., deWinstanley & Bjork, 2004; Richland, Bjork, Finley, & Linn, 2005; Salomon, 1984). According to Alter, Oppenheimer, Epley, and Eyre (2007), introducing difficulties is assumed to stimulate deeper processing, not because of an increase in the objective difficulty, but because of an increase in the *perceived* difficulty of the task. How the perceived difficulty associated with a cognitive task, aptly named 'disfluency', may relate to learning outcomes is described within disfluency theory (Alter et al., 2007). At this, disfluency theory leans on considerations of William James (1890/1950), who stated that humans possess two distinct processing systems: one that is quick, effortless, associative, and intuitive (System 1) and another that is slow, effortful, analytic, and deliberate (System 2). Whether System 1 or System 2 is used to process information might also depend on the perceived ease or difficulty associated with a cognitive task (Alter et al., 2007). If information processing is perceived as easy, it is more likely that System 1 is activated, leading to an effortless and intuitive processing (cf. fluency research; Alter & Oppenheimer, 2009). If, on the other hand, information processing is perceived as difficult, System 2 will be more likely activated, resulting in more invested mental effort and analytic processing. In a series of

*Correspondence to: Alexander Eitel, Knowledge Media Research Center, Schleichstrasse 6, 72076 Tuebingen, Germany.
E-mail: a.eitel@iwm-kmrc.de

†These authors contributed equally to this work and should both be considered as first authors.

Copyright © 2014 John Wiley & Sons, Ltd.

experiments, Alter and colleagues (2007) have shown that people engaged in more analytic and elaborative reasoning when manipulations were introduced that led to an increase in the perceived difficulty associated with the cognitive task. To increase the perceived difficulty of the task without changing the objective difficulty, information was only *perceptually* harder to process (i.e., disfluent), acting as a cue that guided one's subsequent processing style.

Applied to educational scenarios, engaging in more analytic and elaborative rather than in heuristic and intuitive reasoning is considered to be beneficial to learning (cf. levels of processing; Craik & Lockhart, 1972). Making information perceptually harder to process can be realized by text being presented in *fonts that are slightly more difficult to read* (Alter & Oppenheimer, 2009). Accordingly, Diemand-Yauman, Oppenheimer, and Vaughan (2011) conducted two experiments in which they showed that presenting text in harder-to-read fonts (e.g., **Haettenschweiler**) or in low-quality photocopies, respectively, led to better learning outcomes than presenting text in easy-to-read fonts (e.g., Arial). The authors concluded that making text perceptually harder to process (less legible) functioned as a cue that one may not have mastery over the material, thereby triggering a more effortful and elaborative processing style, and thus leading to better learning. Similar positive effects of disfluency were recently observed by French et al. (2013), where the authors compared recall performance of pupils across a wide range of ability levels in classrooms after presenting them with text in an easy-to-read (Arial) or harder-to-read font (*Monotype Corsiva*).

In other research contexts (e.g., basic memory research), results regarding the impact of disfluency are less straightforward. Whereas Sungkhasettee, Friedman, and Castel (2011) found a positive effect on memory performance from presenting words in a more disfluent way, other studies revealed neutral effects (Song & Schwarz, 2008; Guenther, 2012; Rhodes & Castel, 2008) or partly even a negative effect of disfluency on memory performance (Yue, Castel, & Bjork, 2013). These results can be interpreted in a way that harder-to-read fonts do not necessarily foster learning. On the other hand, one might cautiously argue that disfluency effects might especially unfold with more complex materials such as coherent texts in the educational context (Diemand-Yauman et al., 2011; French et al., 2013), because a more analytic processing mode (triggered by disfluency) would allow understanding more complex materials on a deeper level. Accordingly, especially measures addressing deeper understanding (i.e., transfer) should benefit from disfluency.

However, to our knowledge, research on disfluency in educational contexts has been restricted to text processing and retention. No study has yet been conducted in the context of learning with multimedia, in which the processing of pictures is crucial for achieving a deeper understanding of the instructional material, and in which performance in a transfer test is the measure of major interest (cf. Mayer, 2009). It is unclear which role disfluency plays when learning from text and pictures. According to disfluency theory, one might derive that not only a disfluent text but also a disfluent picture may function as a cue for deeper processing, triggering a more effortful and elaborative processing style, and thus resulting in more invested mental effort and in deeper understanding.

In contrast, even though not explicitly stated by the authors of CLT, one may assume that reducing the perceptual legibility of instructional material (i.e., making it harder to read) would be considered an increase in ECL. Making instructional material less legible would require processing demands to deal with the low legibility of the instructional format, but would not contribute to a deeper processing of the content. Unlike research on desirable difficulties, particularly the generation effect (e.g., deWinstanley & Bjork, 2004; Maki, Foley, Kajer, Thompson, & Willert, 1990), learners receiving a less legible text would not be forced to actively generate new information, which would—in terms of CLT—thus not be associated with an increase in GCL. Because, according to CLT, cognitive resources to process a multimedia instruction are limited but crucial to learning, an increase in ECL (without an increase in GCL) caused by presenting a disfluent text or picture should lead to worse learning outcomes.

To test the contrasting predictions derived from disfluency theory and CLT in multimedia learning, we manipulated the legibility of text and pictures (instructional material) in the present experiments, and investigated the impact on learning outcomes as well as on subjective ratings of mental effort and cognitive load. Beyond investigating the effects of disfluent text (cf. Diemand-Yauman et al., 2011; French et al., 2013), we tested whether disfluent pictures could also act as a cue for deeper processing. With respect to learning outcomes, we did not solely administer a text-based retention test (as done in disfluency research), but also a transfer test. This was carried out to test whether deeper processing triggered by a disfluent instruction would particularly benefit transfer performance. In addition, because the pictures were subject to experimental manipulation of the instructional material (at least in Experiment 1), performance on a pictorial recall test was assessed.

According to disfluency theory, better learning outcomes would be assumed in conditions with disfluent compared with fluent text (and pictures) in Experiments 1 to 4. According to CLT, better learning outcomes would be expected in conditions with fluent compared with disfluent text (and pictures) in Experiments 1 to 4.

EXPERIMENT 1

In this experiment, we manipulated the legibility of the text and pictures that were used as instructional material (Figure 1), and investigated the impact on learning outcomes as well as on subjective ratings of mental effort and cognitive load. Tests of retention, transfer, and pictorial recall were used to measure learning outcomes.

Method

Participants and design

Eighty-four undergraduate students from a university in the south-western part of Germany participated either for course credit or payment in the study. There were 66 female and

Copyright © 2014 John Wiley & Sons, Ltd.

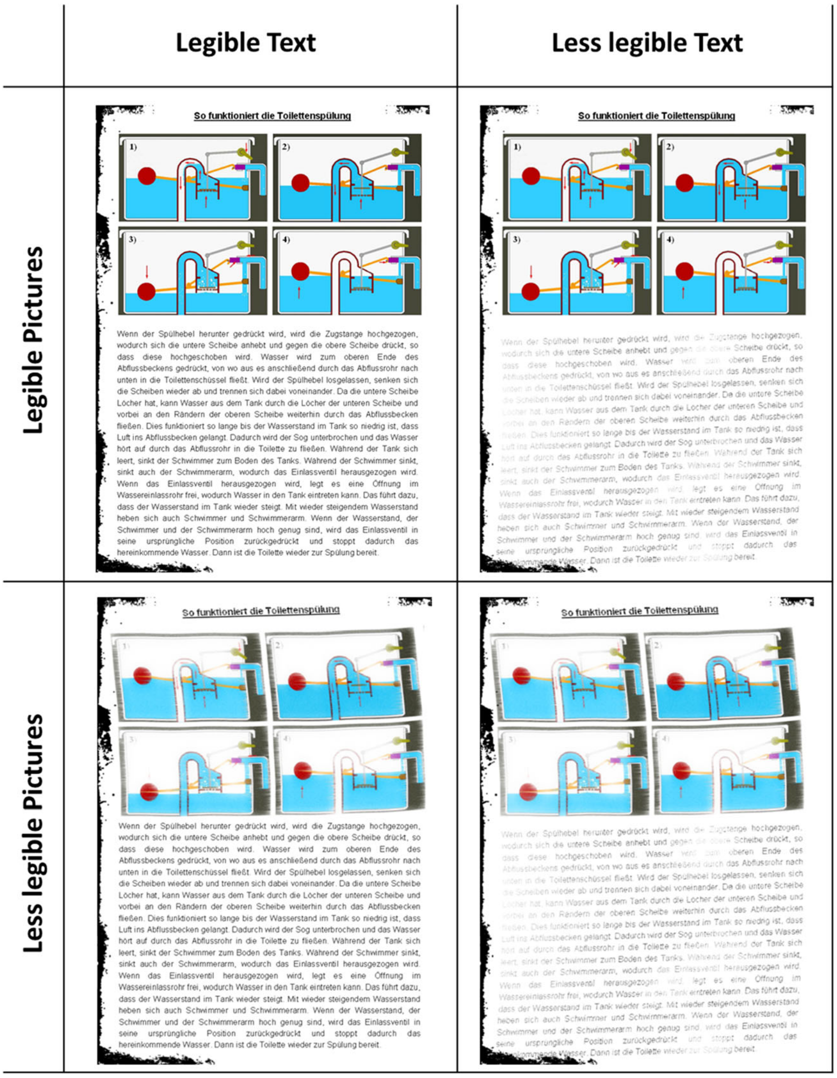

Figure 1. Layout of the second page of the multimedia instruction in the four experimental conditions in Experiment 1. Instructional materials were adapted from Mayer et al. (2005)

18 male participants ($M = 24.23$ years, $SD = 2.09$). They were randomly assigned to one of four conditions, which resulted from a 2×2-design with text legibility (legible vs. less legible) and picture legibility (legible vs. less legible) as independent variables. Twenty-one participants served in each condition.

Materials

The materials comprised a spatial ability test, a demographic questionnaire, and the instructional materials. All materials were printed on sheets of paper. To control for individual differences in spatial abilities—which have been shown to play an important role for this specific type of instructional material (cf. Hegarty & Kriz, 2008)—a shortened version of the Paper Folding Test (PFT; Ekstrom, French, & Harman, 1976) was applied. The PFT consisted of ten items. For each correct answer, one point was given and for each incorrect, answer one point was subtracted, resulting in a minimum of -10 and a maximum of 10 points.

The instructional material was adapted from a study by Mayer, Hegarty, Mayer, and Campbell (2005, Exp. 2). It consisted of two printed pages; the first in landscape format, and the second in high size. The first page depicted an introductory diagram comprising a colored line drawing of a toilet tank, in which each of the ten parts of the toilet tank were labeled (cf. pre-training phase; Mayer, 2009). The second page consisted of four colored key pictures on the top depicting the mechanisms of a toilet flush and an accompanying text that was placed below the pictures describing how a toilet flush works (simultaneous presentation of text and picture; Figure 1).

For the introductory diagram on the first page, text legibility was manipulated by presenting text either in easier-to-read font (Arial, 14 pt, black; legible text), or in harder-to-read font (Haettenschweiler, 14 pt, grayscale 50%; less legible text). A similar manipulation was successfully applied in Diemand-Yauman et al. (2011). Picture legibility was manipulated by presenting either a legible picture or a picture that looked like a low-quality photocopy of the legible picture (less legible; i.e., wavily deformed and blurred).

For the second page, text legibility was manipulated by presenting either a legible text (Arial 14 pt, black) or a less legible text that looked like a low-quality photocopy of the legible text by wavily deforming and blurring the legible text (Figure 1). Different legibility manipulations were used for the first and the second page to rule out the possibility that participants would adapt to the legibility manipulation after

Copyright © 2014 John Wiley & Sons, Ltd.

having seen the first page.[1] It is important to note that care was taken that each letter of a word was still perceivable and had not to be inferred so that the manipulation aimed at disfluency and not at the generation or completion of words—even though it may look otherwise in this reduced view of Figure 1. The experimental manipulation of the picture was identical to page 1. To create the less legible pictures and the less legible text for the second page of the instructional material, the software Adobe® Photoshop® was used.

Measures

Measures comprised subjective ratings of mental effort and cognitive load concerning the learning phase, a knowledge test that consisted of a retention, a transfer, and a pictorial recall test, and items for evaluating the learning materials (i.e., manipulation check items). All measures were assessed in a paper-based format.

Mental effort was assessed by the item 'How much mental effort did you invest?' (cf. Paas, 1992). Besides the widely applied and more general 'mental effort'-item, we additionally assessed subjective cognitive load ratings during learning by one item that is supposed to measure ECL ('How difficult was it for you to learn with the given material'; cf. Cierniak, Scheiter, & Gerjets, 2009). Each item had to be rated on a seven-point Likert scale.

The retention test comprised one question (cf. Mayer et al., 2005): 'Describe how a toilet tank works. Imagine that you push down on the handle of the toilet tank. Describe step-by-step what happens to each of the other parts of the tank as it flushes.' The transfer test comprised four questions (cf. Mayer et al., 2005), each on a separate piece of paper. The four questions were: (1) 'Suppose you push down on the handle of the toilet tank but water does not flush into the toilet bowl. Explain all the possible things that could be wrong.' (2) 'Suppose that after flushing the toilet, you notice that water is continuously running into the toilet tank. Explain all the possible things that could be wrong.' (3) 'Suppose that after you flush the toilet, water continues to run into the toilet bowl without stopping. Explain all of the possible things that could be wrong.' and (4) 'What would happen if the float were to break off from the float arm? What would happen if the upper and lower disks were to stick to each other in the siphon bell? What stops the water from flushing out of the tank?'.

The pictorial recall test comprised three tasks. In the first task, participants had to label the 10 different parts of the toilet tank in a diagram of the toilet flush. In the second and third task, learners were given two pictures of the toilet tank depicting two different steps in the toilet flushing process. For the second task, the connecting rod, the upper, and the lower disc were removed from the pictures, and learners had to draw these elements in the correct way. For the third task, the float, the float arm, the inlet valve, and the incoming water were removed from the pictures and learners had to draw these elements in the correct way.

For the scoring of the retention test, a list of 19 major idea units developed by Hegarty, Kriz, and Cate (2003, Exp. 1) was used. Participants received one point for each major idea unit that they included in their answer to the retention question, regardless of wording. For each transfer question, there was also a list of possible correct answers. One point was given for each correct answer to the transfer questions and the final score of the transfer test was determined by adding up all points given for the transfer questions (cf. Mayer et al., 2005; Exp. 2). Similarly, for each of the three pictorial recall tasks, a list of correct answers and criteria concerning the correctness of the respective drawing was produced beforehand. For each correct aspect, participants received one point and all points were subsumed to an overall score for pictorial recall. Each task of a knowledge test was scored by the same rater who was always blind to the experimental condition.

Two items were administered to check whether the implemented manipulations had been successful, and thus to serve as an indicator for whether text and pictures were less legible (on a superficial level) in the disfluent compared with the fluent conditions. One item asked for the legibility of the text ('I perceived the text layout as well designed'), and one item asked for the legibility of the pictures ('I perceived the picture layout as well designed'). Each of these items had to be rated on a seven-point Likert scale. These items were surveyed after the knowledge test so that they would not influence participants' learning outcomes.

Procedure

Participants were tested in groups of two to six persons per session. Participants were seated at desks in individual cubicles that blocked visual contact with other participants. First, the experimental procedure was briefly described to the participants. Then, participants started to work on the PFT simultaneously. PFT time was limited to 3 min. Subsequently, participants had to fill in the demographic questionnaire.

After filling in the demographic questionnaire, the learning phase began for all participants simultaneously. Depending on the experimental condition, different versions of the introductory diagram sheet were handed out for 1 min and removed afterwards. Then, depending on their experimental condition, participants were given the second page of the instructional material that consisted of a simultaneous presentation of text and pictures describing how a toilet works (Figure 1) and were allowed 5 min to study it. Even though time was restricted, it was sufficient for learners to have the opportunity to engage in valuable cognitive processing and deal with the instructional material in a thorough manner. After the learning phase had finished, students had to rate the mental effort and cognitive load that they experienced during learning once by responding to the respective items. Thereafter, students had to work on tests of retention (6 min), then transfer (11.5 min in total), and then pictorial recall (7 min in total). Learners were given 2.5 min for each

[1] Please note that black margins were added to the second page of the learning materials, which was due to the group testing situation. In case a participant in a condition with disfluent text or pictures would have complained about the bad (superficial) design of the learning materials in the group testing situation, participants in the condition with fluent text and pictures might have been surprised by the complaint, and thus might have started to wonder whether the other participants received other materials. To prevent from such an influence, black margins were included in all conditions so that participants in the condition with fluent text and picture might have concluded that the complaint goes back to the black margins, which, however, did not influence the legibility of the text in the legible conditions.

Copyright © 2014 John Wiley & Sons, Ltd.

of the first three transfer questions, and 4 min for the fourth transfer question. For pictorial recall, learners were given 2 min to work on each of the first two tasks, and 3 min to work on the third task.

Because there were time restrictions to each of the questions for retention, transfer, and pictorial recall, participants started completing the tasks simultaneously. They were instructed to stop writing as soon as the time to work on a task was up (signaled by alarm clock). Participants who finished working on a task early (i.e., before time was up) were not allowed to continue working on the next task, they had to wait until all participants were asked to start with the next task. Participants were not allowed to return to previous pages to review their previous answers. After the pictorial recall test, the items of the manipulation check were given. Finally, students were debriefed and thanked for participation. The whole experiment lasted approximately 1 h.

Results

The analysis followed a three-step procedure. First, an analysis of variance (ANOVA) was conducted to check whether spatial abilities of participants were equally distributed across experimental conditions. Second, to test the contrasting predictions of disfluency theory and CLT against each other, two-factorial analysis of covariances (ANCOVAs) with spatial ability as covariate were conducted to assess how learning outcomes and cognitive load were affected by the experimental manipulation. Third, two-factorial ANCOVAs were performed on the items of the manipulation check to assess whether the intended manipulations were successfully implemented. Because spatial abilities are known to play an important role when learning about mechanical systems (cf. Hegarty & Kriz, 2008), it was included as a covariate in the analyses. Means and standard deviations are reported in Table 1. Partial eta-squared (η_p^2) is reported as a measure of effect size. For eta-squared, effect sizes of 0.01, 0.06, and 0.14 correspond to small, medium, and large effect sizes, respectively (Cohen, 1988).

Spatial ability

A two-factorial ANOVA with text legibility and picture legibility as independent variables and scores in the PFT as the dependent variable were conducted. There was neither a main effect for text layout nor for picture layout nor an interaction (all $F_s < 1$). Thus, spatial abilities were equally distributed across participants from the four conditions.

Learning outcomes

Two-factorial ANCOVAs with text layout and picture layout as independent variables and scores for retention, transfer, and pictorial recall as dependent variables were conducted, with spatial abilities as covariate. Concerning retention, a 2 × 2 ANCOVA revealed neither a main effect of text legibility, $F < 1$, nor of picture legibility, $F(1, 79) = 1.28$, $p = 0.26$, $\eta_p^2 = 0.02$, nor an interaction, $F < 1$. Concerning transfer, in line with disfluency theory, a 2 × 2 ANCOVA revealed a main effect of text legibility, $F(1, 79) = 5.62$, $p = 0.02$, $\eta_p^2 = 0.07$, with learners receiving less legible text outperforming learners receiving legible text. There was no main effect of picture legibility, $F < 1$, and no interaction, $F < 1$. With regard to pictorial recall, in line with CLT, a 2 × 2 ANCOVA showed a main effect of picture layout, $F(1, 79) = 4.01$, $p = 0.049$, $\eta_p^2 = 0.05$, with learners receiving legible pictures outperforming those receiving less legible pictures. There was no main effect of text legibility, $F < 1$, and no interaction, $F(1, 79) = 1.66$, $p = 0.20$, $\eta_p^2 = 0.02$.

Spatial abilities had an impact on retention, $F(1, 79) = 8.68$, $p < 0.01$, $\eta_p^2 = 0.10$, on transfer, $F(1, 79) = 9.88$, $p < 0.01$, $\eta_p^2 = 0.11$, and pictorial recall, $F(1, 79) = 20.50$, $p < 0.001$, $\eta_p^2 = 0.21$. The higher students' spatial abilities, the better they performed on each knowledge test.

Cognitive load

In line with disfluency theory, a 2 × 2 ANCOVA for mental effort revealed a main effect for text legibility, $F(1, 79) = 4.06$, $p = 0.047$, $\eta_p^2 = 0.05$, with learners receiving less legible text reporting that they invested more mental effort than learners receiving legible text. There was no main effect for picture legibility, $F < 1$, and no interaction, $F < 1$. In line with CLT, a 2 × 2 ANCOVA for perceived difficulty (ECL) revealed a main effect for picture legibility, $F(1, 79) = 4.16$, $p = 0.045$, $\eta_p^2 = 0.05$, with learners receiving less legible pictures reporting more difficulties than learners receiving legible pictures. There was no main effect for text legibility, $F < 1$, and no interaction, $F < 1$. Spatial abilities had neither an impact on mental effort,

Table 1. Means (and SD/SE) as a function of text legibility and picture legibility in Experiment 1

Text layout	Legible text		Less legible text	
Picture layout	Legible picture ($n = 21$)	Less legible picture ($n = 21$)	Legible picture ($n = 21$)	Less legible picture ($n = 21$)
Spatial abilities	4.10 (3.70)	5.05 (3.49)	4.62 (4.57)	4.29 (4.20)
Learning outcomes[a]				
Retention	11.35 (0.68)	11.10 (0.68)	11.50 (0.68)	10.20 (0.68)
Transfer	10.19 (0.71)	9.99 (0.71)	11.59 (0.71)	11.97 (0.71)
Pictorial recall	19.80 (0.76)	17.29 (0.77)	18.62 (0.76)	18.08 (0.76)
Cognitive load[a]				
Mental effort	5.22 (0.25)	5.07 (0.25)	5.72 (0.25)	5.56 (0.25)
ECL	2.50 (0.31)	2.89 (0.31)	2.38 (0.31)	3.27 (0.31)
Manipulation check[a]				
Text design	4.53 (0.37)	3.10 (0.37)	2.77 (0.37)	3.07 (0.37)
Picture design	5.22 (0.34)	3.73 (0.34)	5.18 (0.34)	4.20 (0.34)

ECL, extraneous cognitive load.
[a]Means and standard errors are corrected for the influence of spatial abilities.

Copyright © 2014 John Wiley & Sons, Ltd.

$F(1, 79) = 1.68$, $p = 0.20$, $\eta_p^2 = 0.02$, nor on perceived difficulty, $F(1, 79) = 2.34$, $p = 0.13$, $\eta_p^2 = 0.03$.

Manipulation check

Two-factorial ANCOVAs with the two manipulation check items as dependent variables and spatial abilities as covariate were conducted. Concerning the item referring to the legibility of text, the ANCOVA revealed a main effect of text legibility, $F(1, 79) = 5.93$, $p = 0.02$, $\eta_p^2 = 0.07$, with learners receiving legible text perceiving the layout as being better designed than learners receiving less legible text. There was no main effect of picture legibility, $F(1, 79) = 2.35$, $p = 0.13$, $\eta_p^2 = 0.03$. Moreover, a significant interaction was observed, $F(1, 79) = 5.47$, $p = 0.02$, $\eta_p^2 = 0.07$. For the legible pictures conditions, learners receiving legible text perceived the layout of the text as being better designed than learners receiving less legible text ($p < 0.01$). For the less legible picture conditions, there were no differences between learners receiving legible or less legible text ($p > 0.20$).

For the item asking for the legibility of the picture, a 2×2 ANCOVA revealed a main effect of picture legibility, $F(1, 79) = 13.49$, $p < 0.001$, $\eta_p^2 = 0.15$, with learners receiving legible pictures finding the layout better designed than learners receiving less legible pictures. There was no main effect of text legibility, $F < 1$, and no interaction, $F < 1$. Spatial abilities had no impact on the item referring to picture layout, $F(1, 79) = 2.01$, $p = 0.16$, $\eta_p^2 = 0.03$, but a significant effect on the item for text layout, $F(1, 79) = 5.06$, $p = 0.03$, $\eta_p^2 = 0.06$. The higher students' spatial abilities, the less well designed they perceived the text layout.

Discussion

The manipulation check item revealed that less legible text was perceived as being less well designed. In line with disfluency theory (cf. Alter et al., 2007), results of this experiment showed that a less legible text layout led to more mental effort and to a better understanding (as measured by the transfer test). However, learners did not find it more difficult to learn with the less legible text, even though one might derive from disfluency theory that an increase in perceived difficulty should trigger deeper processing and hence foster understanding. This issue will be subject to the general discussion. The manipulation check item for less legible pictures showed that less legible pictures were perceived as being less well designed. Designing pictures in a disfluent way, however, did not lead to better understanding (as measured by the retention and transfer test). By contrast, and in line with CLT, less legible pictures led to a higher perceived difficulty of the instructional material as well as to lower performance in the pictorial recall test than legible pictures. In conclusion, making pictures disfluent in this experiment might have not functioned as a cue for deeper processing, which in turn should have led to better learning according to disfluency theory. Thus, in Experiment 2, we refrained from manipulating the fluency of pictures. Instead, we concentrated on investigating beneficial effects from making text less legible. In particular, we tried to distinguish whether beneficial effects in Experiment 1 were mainly because of a deeper processing of text or additionally because of a more frequent and more intense processing of the pictures (triggered by less legible text), which could be assumed because a more thorough processing of pictures has been shown to foster learning outcomes (e.g., Eitel, Scheiter, & Schüler, 2013).

Therefore, in Experiment 2, besides manipulating the legibility of text, we additionally varied whether pictures were added to text or not. Another reason for why conditions without pictures were included was the fact that they resemble more closely to the experimental conditions reported in prior studies on the disfluency effect (Diemand-Yauman et al., 2011; French et al., 2013).

EXPERIMENT 2

As in Experiment 1, we manipulated the legibility of the text, and investigated the impact on learning outcomes as well as on subjective ratings of mental effort and cognitive load. In addition, we investigated how effects from manipulating the text legibility will be affected when no pictures are presented in multimedia instructional material (cf. Diemand-Yauman et al., 2011; French et al., 2013).

Method

Participants and design

Eighty-five undergraduate students from a university in the south-western part of Germany participated either for course credit or payment in the study. There were 70 female and 15 male participants ($M = 22.32$ years, $SD = 2.67$). They were randomly assigned to one of four conditions, which resulted from a 2×2 design with text legibility (legible vs. less legible) and multimedia (text and picture vs. text only) as independent variables. Three conditions comprised 21 participants, whereas in one condition (less legible text; text only), there were 22 participants.

Materials

The materials were identical to Experiment 1 with the exception of the introductory diagram (pre-training phase) as well as partially the learning materials, which will be explicated in the following.

Because there were text-only conditions, a text was created for the pre-training phase that described the spatial relations of the elements of the introductory diagram (i.e., introductory text). Participants in the text-only conditions saw just the introductory text to familiarize themselves with the components of the to-be-learned system (toilet flush) without having access to a diagram. For the text and picture conditions, as in Experiment 1, learners received the introductory diagram. However, unlike in Experiment 1, the introductory text was additionally given below the labeled diagram. The experimental manipulation of the legibility of the labels in the introductory diagram was identical to Experiment 1. The legibility of the additional text of page 1 was manipulated by presenting text (Arial, 11 pt, black) either in an unaltered fashion (i.e., legible text) or wavily deformed and blurred (i.e., less legible text). The font size of Arial 11 pt was used so that text and introductory diagram fitted on one page.

Copyright © 2014 John Wiley & Sons, Ltd.

As in Experiment 1, in conditions with text and pictures, the second page consisted of four colored key pictures on the top depicting the mechanisms of a toilet flush and an accompanying text that was placed below the pictures describing how a toilet flush works. The four key pictures were presented in a legible way (as in legible picture conditions of Experiment 1; see Figure 1, upper row). In the text-only conditions, the colored pictures were removed. The position of the text on the page was identical to the conditions with text and pictures—only the headline was moved down. The manipulation of the text legibility was the same as in Experiment 1.

Measures

The measures were identical to those used in Experiment 1 with the exception that there was only one manipulation check item asking for the design of the text layout (manipulation check item for picture layout was removed).

Procedure

The procedure was identical to that of Experiment 1 with the exception that participants were given 3 min (instead of 1 min) to study the introductory diagram and the introductory text in the pre-training phase, because more information (by including text) was given compared with the pre-training phase in Experiment 1. The whole experiment lasted approximately 1 h.

Results

The analysis followed the same procedure as in Experiment 1. Means and standard deviations are reported in Table 2.

Spatial ability

To control for whether spatial abilities were equally distributed across participants from the four conditions, a two-factorial ANOVA with text legibility and multimedia as independent variables and scores in the PFT as the dependent variable were conducted. There was neither a main effect for text layout nor for picture layout nor an interaction (all $F_s < 1$).

Learning outcomes

Two-factorial ANCOVAs with text layout (legible vs. less legible) and multimedia (text and picture vs. text only) as independent variables and scores for retention, transfer, and pictorial recall as dependent variables were conducted with spatial abilities as covariate. Concerning retention, a 2 × 2-ANCOVA revealed neither a main effect of text legibility, $F < 1$, nor of multimedia, $F(1, 80) = 1.03$, $p = 0.31$, $\eta_p^2 = 0.01$, nor an interaction, $F < 1$. With regard to transfer, a 2 × 2ANCOVA revealed neither a main effect of text legibility, $F < 1$, nor of multimedia, $F(1, 80) = 2.08$, $p = 0.15$, $\eta_p^2 = 0.03$, nor an interaction, $F(1, 80) = 1.34$, $p = 0.25$, $\eta_p^2 = 0.02$. Concerning pictorial recall, a 2 × 2 ANCOVA showed a significant main effect of multimedia, $F(1, 80) = 18.75$, $p < 0.001$, $\eta_p^2 = 0.19$, with learners receiving text and pictures outperforming learners receiving text only. There was no main effect of text legibility, $F < 1$, and no interaction, $F < 1$.

Spatial abilities had an impact on retention, $F(1, 80) = 8.69$, $p < 0.01$, $\eta_p^2 = 0.10$, on transfer, $F(1, 80) = 7.96$, $p < 0.01$, $\eta_p^2 = 0.09$, and on scores in the pictorial recall test, $F(1, 80) = 14.04$, $p < 0.001$, $\eta_p^2 = 0.15$. The higher students' spatial abilities, the better they performed on each knowledge test.

Cognitive load

Concerning the mental effort item, a 2 × 2 ANCOVA revealed no main effects for text legibility and multimedia, and no interaction (all $F_s < 1$). With respect to the difficulty-item (ECL), a 2 × 2-ANCOVA revealed a main effect for multimedia, $F(1, 80) = 18.08$, $p < 0.001$, $\eta_p^2 = 0.18$, with learners receiving just text reporting more difficulties than learners receiving text and pictures. There was no main effect for text legibility, $F < 1$, and no interaction, $F(1, 80) = 1.44$, $p = 0.23$, $\eta_p^2 = 0.02$. Spatial abilities had no impact on mental effort, $F(1, 80) = 1.19$, $p = 0.28$, $\eta_p^2 = 0.02$ as well as no impact on perceived difficulty, $F < 1$.

Manipulation check

Two-factorial ANCOVAs with the manipulation check item as dependent variable and spatial abilities as covariate were conducted. Results revealed a significant main effect of multimedia, $F(1, 80) = 60.11$, $p < 0.001$, $\eta_p^2 = 0.43$, with participants in the conditions with text and pictures judging the text to better designed than participants in the conditions with text only. Moreover, there was a significant main effect

Table 2. Means (and SD/SE) as a function of text legibility and multimedia in Experiment 2

Text layout	Legible text		Less legible text	
Multimedia	Text and pictures ($n = 21$)	Text only ($n = 21$)	Text and pictures ($n = 21$)	Text only ($n = 22$)
Spatial abilities	5.24 (3.94)	5.05 (3.04)	5.48 (3.06)	6.05 (3.65)
Learning outcomes[a]				
Retention	9.83 (0.70)	9.46 (0.70)	9.95 (0.70)	8.91 (0.68)
Transfer	10.44 (0.66)	8.73 (0.66)	9.61 (0.66)	9.43 (0.65)
Pictorial recall	18.89 (0.82)	15.47 (0.83)	18.75 (0.82)	15.07 (0.81)
Cognitive load[a]				
Mental effort	5.58 (0.28)	5.40 (0.28)	5.38 (0.28)	5.43 (0.28)
ECL	3.48 (0.32)	4.44 (0.32)	3.10 (0.32)	4.80 (0.31)
Manipulation check[a]				
Text design	4.81 (0.32)	2.67 (0.32)	4.38 (0.32)	1.59 (0.31)

ECL, extraneous cognitive load.
[a]Means and standard errors are corrected for the influence of spatial abilities.

Copyright © 2014 John Wiley & Sons, Ltd.

of text legibility, $F(1, 80) = 5.62$, $p = 0.02$, $\eta_p^2 = 0.07$, with learners in the less legible condition perceiving the text as less well designed. There was no significant interaction, $F(1, 80) = 1.05$, $p = 0.31$, $\eta_p^2 = 0.01$. Spatial abilities had no impact on the manipulation check item, $F < 1$.

Discussion

Eliminating pictures from the instructional material led to a higher amount of perceived difficulty and to worse performance on the pictorial recall test, indicating that the text alone was not sufficient for constructing an adequate pictorial model of a toilet flush. Somewhat surprising, this had no significant impact on retention and transfer performance, which might have been due to the harmful redundancy of text and pictures in the pre-training phase (e.g., Schmidt-Weigand & Scheiter, 2011). More interesting concerning the scope of the present paper is that regardless of whether pictures were presented in addition or not, making text disfluent had no impact on the perceived difficulty, and did neither lead to more invested mental effort nor to better learning outcomes (retention, transfer, or pictorial tests) compared with a fluent text version. Thus, the effects from Experiment 1 could not be replicated, even though partially the same learning material—specifically legible pictures and text—was used in these two experiments; just not the same pre-training phase. Hence, this inconsistent finding might have been due to the longer pre-training phase in Experiment 2 (3 min) than in Experiment 1 (1 min). Because of the highly redundant information between introductory text and introductory diagram in Experiment 2, especially participants receiving less legible text might have realized that the effort in reading the disfluent font might not be worthwhile (i.e., inspecting introductory diagram is sufficient). To examine this possible explanation of inconsistent results between Experiments 1 and 2, Experiment 3 was conducted.

EXPERIMENT 3

In this experiment, we investigated whether the discrepant findings between Experiments 1 and 2 could be traced back to the differences in the pre-training phases. As in Experiments 1 and 2, we manipulated the legibility of the text that was accompanied by legible pictures and investigated the impact on learning outcomes as well as on subjective ratings of mental effort and cognitive load.

Method

Participants and design

Ninety-three undergraduate students from a university in the south-western part of Germany participated either for course credit or payment in Experiment 3. Because of problems in the procedure, the incomplete data of one participant had to be removed. Of the remaining 92 participants, 62 were female and 30 were male ($M = 24.45$ years, $SD = 3.85$). They were randomly assigned to one of four conditions, which resulted from a 2×2 design with text legibility (legible vs. less legible) and pre-training phase (short vs. long) as independent variables.

Materials

The materials were identical to parts of Experiment 1 (i.e., conditions with legible pictures) and parts of Experiment 2 (i.e., conditions with pictures). For the pre-training phase, the introductory diagram contained labels and was either shown without text, as in Experiment 1, or with the introductory text as in Experiment 2. The labels of the introductory diagram both in the short and in the long pre-training phase were subject to the same legibility manipulation as in Experiment 1 and 2, whereas the text below the introductory diagram in the long pre-training phase was subject to the same legibility manipulation as in Experiment 2. The experimental manipulation of the text layout for the second page was identical to the two conditions with legible pictures in Experiment 1 as well as to the two conditions with pictures in Experiment 2 (cf. Figure 1, upper row).

Measures

The measures were identical to those used in Experiment 2.

Procedure

The procedure was identical to that of Experiment 1 and 2, with the exception that depending on the condition participants were given 1 min (short pre-training) versus 3 min (long pre-training) to study the introductory diagram sheet before it was removed by the experimenter. Therefore, only participants in conditions with either short (1 min) or long (3 min) study time for the introductory diagram were tested in the same group (of two to six persons per session). Thereafter, the second page was handed out for 5 min irrespective of the conditions participants were assigned to, followed by the cognitive load and mental effort items, the knowledge tests, and the manipulation check item. The whole experiment lasted approximately 1 h.

Results

The analysis followed the same procedure as Experiment 2. Means and standard deviations are reported in Table 3.

Spatial ability

To control for whether spatial abilities were equally distributed across participants from the four conditions, a two-factorial ANOVA with text legibility and pre-training phase as independent variables and scores in the PFT as the dependent variable were conducted. There were no statistical differences: neither for text layout, $F < 1$, nor for the pre-training phase, $F(1, 88) = 1.17$, $p = 0.28$, $\eta_p^2 = 0.01$, nor for an interaction, $F < 1$.

Learning outcomes

Two-factorial ANCOVAs with text layout and pre-training phase as independent variables and scores for retention, transfer, and pictorial recall as dependent variables were conducted, with spatial abilities as covariate. Concerning retention, a 2×2 ANCOVA revealed neither a main effect of text legibility, nor of pre-training phase, nor an interaction (all $Fs < 1$). Concerning transfer, a 2×2 ANCOVA revealed no significant main effects of text legibility, $F(1, 87) = 2.21$, $p = 0.14$, $\eta_p^2 = 0.03$, or pre-training phase, $F < 1$, and no interaction

Table 3. Means (and SD/SE) as a function of text legibility and pre-training phase in Experiment 3

Text layout	Legible text		Less legible text	
Pre-training phase	Short ($n=24$)	Long ($n=21$)	Short ($n=25$)	Long ($n=22$)
Spatial abilities	3.79 (4.20)	5.14 (2.94)	3.44 (3.81)	3.86 (4.54)
Learning outcomes[a]				
Retention	10.03 (0.68)	10.15 (0.74)	9.61 (0.67)	9.96 (0.71)
Transfer	10.02 (0.58)	10.71 (0.63)	9.51 (0.57)	9.45 (0.60)
Pictorial recall	17.74 (0.72)	19.63 (0.78)	18.02 (0.71)	17.60 (0.75)
Cognitive load[a]				
Mental effort	5.08 (0.26)	5.75 (0.29)	5.30 (0.26)	5.45 (0.28)
ECL	2.69 (0.26)	2.69 (0.28)	2.80 (0.26)	3.13 (0.27)
Manipulation check[a]				
Text design	4.07 (0.35)	4.51 (0.38)	3.92 (0.34)	2.94 (0.36)

ECL, extraneous cognitive load.
[a]Means and standard errors are corrected for the influence of spatial abilities.

$F < 1$. With regard to pictorial recall, a 2×2 ANCOVA showed no main effect of text legibility, $F(1, 87) = 1.39$, $p = 0.24$, $\eta_p^2 = 0.02$, and no main effect of pre-training phase, $F < 1$, as well as no significant interaction, $F(1, 87) = 2.46$, $p = 0.12$, $\eta_p^2 = 0.03$. Spatial abilities had an impact on retention, $F(1, 87) = 10.58$, $p < 0.01$, $\eta_p^2 = 0.11$, on transfer, $F(1, 87) = 11.34$, $p = 0.001$, $\eta_p^2 = 0.12$ and on pictorial recall, $F(1, 87) = 25.43$, $p < 0.001$, $\eta_p^2 = 0.23$. The higher students' spatial abilities, the better they performed on each knowledge test.

Cognitive load

Concerning the mental effort item, a 2×2 ANCOVA revealed neither a significant effect of pre-training phase, $F(1, 87) = 2.26$, $p = 0.14$, $\eta_p^2 = 0.03$, nor of text legibility, $F < 1$, nor an interaction $F < 1$. Similarly, there were no significant effects observable for the difficulty-item: Neither for text legibility, $F(1, 87) = 1.02$, $p = 0.31$, $\eta_p^2 = 0.01$, nor for pre-training phase, $F < 1$, nor for an interaction, $F < 1$. Spatial abilities had no impact on mental effort, $F < 1$, but a marginal significant effect on perceived difficulty, $F(1, 87) = 3.50$, $p = 0.07$, $\eta_p^2 = 0.04$. The higher students' spatial abilities, the less difficult they perceived the instruction material.

Manipulation check

Two-factorial ANCOVAs with the manipulation check item as dependent variable and spatial abilities as covariate were conducted. The results revealed no significant effect for pre-training phase, $F < 1$, but a significant effect of text legibility, $F(1, 87) = 5.69$, $p = 0.02$, $\eta_p^2 = 0.06$, with learners in the less legible condition perceiving the text as less well designed. Moreover, there was a significant interaction between text legibility and pre-training phase, $F(1, 87) = 3.93$, $p = 0.05$, $\eta_p^2 = 0.04$. Pairwise comparisons revealed that although for the short pre-training phase text legibility had no impact on the item that referred to the design of the text layout ($p = 0.76$), text legibility had a significant impact on the item that referred to the design of the text layout for the long pre-training phase ($p < 0.01$), with learners in the less legible text version perceiving the text layout as worse designed than their counterparts in the legible text version. Spatial abilities had no impact on the manipulation check item, $F(1, 87) = 2.17$, $p = 0.14$, $\eta_p^2 = 0.02$.

Discussion

The manipulation check item again revealed that the text layout was perceived as less well designed in the less legible text conditions. However, once again, this had no impact on the perceived difficulty to learn with the material. Furthermore, other than in Experiment 1, but in line with Experiment 2, a less legible text (i.e., disfluent) did not trigger participants to invest more mental effort, and connected with this, did not lead to a better understanding. Thus, the disfluency effect (for transfer) as found in Experiment 1 could not be replicated in Experiments 2 and 3, which used the same instructional materials and mostly the same manipulation. Nonetheless, even though we used established multimedia learning materials (i.e. how a toilet flush works; cf. Mayer et al., 2005), one may construe reasons for why the toilet flush materials might not be best suited to examine the impact of disfluency. Therefore, in Experiment 4 of the present research, we switched the materials. In Experiment 4, we used another material that has often been used in studies conducted in the context of multimedia learning; that is, 'how lightning develops' (cf. Mayer, 2009).

EXPERIMENT 4

As in previous experiments, we manipulated the legibility of the text, and investigated the impact on learning outcomes as well as on subjective ratings of mental effort and cognitive load.

Method

Participants and design

Forty-seven undergraduate students from a university in the south-western part of Germany participated for payment in the study. There were 36 female and 11 male participants ($M = 22.15$ years, $SD = 4.30$). They were randomly assigned to the condition with legible text (i.e., fluent) or less legible text (i.e., disfluent). Twenty-two participants served in the disfluent text condition and 25 participants served in the fluent text condition.

Materials

The materials comprised a demographic questionnaire, a test for assessing visual working memory (VWM) capacity and

Copyright © 2014 John Wiley & Sons, Ltd.

the instructional materials. VWM capacity was assessed, because it may influence effects of manipulating superficial text features (i.e., disfluency manipulations). To assess VWM capacity, an adapted computerized form of the Visual Patterns Test (VPT) was applied (Della Sala, Gray, Baddeley, & Wilson, 1997). At this, 30 items consisting of grids of black and white squares were presented in increasing complexity for 3 s on a computer. Participants received a response sheet that contained empty grids for each presented item. After having seen a grid with black and white squares on the computer, participants had to mark on their response sheet where the respective black squares had been. Thereafter, the next item was presented. The total score of the VPT was calculated by summing up the number of correctly solved items, resulting in a minimum of 0 and a maximum of 30 points.

The instructional material was about how lightning develops and was adapted from Mayer (2009). It consisted of eight key pictures about the formation of lightning that were distributed on two printed pages in high size (four key pictures on each page). The corresponding text to each of the eight key pictures was presented below each picture (see Figure 2). As in the study by Diemand-Yauman et al. (2011), a harder-to-read font (*Haettenschweiler, 10pt, italic, grayscale 35%*) was used to make text less legible (i.e., disfluent) compared with a legible text, which was printed in an easier-to-read font (Arial, 10pt, black). Care was taken that the line breaks on the printed pages were the same for the two text versions. Wording of the text and the accompanying pictures were identical for both conditions.

Measures

Measures comprised subjective ratings of mental effort and cognitive load concerning the learning phase, a knowledge test consisting of a retention, a transfer, and a pictorial recall test, and items for evaluating the learning materials (i.e., manipulation check items). All measures were assessed in a paper-based format. The mental effort and perceived difficulty items as well as the manipulation check item concerning the design of the text layout were the same as in Experiments 1 to 3.

The retention test comprised one question (cf. Mayer, 2009): 'Write down everything you know about the formation of lightning. Please be as accurate as possible and use the knowledge you acquired during the learning phase'. Participants received one point for each of 20 major idea units that they included in their answer to the retention question, regardless of wording.

The transfer test comprised four questions, each on a separate piece of paper (cf. Mayer, 2009). The four questions were the following: (1) 'What could be done to decrease the intensity of a lightning storm?' (2) 'Suppose you see clouds in the sky, but no lightning. Why not?' (3) 'What does air temperature have to do with lightning?' and (4) 'What causes lightning?' One point was given for each correct answer to the transfer questions and the final score of the transfer test was determined by adding up all points given for the transfer questions.

The pictorial recall test consisted of five test items (cf. Schmidt-Weigand & Scheiter, 2011). For each item, learners had to draw the most relevant aspects in the formation of lightning. Answers had to be given on five sheets containing a simplified background scene. One point was given for each element (e.g., positively charged particle) that was appropriately drawn.

Procedure

The procedure was identical to that of Experiments 1 to 3, with the exceptions that participants did not receive the

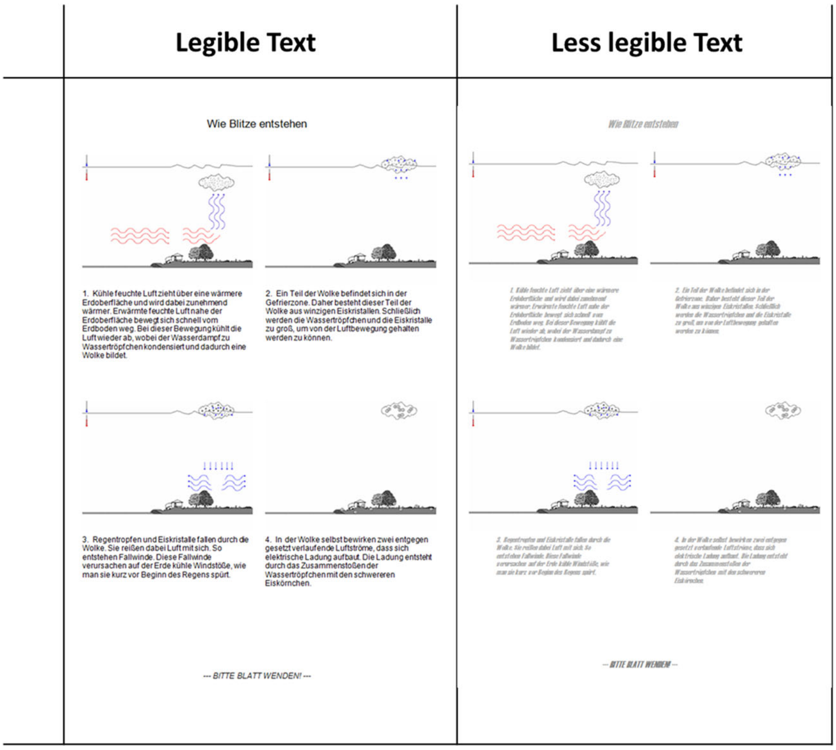

Figure 2. Layout of the multimedia instruction (first of two pages) in the two experimental conditions in Experiment 4. Instructional materials were adapted from Mayer (2009)

PFT, that participants had only 3 min for the learning phase, 3 min for each of the transfer tasks and 8 min for the pictorial test, and that they had to complete the VPT after working on the knowledge tests.

Results

The means and standard deviations for the two groups are reported in Table 4. Separate *t*-tests for independent means were conducted for all variables listed in Table 4. Cohen's *d* is reported as a measure of effect size, with $d = 0.2$ indicating a small effect, $d = 0.5$ a medium effect, and $d = 0.8$ a large effect, respectively (Cohen, 1988). First, it was examined whether participants differed between the two conditions with respect to VWM capacity. The *t*-tests for independent means showed no differences between students in the fluent and disfluent conditions, $t(41.12) = -1.13$, $p = 0.28$, $d = -0.28$.

Learning outcomes

Separate *t*-tests for independent means for the scores in the retention, transfer, and pictorial recall test were conducted between the fluent and disfluent condition. No statistical significant differences were observable: Neither for retention, $t(45) = 0.01$, $p = 0.99$, $d = 0.003$, nor for transfer, $t(45) = 0.14$, $p = 0.89$, $d = 0.04$, nor for the pictorial recall test, $t(45) = 0.732$, $p = 0.47$, $d = 0.22$, indicating that the manipulation of text legibility had no impact on learning outcomes.

Cognitive load

Separate *t*-tests for independent means for the mental effort item and the difficulty item were conducted between the fluent and disfluent condition. No statistical significant differences were observable: neither for mental effort, $t(45) = 0.57$, $p = 0.57$, $d = 0.17$, nor for difficulty, $t(45) = -1.17$, $p = 0.25$, $d = -0.35$, indicating that the manipulation of text legibility had no impact on subjective experienced cognitive load.

Manipulation check

To check whether the text layout in the less legible condition was perceived as being less well designed compared with the legible text condition, a *t*-test for independent means was conducted for the manipulation check item. Results revealed no statistical significant differences between the fluent and disfluent condition, $t(45) = 1.09$, $p = 0.28$, $d = 0.33$.

Table 4. Means (and SD) as a function of text legibility in Experiment 4

	Legible text (fluent) ($n = 25$)	Less legible text (disfluent) ($n = 22$)
Visual patterns test	22.20 (4.45)	23.41 (2.82)
Learning outcomes		
Retention	8.24 (3.68)	8.23 (3.77)
Transfer	2.76 (1.85)	2.68 (1.89)
Pictorial recall	7.44 (3.03)	6.77 (3.22)
Cognitive load		
Mental effort	5.56 (0.87)	5.41 (0.96)
ECL	3.12 (1.39)	3.59 (1.37)
Manipulation check	4.32 (1.44)	3.86 (1.42)

ECL, extraneous cognitive load.

Discussion

Both experimental groups (in fluent and disfluent condition) scored similar on each of the learning outcome measures in Experiment 4 (i.e., retention, transfer, pictorial recall), and reported the same amount of mental effort and perceived difficulty. Hence, as in Experiments 2 and 3, there was no disfluency effect, which weakens the argument that the lack of an effect can be traced back to the unique characteristics of the toilet flush materials. One may counter that for Experiment 4, the manipulation check item did not reveal differences between the two conditions. However, the disfluency manipulation (i.e., Arial vs. **Haettenschweiler**) was very similar to the reported manipulations of the study by Diemand-Yauman et al. (2011), who in turn reported differences in learning outcomes. Moreover, by using this relatively subtle manipulation, the objection of using a too strong manipulation in Experiments 1–3 can be repelled.

GENERAL DISCUSSION

In the present experiments, we investigated the role of disfluency in multimedia learning with respect to learning outcomes and cognitive load. According to disfluency theory (Alter & Oppenheimer, 2009), a less legible text should have led to a deeper processing of the instructional material, resulting in better learning outcomes (especially regarding transfer performance). Against the backdrop of CLT (Sweller et al., 2011), learners would have needed cognitive resources to deal with the low legibility format of instruction (i.e., text and pictures) that was extraneous to the instructional contents. This would mean an increase in ECL, being detrimental to the learning success.

In line with *disfluency theory*, students learning with a disfluent text invested more mental effort and had better transfer scores than students learning with a fluent text in Experiment 1. These results suggest that making text harder-to-read on a perceptual level might have acted as a cue for participants to process materials more deeply, in turn leading to better comprehension of the instructional contents. In Experiments 2, 3, and 4, however, making text harder-to-read on a perceptual level did not have beneficial effects on self-reported mental effort, retention, and transfer performance. In contrast, descriptively, results from Experiments 2, 3, and 4 went in the opposite direction (Figure 3). Making text disfluent tended to hamper learning outcomes (retention, transfer, and pictorial recall) compared with a legible text. This was true even though in the first three experiments we did basically the same; namely, the same instructional materials were presented with the same disfluency manipulations to participants that were taken from the same pool of subjects (i.e., undergraduate students).

Therefore, a Bayesian analysis (cf. Masson, 2011) was conducted for conditions with legible versus less legible text across all four experiments to test whether there was an overall effect of disfluent text on learning outcomes. This analysis revealed that the conditional probabilities for the null hypothesis being true given the present data were 0.93, 0.93, and 0.94 for retention, transfer, and pictorial recall,

Copyright © 2014 John Wiley & Sons, Ltd.

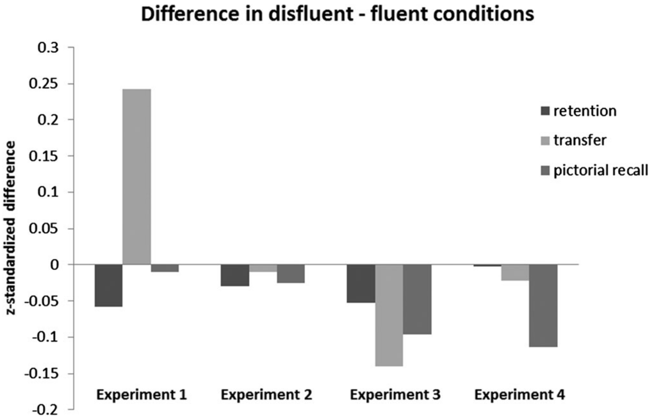

Figure 3. Differences between conditions with less legible (disfluent) versus legible (fluent) text for the three learning outcome measures across all four experiments. For Experiments 1 to 3, learning outcomes corrected for the influence of spatial abilities are reported

respectively (Table 5).[2] Thus, given the present data, making text disfluent is unlikely to having produced reliable differences on learning outcomes in the present experiments (below 10% likelihood for all outcome measures). This shows that introducing perceptual difficulties was not desirable for learning. Moreover, across the four, experiments perceived, difficulty was not correlated with self-reported mental effort ($r = 0.06$, $p = 0.27$) and negatively correlated with all learning outcome measures ($r = -0.20$ to $r = -0.35$, all $p_s < 0.001$). Furthermore, mental effort was not associated with any of the learning outcome measures ($r < 0.01$ to $r = 0.08$, all $p_s > 0.15$) across the four experiments. Hence, neither did a less legible text lead to an increase in perceived difficulty, nor did an increase in perceived difficulty result in more invested mental effort or a better understanding, thereby not supporting assumptions made by disfluency theory in educational settings. Besides, the missing effects of disfluency are unlikely to be traced back to a too weak manipulation, because the items of the manipulation check revealed that disfluent text was perceived as being less well designed (i.e., less legible on a superficial level) than fluent/regular text. Moreover, the non-significant results obtained in Experiment 4 suggest that missing effects of disfluency were not bound to the specific instructional material of the toilet flush and that also a more subtle disfluency manipulation did not lead to better learning outcomes. Nevertheless, it should be noted that the manipulation check items solely asked about the design of the text layout to clarify whether the differences in text layouts (fluent vs. disfluent) were noticeable. Thus, in future studies it might be worthwhile to ask more directly about the (subjective) legibility in conditions with fluent and disfluent text.

Following *CLT*, in Experiment 1, students reported higher ECL when learning with less legible *pictures*, leading to worse pictorial recall. Making *text* less legible, by contrast, had no (significant) detrimental impact on cognitive load and on learning outcomes when considering all four experiments. Thus, the present results suggest that although making pictures disfluent is an undesirable difficulty (possibly making it harder to extract relevant information), making text less legible does not negatively affect cognitive load by an increase in ECL. One may argue that in a more restricted view of CLT, ECL is mainly caused by a bad instructional design that forces learners to hold a representation in working memory while matching it with another external representation (e.g., as in the split-attention effect; cf. Ayres & Sweller, 2005), thereby increasing *cognitive* load. By taking such a more restricted perspective of CLT, it might be subject to discussion whether CLT would actually predict negative effects from introducing *perceptual* difficulties in text by manipulating the text's visual appearance. Making text less legible in the present experiment may have led to a more intense *perceptual* processing of single words and sentences. Thus, learners may have experienced a higher *perceptual load* through a less legible text layout (cf. Lavie, 1995) while their cognitive load remained unaffected.

Potential boundary conditions of the disfluency effect

Given the inconsistent results from making text disfluent in the present experiments, one might argue with statistical artifacts or chance, respectively. On the one hand, the positive disfluency effect for transfer performance in Experiment 1 of the present research (and from other published experiments like from Diemand-Yauman et al., 2011 or French et al., 2013) might have been due to a Type I error (alpha-error). On the other hand, not finding beneficial effects of disfluency in the other three experiments (see also Carpenter, Wilford, Kornell, & Mullaney, 2013; Yue et al., 2013) might have been due to a Type II error (beta-error)—even though particularly the Type II error is quite unlikely, because the descriptive results of Experiments 2, 3, and 4 of the present research were in disadvantage of disfluent text.

More interestingly is the notional explanation that disfluency may be beneficial only for certain *learner characteristics* that we did not assess in the current experiments and that these learner characteristics were more accentuated in the study sample of the first experiment compared with the study sample of the other three experiments. This may explain why disfluency acted as a cue to invest more mental effort and thus to result in better transfer performance only in Experiment 1. Moreover, this may explain why participants in the disfluent text conditions in Experiment 1 had better performance in the transfer test than participants in the disfluent text conditions in Experiments 2 and 3, whereas for the fluent text conditions, participants did not differ between Experiments 1 to 3 with respect to transfer performance. One might interpret these results in a way that participants in Experiment 1 differed from participants in Experiment 2 and 3 with respect to learner characteristics that we did not assess, but which were important regarding the disfluency effect. However, participants in all four experiments were students from the same university, and we assessed typical learner characteristics such as spatial abilities, age, and gender. Participants were comparable regarding these characteristics, which is why this cannot explain the inconsistent data between the present experiments. This may call for further experiments with more emphasis on learner

[2] Similar results were obtained when including only the first three experiments in the analysis. Please contact the corresponding author for more details concerning the Bayesian analysis of the present data.

Copyright © 2014 John Wiley & Sons, Ltd.

Table 5. Mean z-standardized scores, analysis of variance results, and Bayesian analysis (cf. Masson, 2011) for retention, transfer, and pictorial recall aggregated across Experiments 1 to 4 for legible (fluent) versus less legible (disfluent) text

Learning Outcome	$M_{\text{legible text}}$ ($n = 154$)	$M_{\text{less legible text}}$ ($n = 154$)	$F_{1,306}$	Bayes Factor[a]	$p(H_0/D)$[b]	$p(H_1/D)$
Retention	0.04 (0.98)	−0.04 (1.03)	0.40	14.34	0.93	0.07
Transfer	−0.04 (1.00)	0.04 (1.00)	0.39	14.44	0.93	0.07
Pictorial recall	0.03 (1.00)	−0.03 (1.01)	0.21	16.50	0.94	0.06

[a]The Bayes factor provides the odds ratio for the null/alternative hypothesis. A value of 14 indicates that the evidence for the null hypothesis is 14 times stronger than that for the alternative hypothesis.
[b]Values between 0.75 and 0.95 yield positive evidence in favor of accepting the null hypothesis (Masson, 2011).

characteristics (e.g., prior knowledge levels or learning styles) when learning with disfluent material.

Another factor that possibly influences the impact of disfluency on the learning success is the *explicit learning instruction*. In the study of French et al. (2013), pupils did not know that they were tested afterwards, and we do not know what the exact instructions were in the studies of Diemand-Yauman et al. (2011). In our studies, however, learners were told that they should learn the materials in preparation for a test (cf. Szpunar, McDermott, & Roediger, 2007). Thus, learners might have not been endangered to process the materials superficially, even in the case of a fluent font. As a result, the impact of the disfluency manipulation to act as a cue for a deeper processing of the materials might have been weakened.

Irrespective of these theoretical explanations, the results of the current studies do not support the assumption that disfluency is especially beneficial when learning with more complex materials in educational contexts. When taking only results from Experiment 1 into account, one might have drawn the (faulty) conclusion that making text harder-to-read is beneficial to learning with text and pictures. This conclusion is unlikely to be drawn when considering results from all four of the present experiments (cf. Figure 3). Therefore, the present pattern of results illustrates why replications are important to psychological science (see Pashler & Wagenmakers, 2012 for a special issue on replicability), and thus it is important to report these results. It is crucial to inform other researchers that there is no strong evidence in favor of disfluency within the educational context. Rather, the present results do not offer support for using hard-to-read fonts as a device for improving learning from text (and pictures). In this regard, we can report anecdotally that we were in exchange with other researchers who recently conducted disfluency studies, and whose results were rarely in favor of making text disfluent. Thus, publishing the present results might help preventing from (or attenuate) a possible publication bias regarding effects of disfluency on educational outcomes. Otherwise, in the worst case, based on published studies in the field of educational psychology, one might recommend to make instructional material hard-to-read, which would not stand on solid empirical grounds if the unpublished studies were taken in account. In consequence, in our opinion, any recommendation for educational practice is not (yet) justified by empirical data regarding the disfluency effect.

ACKNOWLEDGEMENT

This research was funded by the Pact for Research and Innovation of the Competition Fund of the Leibniz Gemeinschaft. We thank Raphael Frank for the help in data collection.

REFERENCES

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*, 219–235. doi:10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*, 569–576. doi:10.1037/0096-3445.136.4.569

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 134–146). New York, NY: Cambridge University Press.

Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559. doi:10.1126/science.1736359

Bjork, R. A. (2013). Desirable difficulties perspective on learning. In H. Pashler (Ed.), *Encyclopedia of the Mind*. (Vol *4*, pp. 134–146). Thousand Oaks: Sage Publication.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, 1–7. doi:10.3758/s13423-013-0442-z

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*, 315–324. doi:10.1016/j.chb.2008.12.020

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edn). Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & Lockhart, S. E. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. doi:10.1016/S0022-5371(72)80001-X

De Croock, M. B. M., van Merriënboer, J. J. G., & Paas, F. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: effects on transfer performance and invested mental effort. *Computers in Human Behavior*, *14*, 249–267. doi:10.1016/S0747-5632(98)00005-3

Della Sala, S., Gray, C., Baddeley, A. D., & Wilson, L. (1997). *Visual Pattern Test*. Bury St. Edmunds: Thames Valley Test Company.

DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: implications for making a better reader. *Memory and Cognition*, *32*, 945–955. doi:10.3758/BF03196872

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): effects of disfluency on educational outcomes. *Cognition*, *118*, 114–118. doi:10.1016/j.cognition.2010.09.012

Eitel, A., Scheiter, K., & Schüler, A. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology*. doi:10.1002/acp.2922

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for kit of Factor-Referenced Cognitive Tests*. Princeton: Educational Testing Service.

French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, A., … Tabor, J. (2013). Changing fonts in education: How the benefits vary with ability and dyslexia. *The Journal of Educational Research*, *106*, 301–304. doi:10.1080/00220671.2012.736430

Guenther, R. K. (2012). Does the processing fluency of a syllabus affect the forecasted grade and course difficulty? *Psychological Reports*, *110*, 946–954. doi:10.2466/01.11.28.PR0.110.3.946-954

Hegarty, M., & Kriz, S. (2008). Effects of knowledge and spatial ability on learning from animation. In R. K. Lowe & W. Schnotz (Eds.), *Learning*

Copyright © 2014 John Wiley & Sons, Ltd.

with Animation: Research and Design Implications (pp. 3–29). New York: Cambridge University Press.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction*, *21*, 209–249. doi:10.1207/s1532690xci2104_1

James, W. (1950). *The Principles of Psychology*. New York: Dover. (Original work published 1890)

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 451–468. doi:10.1037//0096-1523.21.3.451

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 609.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*, 679–690. doi:10.3758/s13428-010-0049-5

Mayer, R. E. (2009). *Multimedia Learning* (2nd edn). Cambridge: Cambridge University Press.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied*, *11*, 256–265. doi:10.1037/1076-898X.11.4.256

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434.

Paas, F., & van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351–371. doi:10.1007/BF02213420

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science—a crisis of confidence? *Perspectives on Psychological Science*, *7*, 528–530. doi:10.1177/1745691612465253

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625. doi:10.1037/a0013684

Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: generation and interleaving effects. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*. (1850–1855). Mahwah, NJ: Lawrence Erlbaum.

Salomon, G. (1984). Television is "easy" and print is "tough": the differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*, 647–658. doi:10.1037/0022-0663.76.4.647

Schmidt-Weigand, F., & Scheiter, K. (2011). The role of spatial descriptions in learning from multimedia. *Computers in Human Behavior*, *27*, 22–28.

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do. *Psychological Science*, *19*, 986–988. doi:10.1111/j.1467-9280.2008.02189.x

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: illusions of competency and desirable difficulties. *Psychonomic Bulletin and Review*, *18*, 973–978. doi:10.3758/s13423-011-0114-9

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. New York: Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296. doi:10.1023/A:1022193728205

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III (2007). Expectation of a final cumulative test enhances long-term retention. *Memory and Cognition*, *35*, 1007–1013. doi:10.3758/BF03193473

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: the influence of typeface clarity on metacognitive judgments and memory. *Memory and Cognition*, *41*, 229–241. doi:10.3758/s13421-012-0255-8

Copyright © 2014 John Wiley & Sons, Ltd.

Copyright of Applied Cognitive Psychology is the property of John Wiley & Sons, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.