

Exploring the lack of a disfluency effect: evidence from eye movements

Alexander Strukelj¹ · Katharina Scheiter² ·
Marcus Nyström³ · Kenneth Holmqvist³

Received: 2 February 2015 / Accepted: 16 October 2015 / Published online: 4 November 2015
© Springer Science+Business Media New York 2015

Abstract An eye-tracking study with 60 native Swedish speakers (18–30 years) was conducted to investigate the positive effects on learning outcomes predicted by the disfluency effect. Subtle low-pass filtering was used as a disfluency manipulation and compared with a control condition using regular text. The text was presented on four separate text presentation screens (TPSs), and eye movements were recorded. A free recall task was given 25 min later, and working memory capacity (WMC) was assessed to test if it would moderate learning outcomes. The disfluency manipulation had no effect on learning outcomes, total reading times on words or lines, first- or second-pass reading on lines, or average fixation durations. Moreover, the disfluency effect was not moderated by students' WMC or self-reported prior knowledge of the topic. However, an adaptation to the disfluency manipulation was found. Total reading times on both words and lines were shorter in TPS 1 and 2 in the disfluency condition compared with the control condition, whereas reading times were longer in TPS 3 and 4. It is discussed if failures to replicate the disfluency effect arise from material features, with positive adaptations to disfluency (i.e., higher effort investment) possibly requiring more comprehensive materials.

Keywords Desirable difficulty · Disfluency effect · Effort regulation · Eye tracking · Low-pass filtering · Reading comprehension · Working memory capacity

✉ Alexander Strukelj
alexander.strukelj@englund.lu.se

¹ Centre for Languages and Literature and The Humanities Lab, Lund University, Box 201, SE-22100 Lund, Sweden

² Leibniz-Institut für Wissensmedien and Eberhard Karls University of Tübingen, Tübingen, Germany

³ The Humanities Lab, Lund University, Lund, Sweden

Introduction

When reading text that has been visually manipulated to be slightly more difficult to read (e.g., by using hard-to-read fonts), will this in fact *benefit* processing, and consequently understanding, of the text? According to previous research, the answer is “yes” (Diemand-Yauman et al. 2011; French et al. 2013). Diemand-Yauman et al. (2011) showed that learning outcomes improved when making text more difficult to read, not only for simple memory tasks but also in real-life educational contexts. In line with previous research, we refer to findings that show better memory and/or comprehension for visually manipulated materials as the *disfluency effect*. A manipulation to a text that is assumed to yield this effect, i.e., a manipulation that makes encoding at a perceptual level difficult, is referred to as a *disfluency manipulation*. Text that has undergone a disfluency manipulation is referred to as *disfluent text*.

The disfluency effect is explained as a phenomenon resulting from metacognitive and self-regulatory processes. In particular, it is assumed that students are metacognitively aware of their processing behavior, and will monitor it. Consequently, they are assumed to notice the difficulties that occur during encoding of disfluent text. The experience that encoding does not proceed as smoothly as expected will serve as a metacognitive cue that regulation of study behavior is required (Bjork et al. 2013). As a consequence, students will invest more effort into the processing of the text, using processing strategies that are suited to overcome these difficulties, which in turn should yield better recall of the information (Diemand-Yauman et al. 2011). Creating hard-to-read text has been suggested to serve as a desirable difficulty during learning, triggering more effortful processing to compensate for the aversive effects of hard-to-read text, ultimately leading to better learning outcomes.

However, recent research suggests that the disfluency effect found by Diemand-Yauman et al. (2011) is very difficult to replicate, suggesting that it might be less robust than initially hypothesized (for failed replications see Eitel et al. 2014; Meyer et al. 2015; Yue et al. 2013). The aim of the current study was to address two explanations of these replication failures. These explanations along with their implications are discussed in the following sections.

Why the disfluency effect may be difficult to replicate

One reason for recently observed failures to replicate the disfluency effect might be that students’ responses to disfluent materials are often too subtle to manifest themselves at a more coarse-grained outcome level. When reading disfluent text, it is possible that students become metacognitively aware of the difficulties associated with perceptually encoding the text, and regulate their reading behavior to accommodate these difficulties. However, these changes in self-regulation behavior might not be sufficiently strong to affect learning outcomes. Therefore, the current study investigates how disfluent texts affect students’ learning outcomes, as indicated by results on a free recall task, and information-processing behavior, as indicated by eye movement behavior recorded during reading.

Eye movements are tightly linked to visual attention and have been used extensively to measure cognitive processes in a variety of research fields such as concept learning (Rehder et al. 2009; Rehder and Hoffman 2005), oculomotor control in visual search (Hooge and Erkelens 1999), web usability (Ehmke and Wilson 2007), and learning and instruction (Eitel et al. 2013; Johnson and Mayer 2012; Scheiter and Eitel 2015; Schwonke et al. 2009). Eye tracking has been particularly useful in reading research, where the fine movements of the eyes

contain important information about how the text is read and processed (see Clifton et al. 2007; Rayner 1998, for comprehensive reviews).

During reading, the eyes make small, very quick jumps across the text called *saccades*. The saccades are interspersed with *fixations*, where the eyes remain relatively still over a period of time. During a fixation, information intake is open and visual information can be processed by the brain. Reading is a complex multi-level process, composed of low-level processes such as fixations of single words, to higher-level events such as rereading portions of text to verify and possibly repair one's understanding of the text (i.e., regressions). This is reflected in different units of analysis (e.g., words, paragraphs, text, etc.). The intention of disfluent text is to exert an influence on low-level processes by increasing the difficulty of perceptually encoding written language.

The duration of a fixation is often considered a measure of processing time on a word, and can be related to factors such as word frequency (Rayner and Duffy 1986), text difficulty (Reingold and Rayner 2006), and cognitive load (Rayner 2009). A global eye-tracking measure such as the *average fixation duration*, signifying the total duration of all fixations on a word divided by the total number of fixations, approximates the mean amount of processing during each fixation on a word during a trial (Rayner et al. 2009). However, the measure may underestimate the actual time the word is inspected (Rayner 1998). Therefore, the measure *total reading time* is often used in reading research. Total reading time is the total time a word has been read, that is, the sum of all fixations. This measure gives an indication of how long each word was processed during reading, and when used on all words in a text, it gives an approximation on how long participants spent processing the text (Rayner 1998). Total reading times can also be used for entire lines of a text, with a higher value indicating a higher amount of re-reading.

In the current study, eye-tracking measures were used as process measures to investigate if increased difficulty of the visual decoding of text also stimulates more effortful processing of the text contents. A more effortful processing of difficult-to-read text should be reflected in longer total reading times, longer average fixation durations as well as in more re-reading behavior (Hyönä et al. 2003; Mason et al. 2013; Reingold and Rayner 2006).

A second explanation for the aforementioned replication failures might be that the disfluency effect does not reliably occur for all learners. Instead, it is largely moderated by the cognitive prerequisites a learner brings to the task. If the disfluency effect was based on students' additional effort investment resulting from their metacognitive regulation, this would require sufficient cognitive resources to adaptively regulate their own learning as well as visually decoding the hard-to-read text. From this perspective, the disfluency effect should become more pronounced for students with better resource availability. Accordingly, disfluency effect could be moderated by students' availability of cognitive resources, namely, their working memory capacity.

Working memory capacity (WMC) has been shown to predict the outcome of a wide variety of cognitive tasks (e.g., Kane et al. 2001). Therefore, much research has examined the influence of individual differences in WMC on reading comprehension (Daneman and Carpenter 1980; Daneman and Merikle 1996) and learning (Unsworth and Engle 2005). Daneman and Carpenter (1980) found that reading comprehension for high capacity readers was significantly greater than for low capacity readers, which was further substantiated by a meta-analysis of 77 studies examining the effect of WMC on language comprehension (Daneman and Merikle 1996). Furthermore, Unsworth and Engle (2005) showed that *intentional* learning was helped by high WMC, but that automatic processing seemed to be

unaffected. Thus, it is expected that WMC will contribute positively to learning in the reading task used in the current study.

With respect to the disfluency effect, WMC may be important in two ways. First, if the disfluency effect was a result of the monitoring of processing difficulties, with subsequent effort regulation to compensate for these difficulties, students would require sufficient cognitive resources to engage in effective self-regulation (cf. Ilkowska and Engle ; Van Merriënboer and Sluijsmans). Second, the disfluency effect is assumed to arise from students investing more effort, requiring free cognitive resources that can be devoted to the reading task. Against this backdrop, it can be expected that the disfluency effect is more likely to occur for students with high compared with low WMC, who have sufficient cognitive resources to self-regulate their processing and to invest them into more effortful study (cf. Lehmann et al. *in press*).

WMC can be measured in a variety of ways (see Conway et al. 2005, for a review). A distinction is commonly made between simple and complex WMC tasks. A simple task only involves storage and recall of items, where complex tasks include an additional processing task performed before each item to be stored in memory and subsequently be recalled serially. One of the most common complex tasks is the operation span, or OSPAN, task. The OSPAN task is a cognitively demanding task, where participants need to memorize letters (the storage task) while solving math problems (the processing task). It measures the participant's capacity and ability of simultaneously performing two separate tasks. In their meta-analysis, Daneman and Carpenter (1980) show that complex process-and-storage-tasks such as the OSPAN span task are superior predictors of comprehension, compared to simple WMC tasks. Unsworth et al. (2005) created an automated version of the OSPAN task, which was validated to correctly assess WMC.

Overview of experiment and hypotheses

In the current study, students received a written explanatory text that either was printed regularly (control condition) or manipulated so that it became harder to read (disfluency condition). Several different types of disfluency manipulations have been used in previous studies, such as italicizing the font, using a different font, or moving the paper in the copier while making the copies of the study material (Diemand-Yauman et al. 2011). At least the latter manipulation deployed by Diemand-Yauman et al. in their second study is very hard to replicate. Moreover, the effects on reading by adding a visual manipulation on the text has been investigated using eye-tracking in several studies. However, these studies did not aim at studying the disfluency effect. Manipulating the contrast of an isolated word in a text has been shown to trigger longer fixation durations on the difficult-to-read word (Reingold and Rayner 2006). Also, blurred sentences trigger longer fixation times, when compared to non-blurred sentence (Jainta et al. 2011). Crucially, the comprehension when reading blurred sentences was also measured in Jainta et al. (2011), and was shown to not differ significantly between conditions. The manipulation in Jainta et al. (2011) looks very similar to low-pass filtering the text, which degrades the image by evening out frequencies in the image above a set threshold, thereby creating a blurring effect.

Research has shown that low-pass filtering the text slows reading of text, but only when filtering was done above a certain threshold. This suggests that only one spatial-frequency channel is required for effective or uninterrupted reading (Legge and Cheung 2004; Legge et al. 1985). This has been further investigated using the *gaze-contingent moving-window*

paradigm, where stimuli change in real-time depending on where the participant is fixating. A moving window can be defined to either unmask a section of a manipulated stimulus, or in some way manipulate a section of an otherwise unmasked stimulus. This window moves depending on the eye movement behavior of the participant, and its position can be located directly at the center of the current fixation or anywhere else on the stimulus presentation screen. To investigate the effect of different types of visual filtering on reading behavior, a gaze-contingent moving-window was set to low-, band-, and high-pass filter the *foveal* text (i.e., only a small area in the center of the current fixation), providing “coarse”, “medium”, or “fine” visual input, while the rest of the text was left unfiltered (Jordan et al. 2012). In the study, reading times were only found to differ significantly for the coarse (i.e., low-pass filtered) text using a window of five characters or more, with the low-pass filtering producing longer reading times. Chung et al. (2007) found that dioptic blur decreased reading speed as blur intensity increased, but interestingly also found a small reading speed *increase* for the lowest blur levels. Compensating for visually degraded text has also been shown to alter reading processes, such as increasing the reliance on context. This has been shown using lexical decision tasks, producing stronger semantic priming effects when words are degraded (Becker and Killion 1977; Holcomb 1993).

Against the backdrop of this research, low-pass filtering was used to create the disfluent text, as the current study aims to reveal longer reading times from the disfluency manipulation. Furthermore, as visual filtering does not alter the size or position of text, the manipulation suitable was for eye-tracking analysis. Several hypotheses were posed for the current study.

First, according to the disfluency hypothesis, participants reading a disfluent text were expected to have higher scores on the learning outcome questions compared to participants in the fluent condition. Moreover, disfluency effects were also expected at the information-processing level, with more effortful processing evident in eye movements. That is, the text with the disfluency manipulation should take longer to read compared to reading the fluent text, and this decoding of the visual disfluency would result in longer fixations on the words and more re-reading behavior. This would be evidenced in longer total reading times on the words and lines in the text, and possibly longer average fixation durations. Altogether, these changes in processing behavior reflect an alteration of students’ self-regulation. While monitoring their reading, they become aware of difficulties resulting from the disfluent text and adjust their reading to overcome these difficulties, for instance by reading more slowly or revisiting parts of the text more frequently.

Second, a main effect regarding the effects of WMC was expected, with students having higher WMC showing better learning outcomes.

Third, WMC was expected to moderate the disfluency effect on learning outcomes, with the effect being more pronounced for students with higher WMC.

In addition, we explored changes in eye movement measures over time, with participants possibly adapting to the disfluency manipulation during reading. The disfluency manipulation and the fluent condition were expected to become increasingly similar during the experiment. Adaptation over time has been observed by Yue et al. (2013) Exp. 1a, who found that students’ judgments of learning became more accurate over time in the disfluent condition, but became less accurate in the fluent condition. Thus, students in the disfluent conditions reduced their overconfidence, once they noticed that the text was hard to decipher. In the current study, we were interested in whether adaption to using hard-to-read text would also be evident with regard to students’ text processing behavior.

Method

Participants and design

A total of 60 native speakers of Swedish (41 female) between 18 and 30 years of age took part in the experiment. The participants were recruited through student lists and advertisements at the Lund University campus. All participants had normal or corrected-to-normal vision. Due to technical problems with the recording and recording equipment, as well as insufficient data, seven participants were excluded from the results. This resulted in 53 participants in total ($M=21.7$ years, $SD=2.96$). The participants were naïve to the purpose of the experiment. They were randomly assigned to one of two experimental conditions: a control condition in which the text students were expected to study was presented regularly, and the disfluency condition, where the text presentation was manipulated to make it perceptually harder to read.

Materials

A text detailing how airplanes achieve lift (Mautone and Mayer 2001) was shown on four consecutive *text presentation screens* (hereby, *TPSs*), with participants skipping to the next TPS in their own pace (see Appendix 1). Overall text length was 280 words. The text had previously been translated into Swedish, and adapted to on-screen reading by using a large font, namely Arial at a 48 point size. The individual lines of the text were separated by one blank line, in order to facilitate accurate eye-tracking measurements. In the control condition the text was presented using regular font. In order to not change the size, structure, or position of the text between the two conditions, the disfluency manipulation was generated by low-pass filtering each color band in the original text with an averaging low-pass filter of size 8×8 pixels ($SD=8$ pixels). This only removed high-frequency information from the font, leaving everything else intact. An example stimulus pair is shown in Fig. 1a and b.

Measures

To assess working memory capacity, an automated OSPAN working memory task (Unsworth et al. 2005) was completed by the participants. Participants were presented with letters in set sizes ranging from 3 to 7. Math problems were shown between each letter presentation, and after solving the problem and continuing to the next screen, participants needed to indicate whether a suggested solution was correct or not. After each set, they needed to correctly recall

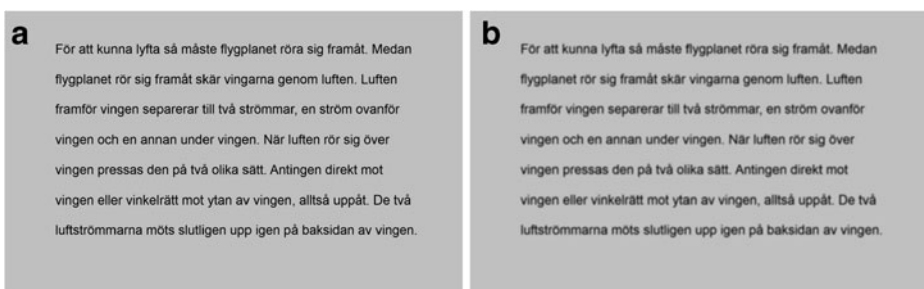


Fig. 1 **a** Example stimuli, fluent condition. **b** Example stimuli, disfluent condition

the letters in order of presentation. WMC is reflected in the measure *OSPAN absolute score*, which signifies the number of perfectly recalled letter sets (max 75).

During the reading task, students' eye movements were recorded and several dependent variables were derived from this data. A word-by-word analysis of eye movement measures is appropriate, as it allows the highest sensitivity with regards to average fixation durations and total reading times. As words are skipped during reading, predominantly function words, removing these will improve results for the recorded data. Reading comprehension in general is often linked to higher-level reading indicators such as first- and second-pass reading times. Therefore, in addition to a word level analysis, individual lines were used in the current study. These analyses contain all data, even lines that were completely skipped by participants (39 lines in total over all trials and all participants).

When investigating each word, the total time spent on a word (total reading time) and the mean duration of the fixations performed on a word (average fixation duration) were used. Only words with a minimum of one fixation was used in the analyses. When investigating each line of text, time spent reading a line of text during the first entry only (first-pass reading time), time spent reading a line of text during the first return to the line only (second-pass reading time), and total reading time were used. All data was used, including lines that were completely skipped. All eye-tracking data were normalized for all statistical models and plots in order to compensate for variance between different TPSs. Using total reading time as an example, the fluent condition was normalized to zero by subtracting the mean total reading time for each individual word or line within one single TPS with the mean total reading time of all words or lines in the corresponding TPS. The disfluent condition was then normalized with respect to the fluent condition by subtracting the mean total reading time for each individual word or line within one single TPS with the mean total reading time of all of the words or lines in the corresponding fluent TPS. This resulting variable is named *normalized total reading time*.

As a dependent performance variable, a free recall task was completed with pen and paper subsequent to reading the text. Participants were prompted (in Swedish) to "Kindly write as much as you remember about how airplanes achieve lift. You have 4 min." Students' responses were analyzed with respect to whether they had correctly mentioned each of the 7 keys ideas introduced in the text found in Appendix 2. They received one point for each correctly mentioned idea. The maximum possible score was 7. A forced-choice verification test was also completed by the participants, consisting of 11 true/false questions. However, due to a very low Cronbach reliability score (0.4), the results from this task is not reported or used in any statistical models.

Prior knowledge about the topic (PK) was self-reported by participants on a scale ranging from 1 to 5, where 1 signified no prior knowledge regarding the topic, and 5 signified knowing everything about the topic (no participant responded with a 5).

Apparatus

Eye movement data were recorded binocularly with the Smart Binocular setting at a sampling rate of 120 Hz, using a RED-m remote video-based eye-tracker from SensoMotoric Instruments (Teltow, Germany) in a one-computer setup with one screen. The recordings took place in the Digital Classroom at the Humanities Laboratory, Lund University. The Digital Classroom uses 25 remote eye-trackers to conduct simultaneous recording of multiple participants, thereby achieving a more classroom-like setting. The distance from each participant's eyes to the stimulus monitor was approximately 600 mm. Stimuli were displayed on a Dell P2210 22" widescreen LCD display at a resolution of 1680×1050 pixels (475×300 mm,

equivalent to approximately 43.2×28.1 degrees of visual angle) with a refresh rate of 60 Hz. The eye-tracking system was controlled with SMI iView RED-m, while stimulus presentation, a 5-point calibration, and a 4-point validation of the calibration accuracy were handled using SMI Experiment Center 3.1. Calibration accuracy was under 1.0 degrees of deviation in both the horizontal and the vertical direction for all participants.

Procedure

The experimental session lasted around 40 min, with no time restrictions imposed during any part of the experiment except for the free recall task. The participants were only instructed that they would participate in a learning study. They were informed that their participation was completely voluntary and that they could stop the experiment at any time. The procedure of the experiment was then quickly explained by the experimenter, but participants were also told that written instructions would be provided during each part of the experiment.

The participants were seated in front of the eye-trackers and told that the experiment consisted of a text presented on four separate text presentation screens. Pressing the space bar would take them to the next portion of the text, and returning to a previous screen was not possible. They were instructed to read the texts carefully, as their retention of the material would be tested after a period of time. An OSPAN task would follow, introduced like a memory task to participants, where letters needed to be recalled in order while solving math problems. After this, a task regarding the previously read text would be needed to be completed on a separate piece of paper, followed by another task regarding the text completed on the computer. Then the eye tracker was calibrated.

The experiment started with the participant completing a questionnaire about their age, gender, educational background, and PK. The instructions were repeated in written form, and the texts followed. After the texts, participants completed an OSPAN working memory span task (Unsworth et al. 2005). The OSPAN task took around 25 min to complete, after which participants completed the free recall task. After this, the experiment was over. The participants were told about the purpose of the experiment, given a debriefing form to take home, thanked again for their participation, and given a movie ticket as compensation.

Data analysis

The eye-tracking data recorded binocularly in iView RED-m with the Smart Binocular setting is automatically transformed into one average using data from both eyes. Velocity-based high-speed event detection was performed using SMI BeGaze 3.4 with default settings (peak velocity threshold = $40^\circ/\text{s}$, minimum fixation duration = 50 ms), which transformed the raw data into fixations and saccades. The data processing, statistical analyses, and plots were made using R (3.1.1) (R Development Core Team 2014) and the lme4 package (1.1–7) in R Studio (0.98.953).

Linear mixed effects models were used when investigating eye movement measures and generalized linear models were used when investigating learning outcomes. Average fixation durations were log-transformed to better fit a Gaussian distribution. OSPAN absolute scores were centered when used in statistical models. All models used a fixed effect with two levels (fluent and disfluent condition). Linear mixed effects models offer several advantages to using ANOVAs when analyzing eye-tracking data, as models compensate for uneven distributions of data, which occur when performing eye-tracking studies (Bates et al. 2012). Also, models can be created to compensate for individual differences (Baayen et al. 2008). All models investigating

eye movements were created using random intercepts and slopes of condition for participants, with intercept-slope correlations. When models did not converge, they were reduced incrementally. As a consequence, first-pass reading times, second-pass reading times, and total reading times for lines of text used random intercepts for participants only. Linear mixed effects models report t -values, with a t -value below -1.96 and above 1.96 generally considered significant. This also depends on degrees of freedom, so Satterthwaite approximation was used to calculate p -values for all linear mixed effect models using the `lmerTest` package.

Results

Learning outcomes

Learning outcomes from the free recall task can be found in Table 1. In addition to results for the entirety of the experiment, scores for the information given in TPS 1 and 2, as well as in TPS 3 and 4, can also be found in Table 1. This is due to an adaption effect found in eye movement measures, which is further explored in the section *Adaptation of Information Processing Over Time* below. In a first step, we tested whether the text would be equal in terms of students' prior knowledge on the topic (PK), with no significant differences found (Estimate=0.592, $SE=0.351$, $t=1.68$, $p=0.10$). We also examined the interaction between condition and PK, with no significant results found (Estimate=-0.121, $SE=0.428$, $t=-0.28$, $p=0.78$). Consequently, the results can be interpreted unambiguously. Contrary to what would be expected according to the disfluency effect, learning outcomes did not differ significantly between the fluent and the disfluent condition (Estimate=-0.439, $SE=1.56$, $t=-0.28$, $p=0.78$). However, participants with higher OSPAN absolute scores had significantly better results compared to participants with lower OSPAN absolute scores (Estimate=0.044, $SE=0.019$, $t=2.33$, $p=0.02$). OSPAN absolute scores did not interact with condition (Estimate=-0.025, $SE=0.025$, $t=-1.01$, $p=0.32$). PK scores and OSPAN absolute scores can be found in Table 2. Figure 2 shows the score in the free recall task as a function of OSPAN score.

Information processing

Mean scores for the eye-tracking variables can be found in Tables 3 and 4. Table 3 shows data using individual words as areas of interest, and Table 4 shows data using individual lines as areas of interest. In Tables 3 and 4, *disfluent-fluent comparison* shows the mean difference between conditions using normalized measures. When examining reading over the whole experiment, there was no significant effect of condition for total reading times (Estimate=0.012, $SE=0.077$, $t=0.16$, $p=0.88$) or average fixation durations (Estimate=0.034, $SE=0.039$,

Table 1 Mean learning outcomes from the free recall as a function of experimental condition

TPSs	Fluent	Disfluent	Both conditions	Maximum score
TPS 1 & 2	1.62 (0.152)	1.53 (0.124)	1.58 (0.097)	3
TPS 3 & 4	1.79 (0.188)	1.57 (0.190)	1.68 (0.133)	4
Total	3.41 (0.283)	3.10 (0.273)	3.25 (0.196)	7

Note: Standard errors in parentheses

Table 2 Self-reported previous knowledge and OSPAN absolute scores

Measure	Fluent	Disfluent	Both conditions	Maximum score
<i>Previous knowledge</i>	1.62 (0.167)	2.04 (0.223)	1.83 (0.142)	5
<i>OSPAN absolute score</i>	36.3 (2.75)	42.5 (3.39)	39.4 (2.20)	75

Note: Standard errors in parentheses

$t=0.88$, $p=0.39$) on the words in the text. No significant effects of condition were found for first-pass reading times (Estimate=84.2, $SE=175$, $t=0.48$, $p=0.63$), second-pass reading times (Estimate=-375, $SE=470$, $t=-0.80$, $p=0.43$), or total reading times (Estimate=-869, $SE=561$, $t=-1.55$, $p=0.12$) on the lines in the text.

Adaptation of information processing over time

To test whether students had adapted their reading behavior to the disfluency manipulation over the course of time, we also analyzed its effects relative to the position of the text part.

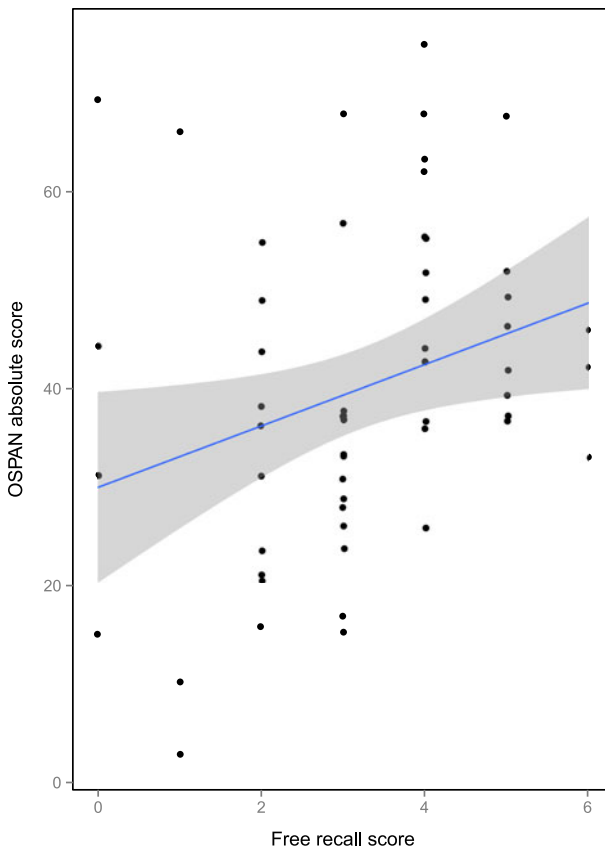


Fig. 2 Free recall task scores as a function of working memory capacity (OSPAN score). The linear regression line includes a 95 % confidence interval

Table 3 Participant mean data per word as a function of experimental condition

Measure	Fluent	Disfluent	Disfluent-fluent comparison
Total reading time			
TPS 1	599 (14.7)	530 (12.0)	-69.5
TPS 2	544 (15.8)	485 (10.5)	-58.8
TPS 3	512 (15.2)	566 (16.8)	54.2
TPS 4	574 (23.2)	614 (21.6)	40.4
Total	560 (8.49)	543 (7.44)	-16.1
Average fixation duration			
TPS 1	206 (2.50)	210 (2.25)	4.39
TPS 2	204 (2.49)	209 (2.22)	4.52
TPS 3	206 (3.74)	206 (3.83)	0.397
TPS 4	193 (2.20)	204 (2.61)	10.8
Total	203 (1.39)	208 (1.36)	4.78

Total reading times and average fixation durations (in ms) for all words in texts per TPS, as well as over entire experiment. Only data for words with a minimum of one fixation was used. Disfluent-fluent comparison signifies the differences between conditions and is shown using the measure *normalized total reading time* in Fig. 3

Note: Standard errors in parentheses

Table 4 Participant mean data per line as a function of experimental condition

Measure	Fluent	Disfluent	Disfluent-fluent comparison
First-pass reading time			
TPS 1	1323 (82.9)	1325 (83.4)	-9.01 (58.8)
TPS 2	991 (62.5)	1310 (70.1)	162 (47.6)
TPS 3	1010 (79.6)	1219 (87.8)	107 (60.0)
TPS 4	1019 (86.8)	1146 (87.1)	64.3 (61.5)
Total	1093 (38.9)	1257 (40.7)	83.8 (28.2)
Second-pass reading time			
TPS 1	2798 (229)	2738 (205)	-30.2 (153)
TPS 2	1953 (176)	1690 (129)	-134 (109)
TPS 3	2146 (197)	2500 (200)	181 (140)
TPS 4	2291 (269)	2621 (246)	168 (182)
Total	2290 (109)	2349 (97.3)	30.0 (72.3)
Total reading time			
TPS 1	4847 (301)	4064 (207)	-432 (178)
TPS 2	3144 (225)	3060 (145)	-45.8 (129)
TPS 3	3316 (241)	3720 (203)	223 (156)
TPS 4	3511 (327)	3768 (254)	141 (203)
Total	3705 (139)	3624 (101)	-44.6 (82.7)

First-pass reading times, second-pass reading times, and total reading times (in ms) for all lines in texts per TPS, as well as over entire experiment. Disfluent-fluent comparison signifies the differences between conditions and is shown using the measure *normalized total reading time* in Fig. 4

Note: Standard errors in parentheses

When taking into account the time course presentation of the four parts of the texts, a significant interaction was found between condition and TPS for total reading times on the words in the text. In TPS 1 and 2, total reading times on the words were shorter in the disfluent condition compared to the fluent condition, but total reading times was longer in TPS 3 and 4 (Estimate=0.069, $SE=0.014$, $t=4.81$, $p<0.0001$), explaining why there had been no overall effect of disfluency on total reading times. This effect was also found when using lines as area of interest (Estimate=371, $SE=128$, $t=2.89$, $p<0.004$).

Figure 3 shows mean normalized total reading times on all words for each TPS as a function of condition, and Fig. 4 shows this data for all lines. No significant interaction was found for average fixation durations on words (Estimate=0.00015, $SE=0.0063$, $t=-0.024$, $p=0.98$), or for first-pass reading times (Estimate=33.3, $SE=46.6$, $t=0.72$, $p=0.47$) or second-pass reading times on lines (Estimate=180, $SE=117$, $t=1.55$, $p=0.12$).

The aforementioned analyses suggest that students responded differently to the disfluency manipulation as a function of text position, thereby adjusting to disfluency. To test whether this adaptation would also have an effect on students' learning outcomes, results from the learning outcome measures were divided into two parts depending on which TPS contained the information necessary to answer the question correctly (TPS 1 & 2 vs. TPS 3 & 4). These results can be found in Table 1. However, no significant effects of condition were found (TPS 1 & 2: Estimate=0.066, $SE=0.558$, $t=0.12$, $p=0.91$; TPS 3 & 4: Estimate=0.359, $SE=0.719$, $t=0.50$, $p=0.62$), suggesting that while students had adapted to the disfluency manipulation at the processing level, this did not have any profound effect on learning outcomes.

Discussion

Learning outcomes did not differ significantly between the two conditions in the free recall task. The pattern of results thus does not confirm the disfluency effect, which predicts that a very subtle visual manipulation on text, such as substituting a font for another or using bold-face (Diemand-Yauman et al. 2011), can lead to positive effects in learning outcomes (Kühl and Eitel *in press*). However, it aligns with a number of failures to replicate the disfluency effect that have accumulated recently (Eitel et al. 2014; Meyer et al. 2015; Yue et al. 2013). The current study took these failures to replicate as a starting point to investigate possible explanations for the unreliability of the disfluency effect.

Fig. 3 Normalized total reading times on individual words. Mean differences in ms between the fluent and disfluent condition as a function of presentation screen. Error bars denote standard errors of the mean

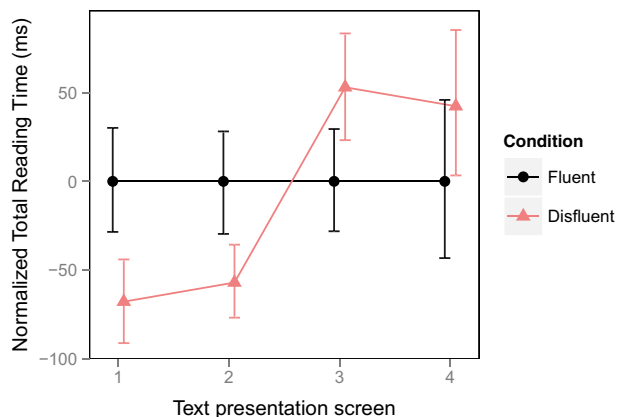
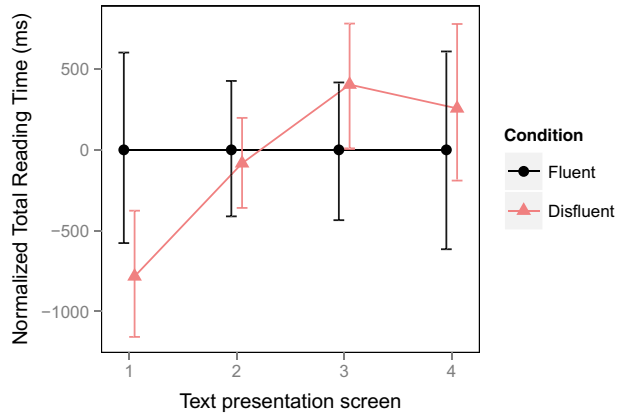


Fig. 4 Normalized total reading times on individual lines of text. Mean differences in ms between the fluent and disfluent condition as a function of presentation screen. Error bars denote standard errors of the mean



According to the first explanation, the disfluency effect may be hard to find at a coarse level of observation, such as in learning outcome measures. Rather, it may sometimes be observable only for more fine-grained information processing measures. To test this assumption, eye tracking was used to test whether students' attention distribution would be affected by manipulating the fluency of the to-be-studied text. Contrary to what would have been expected, no overall disfluency effects were found for total reading times on either words or lines, or for average fixation duration, thus not confirming our first hypothesis. Only when analyses were done separately for each TPS did the disfluency manipulation have an effect on students' eye movement behavior, suggesting that students did alter their reading strategies over time. We will return to these findings later.

According to the second explanation, failures to replicate the disfluency effect may be moderated by the cognitive prerequisites a learner brings to the task, especially by their WMC. In particular, it was expected that the disfluency effect might be more likely to occur for students with high compared with low WMC, who have sufficient cognitive resources to self-regulate their processing and to invest them into more effortful study (cf. Lehmann et al. *in press*; Ilkowska and Engle 2010; van Merriënboer and Sluijsmans 2009). In line with previous research, WMC was predictive for learning outcomes at least with regard to students' performance in the free recall task, thereby at least partially confirming our second hypothesis (cf. Daneman and Merikle). However, in contrast to our third hypothesis, WMC did not moderate the disfluency effect at an outcome level. Thus, from the data we can rule out the WMC-explanation for why other studies did not find a disfluency effect. Prior knowledge was examined in the current study, and shown not to affect free recall performance. Participants reported their subjective estimation of their prior knowledge, which is a limitation when measuring previous knowledge. However, domain-specific questions in the topic could affect the results in the free recall task. As the influence of prior knowledge was not a primary focus, the results for free recall were prioritized.

A possible alternative explanation for why we and others did not find a disfluency effect for learning outcomes may have to do with the nature of the materials used. This brings us back to our findings regarding how students adapted to the disfluency manipulation over the course of reading the four pages of text. According to the underlying explanation of the disfluency effect, learners studying hard-to-read text should invest more effort into their processing of the text (Yue et al. 2013). Thus, we expected them to show longer total reading times on both

words and lines, and possibly longer average fixation durations. However, in contrast to these expectations, students had shorter total reading times during the first two TPSs in the disfluent condition compared to the fluent text condition, but longer total reading times during the last two TPSs. These results were present for both words and lines.

These findings may be interpreted as evidence for the fact that, when encountering disfluent text, students first lowered their effort and only increased it after a period of time. It is possible that participants were unconsciously disturbed by the disfluency manipulation initially, causing them to spend less time on the text overall, before commencing a more effortful reading. Interestingly, a small (non-significant) increase in reading speed was found for text with a very low level of blur in Chung et al. (2007), after which it decreased with increased levels of blur. An analogous case of adaptation over time has been observed by Yue et al. (2013) Exp. 1a, using judgments of learning as main dependent variable. Overall, our findings and those by Yue et al. (2013) suggest that beneficial effects of disfluency may appear only later in the course of instruction, once students have found ways to adequately regulated their effort. It is possible that disfluency serves as a desirable difficulty, but only for longer materials, where sufficient time allows for self-regulatory processes such as effort adaptation to commence. In that sense, disfluency in the present study may have served as a metacognitive cue, but this cue may have been insufficient to induce a change in learning behavior that would pay off in terms of improved learning outcomes. Diemand-Yauman et al. (2011) manipulated students' course materials over a whole semester finding a disfluency effect for complex learning tasks, thus, suggesting that students responded to the disfluency manipulation in ways that affected memory and comprehension. On the other hand, Eitel et al. (2014) manipulated a single-page instruction only and were unable to find a disfluency effect in 3 out of 4 studies.

Future research should investigate this post-hoc explanation further by using longer materials extending over multiple pages, so that beyond becoming aware of the increased processing difficulties students also have a chance of adjusting their learning behavior accordingly. The use of more comprehensive materials, for which students can monitor and regulate their learning, would also better reflect the complexity of real-world learning scenarios. Moreover, it would also allow designing sufficiently comprehensive learning outcome tests so that learning outcomes can be reliably analyzed as a function of different phases of effort regulation. Investigating the emergence of the disfluency effect over time could have important consequences for theory building in this area. So far, the explanation of the disfluency effect suggests a two-stage process consisting of becoming meta-cognitively aware of processing challenges and regulation of learning behavior to accommodate these challenges. However, there is little direct evidence regarding either process. With the present study, we believe to have contributed evidence favoring this two-stage process assumption. In terms of methodological advances, future studies should not only use eye tracking, but also other measures to study adaptation over time, since the interpretation of eye movements as higher-level process indicators is often ambiguous. Combining eye tracking with concurrent verbal protocols would for instance allow investigating whether increased reading times really reflect intentional effort investment to compensate for increasing reading difficulties, in which case students should be able to report on their self-regulatory decisions. Moreover, concurrent verbal protocols could provide another way of finding evidence for the two-state process described earlier.

Another more conceptual issue to be addressed in future studies is whether some of the variability in the disfluency effect is due to the type of fluency manipulation deployed. In the current study, low-pass filtering of letters was used. By visually degrading text using low-pass

filtering, previous studies found a cost in low-level information processing behavior, resulting in longer total reading times when the stimuli needed to be decided (Jainta et al. 2011; Jordan et al. 2012). It might be considered a rather subtle manipulation compared with the manipulations deployed by Eitel et al. (2014) or Diemand-Yauman et al. (2011), who used hard-to-read photocopies produced by moving the paper while photocopying it. It is yet an open question of what constitutes a hard-to-read text and whether any text manipulation at the perceptual level can be assumed to lead to the same regulatory adjustments regarding higher-level processes. Thus, before attempting further conceptual replications of the initial disfluency effect found by Diemand-Yauman et al. (2011), a better definition of the concept in question is needed.

The visual degradation in the current study produced observable eye movement effects, albeit not the hypothesized ones. Visually degraded stimuli has been shown to increase reaction times (Holcomb 1993) and increase reliance on context in lexical decision tasks (Becker and Killion 1977; Holcomb 1993). In the current study, no effects were found in early measures such as first- and second-pass reading times, or in average fixation durations. However, a delay was observed before the hypothesized longer total reading times of the disfluent text, showing that the reading of visually degraded text is a more complex process than previous research on disfluency suggest. This delay is surprising and should be investigated further with other types of visual degradation using text reading over periods of time. Also, future research should investigate the source of this delay. It is possible that the delay depends on the instruction given to participants, which in this study was to learn the text thoroughly. If instructions are to merely read for comprehension, results might be different.

As noted earlier, the present study has its limitations. It used only a relatively short text, which may be an important boundary condition of the disfluency effect as discussed earlier. Even though it is the first study to provide fine-grained processing data regarding the time course of responding to hard-to-read text, the interpretation of the eye movement data in isolation is not unequivocal. Furthermore, the study used a relatively subtle manipulation of the text's visual appearance, which at present cannot be ruled out as a potential reason for the lack of a disfluency effect at the outcome level.

The current study mainly aimed at studying the processes underlying the disfluency effect, which it failed to replicate. While the study can offer only suggestions for future research regarding possible reasons for the heterogeneity of the disfluency effect, it shows that visually degrading text can produce complex effects not necessarily connected to text processing and text comprehension. In terms of implications for education, the current study recommends that disfluency manipulations should not be used as instructional interventions, in contrast to what has been suggested by Diemand-Yauman et al. (2011). Even though hard-to-read text did not harm learning in the present or any of the previous failed replications, it does not provide any grounds for offering badly designed materials to students. This line of research may nevertheless reveal important insights for educational practice in the future, should there be further evidence for the mechanism supposed to underlie the disfluency effect. In this case, further educational research could look at other ways of increasing students' awareness of processing difficulties, which in turn yields more effort investment. Thus, even though the experimental manipulation as such may be unsuited for educational practice, because, among other things, it contradicts students', teachers', and parents' expectations of well-designed and hence effective learning materials, the general principle may be suited to design effective, and more socially accepted instructions.

Acknowledgments This research has been funded by a grant from the *Marcus och Amalia Wallenberg Stiftelse* for the project “EyeLearn: Using visualizations of eye movements to enhance metacognition, motivation and learning.”

Appendix 1: stimuli

Text was given in Swedish to participants, on four separate TPSs. Each TPS is indicated below.

TPS 1

In order to achieve lift, the airplane must move forward. As the airplane moves forward, its wings cut through the air. This allows the air to flow over the wing. The air hitting the front of the wing separates. Some air flows over the wing and some flows under the wing. As the air moves across the wing, the air pushes directly against the wing, or perpendicular to the surface of the wing. The air eventually meets up again at the back of the wing.

TPS 2

Air flows across the top surface of the wing differently than it flows across the bottom surface. The air flowing over the top of the wing has a longer distance to travel in the same amount of time as the air flowing under the wing. As a result, air traveling over the curved top of the wing flows faster than the air that flows under the bottom of the wing.

TPS 3

As a result of the air on the top of the wing traveling faster, the air pressure on the top surface of the wing differs from that on the bottom of the wing because air pressure decreases when air moves faster. As a result, the pressure on the top part of the wing decreases. The top surface of the wing now has less pressure exerted against it than the bottom surface of the wing.

TPS 4

The downward force of the faster-moving air on the top of the wing is not as great as the upward force of the slower moving air under the wing. This creates more upwards-directed force on the wing than downwards-directed force. This results in lift for the airplane, which can take off.

Appendix 2: core idea units in the free recall task

1. The upper surface of the wing is more curved (or longer).
2. The bottom surface is less curved (or shorter).

3. Air flows faster over the top surface of the wing (or air flowing over the top surface of the wing has a longer distance to travel in the same amount of time).
4. Air traveling over the bottom surface of the wing is slower (or has less distance to travel).
5. The top surface of the wing has less pressure (or pressure on the top of the wing decreases, or air on the top of the wing gets more spread out, or the downward force is not as great).
6. The bottom of the wing has more pressure (or the upward force is greater).
7. There is a net upward force on the wing, a lift.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005.
- Bates, D., Mächler, M., & Bolker, B. (2012). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*.
- Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(3), 389–401. doi:10.1037/0096-1523.3.3.389.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. doi:10.1146/annurev-psych-113011-143823.
- Chung, S. T. L., Jarvis, S. H., & Cheung, S.-H. (2007). The effect of dioptric blur on reading performance. *Vision Research*, *47*(12), 1584–1594. doi:10.1016/j.visres.2007.03.007.
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Elsevier Ltd. <http://doi.org/10.1016/b978-008044980-7/50017-3>.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. doi:10.3758/bf03196772.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. doi:10.1016/S0022-5371(80)90312-6.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: a meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. doi:10.3758/BF03214546.
- 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*(1), 111–115. doi:10.1016/j.cognition.2010.09.012.
- Ehmke, C., & Wilson, S. (2007). Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it. Volume 1* (pp. 119–128). British Computer Society. Retrieved from <http://dl.acm.org/citation.cfm?id=1531294.1531311>
- Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: evidence for scaffolding. *Learning and Instruction*, *28*, 48–63. doi:10.1016/j.learninstruc.2013.05.002.
- Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: does harder-to-read mean better-to-understand? *Applied Cognitive Psychology*, *28*(4), 488–501. doi:10.1002/acp.3004.
- 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.
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61. doi:10.1111/j.1469-8986.1993.tb03204.x.
- Hooge, I. T. C., & Erkelens, C. J. (1999). Peripheral vision and Oculomotor control during visual search. *Vision Research*, *39*(8), 1567–1575. doi:10.1016/S0042-6989(98)00213-2.
- Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In R. Radach, J. Hyönä, & H. Deubel (Eds.), *The mind's eye: cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science. doi:10.1016/b978-044451020-4/50018-9.
- Ilkowska, M., & Engle, R. W. (2010). Working memory capacity and self-regulation. In R. H. Hoyle (Ed.), *Handbook of personality and self-regulation* (pp. 265–290). Malden: Wiley-Blackwell. doi:10.1002/9781444318111.ch12.

- Jainta, S., Dehnert, A., Heinrich, S. P., & Jaschinski, W. (2011). Binocular coordination during reading of blurred and nonblurred text. *Investigative Ophthalmology & Visual Science*, *52*(13), 9416–9424. doi:10.1167/iov.11-8237.
- Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *18*, 178–191. doi:10.1037/a0026923.
- Jordan, T. R., McGowan, V. A., & Paterson, K. B. (2012). Reading with a filtered fovea: the influence of visual quality at the point of fixation during reading. *Psychonomic Bulletin & Review*, *19*(6), 1078–1084. doi:10.3758/s13423-012-0307-x.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183. doi:10.1037/0096-3445.130.2.169.
- Kühl, T., & Eitel, A. (in press). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning*.
- Legge, G. E., & Cheung, S. (2004). 49.1: Invited Paper: Psychophysics of reading: Implications for displaying text. In *SID Symposium Digest of Technical Papers* (Vol. 35, No. 1, pp. 1359–1361). Society for Information Display. Blackwell Publishing Ltd. doi:10.1889/1.1825762
- Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading: I. Normal vision. *Vision Research*, *25*(2), 239–252. doi:10.1016/0042-6989(85)90117-8.
- Lehmann, Goussios, C., & Seufert, T. (in press). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning*.
- Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*, 95–109. doi:10.1016/j.compedu.2012.07.011.
- Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, *93*, 377. doi:10.1037/0022-0663.93.2.377.
- Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16–e30. doi:10.1037/xge0000049.
- R Development Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: <http://www.r-project.org/>. Retrieved from <http://www.r-project.org/> (Version 3.0.3)
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi:10.1080/17470210902816461.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. doi:10.3758/bf03197692.
- Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, *20*(1), 6–10. doi:10.1111/j.1467-9280.2008.02243.x.
- Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41. doi:10.1016/j.cogpsych.2004.11.001.
- Rehder, B., Colner, R. M., & Hoffman, A. B. (2009). Feature inference learning and eyetracking. *Journal of Memory and Language*, *60*(3), 393–419. doi:10.1016/j.jml.2008.12.001.
- Reingold, E. M., & Rayner, K. (2006). Examining the word identification stages hypothesized by the E-Z reader model. *Psychological Science*, *17*(9), 742–746. doi:10.1111/j.1467-9280.2006.01775.x.
- Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction*, *36*, 11–26. doi:10.1016/j.learninstruc.2014.11.002.
- Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology*, *23*, 1227–1243. doi:10.1002/acp.1526.
- Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: evidence from the serial reaction time task. *Memory & Cognition*, *33*(2), 213–220. doi:10.3758/bf03195310.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. doi:10.3758/BF03192720.
- Van Merriënboer, J. J. G., & Sluijsmans, D. M. A. (2009). Toward a synthesis of cognitive load theory, four-component instructional design, and self-directed learning. *Educational Psychology Review*, *21*, 55–66. doi:10.1007/s10648-008-9092-5.
- 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 & Cognition*, *41*, 229–241. doi:10.3758/s13421-012-0255-8.