

Null Effects of Perceptual Disfluency on Learning Outcomes in a Text-Based Educational Context: a Meta-analysis

Heping Xie^{1,2} · Zongkui Zhou^{1,2} · Qingqi Liu^{1,2}

Published online: 20 June 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Recent research in a text-based educational context has demonstrated a seemingly paradoxical disfluency effect in reading, namely that learning with hard-to-read (disfluent) materials helps learners recall more details than learning with easy-to-read (fluent) materials. Many follow-up studies using a variety of participants, learning materials, and experimental designs have been conducted to verify the effects of disfluency manipulation on recall, transfer, judgments of learning, and learning time. However, a number of them have failed to replicate this effect and the mixed findings bring into question the generality of the disfluency effect with respect to learning. In this meta-analysis, we tested the overall effect of perceptual disfluency on learning with texts, as well as moderators of this effect, based on 25 empirical articles involving 3135 participants. Results showed that overall, there was no effect of perceptual disfluency on recall ($d = -0.01$) or transfer ($d = 0.03$), but perceptual disfluency did reduce participants' judgments of learning ($d = -0.43$) and increase learning time ($d = 0.52$). Tests of moderation focused on the most commonly studied dependent measure, namely recall. There was no evidence that characteristics of the participants, learning material, or experimental design moderated the effect of perceptual disfluency on recall. In general, though perceptual disfluency can be used as an effective metacognitive cue to reduce judgments of learning and increase learning time, there is not enough evidence to show that it either stimulates analytic processing or increases extraneous cognitive load.

Keywords Perceptual disfluency · Recall · Transfer · Judgments of learning · Learning time · Meta-analysis

✉ Zongkui Zhou
zhouzk@mail.ccnu.edu.cn

¹ Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, 152 Luoyu Street, Wuhan, China

² School of Psychology, Central China Normal University, 152 Luoyu Street, Wuhan, China

Introduction

Perceptual disfluency is an important focus of research in educational psychology (for a systematic review, see Reber and Greifeneder 2017). The term perceptual disfluency, considered to be a form of metacognition, refers to the learner's subjective sense of difficulty while completing a cognitive task. It reflects the ease of perceptual-level processes associated primarily with material form (Reber et al. 2004). For example, if the font of a visual text is irregular, we are aware that it is difficult to process. If the quality of a spoken text is poor, we recognize the challenge in decoding it. A paradoxical finding in some studies is that students learn better when they perceive disfluency—that is, when educational materials are slightly harder, rather than easier, to read. This effect was initially found by Diemand-Yauman et al. (2011), who provided empirical evidence that learning with disfluent (hard-to-read) materials helped learners recall more details of the information than learning with fluent (easy-to-read) materials, not only in a controlled laboratory setting but also in actual classroom environments. These intriguing findings, going against common sense, have attracted a lot of attention, especially in the research field of educational psychology. Up to July 2017, the original article published by Diemand-Yauman et al. (2011) has triggered more than 200 citations (source: Google Scholar®), and there have been a number of follow-up studies aiming at replicating the compelling results in a text-based educational context (e.g., Eitel et al. 2014; French et al. 2013; Strukelj et al. 2016). Researchers' enthusiasm was probably aroused due to the potential for these results to be applied to education. If a simple intervention based on the disfluency effect is proved to foster learning, it could be implemented with minimal investment.

Before widely introducing disfluency-based educational interventions to actual classrooms, however, it is of the essence to evaluate the robustness of the disfluency effect. Since the work of Diemand-Yauman et al. (2011), dozens of studies using a variety of participants, learning materials, and experimental designs have been conducted to verify the effects of disfluency manipulation on recall (French et al. 2013; Guenther 2012; Lee 2013), transfer (Eitel and Kühl 2016; Eitel et al. 2014; Kühl et al. 2014a; Lehmann et al. 2016), judgments of learning (JOLs) (Pieger et al. 2016, 2017; Weissgerber and Reinhard 2017), and learning time (Pieger et al. 2016, 2017; Rummer et al. 2016; Seufert et al. 2017). However, the mixed findings in this body of research call into question the generality of the disfluency effect, leading to a call for an assessment of the overall effects of disfluency (Bjork and Yue 2016; Dunlosky and Mueller 2016). Therefore, the most important aim of the present meta-analysis was to evaluate the effects of perceptual disfluency compared to fluency in educational learning texts on different sorts of performance (i.e., recall, transfer, JOL, and learning time). In addition, many researchers tentatively noted that disfluency might facilitate learning under some circumstances but not others (e.g., Eitel and Kühl 2016; Kühl et al. 2014a; Lehmann et al. 2016). That is, there might be some undiscovered moderators influencing the effects of disfluency (Kühl et al. 2014b; Oppenheimer and Alter 2014). Thus, an exploratory aim of the study was to identify potential factors (related to participants, learning material, and experimental design) that might moderate the effect of disfluency on learning, specifically on the most frequently studied outcome—recall.

Theoretical Frameworks for Explaining Perceptual Disfluency

According to the notion of “desirable difficulties” (Bjork 1994, 2013), facing challenges in the encoding phase can help learners process information more deeply and retrieve it better later. For example, research on the generation effect (e.g., deWinstanley and Bjork 2004; Hirshman

and Bjork 1988) found that asking learners to actively generate letters when they were presented in an incomplete format (e.g., “aff_ct_v_”) resulted in better recall of the letters than asking learners to passively read the group of letters presented in their entirety (e.g., “affective”). This kind of difficulty (e.g., evoked by generation) is desirable, as more elaborative processing is activated.

Increasing the difficulty of a task is likely to trigger deeper processing, not because of an increase in the objective difficulty, but because of an increase in the subjective sense of difficulty (Alter et al. 2007; Eitel et al. 2014). This “perceptual disfluency” can be achieved by presenting words or texts in a format that makes them harder to read. An explanation of the metacognitive mechanism of perceptual disfluency has recently been proposed by the disfluency theory (Alter et al. 2007; Köhl and Eitel 2016). This theory is based on the considerations of William James (1890/1950), who proposed two distinct processing systems: one that leads to quick, effortless, associative, and intuitive processing (system 1) and another that works slowly, effortfully, analytically, and deliberately (system 2). The activation of systems 1 and 2 depends on the subjective sense of ease or difficulty of a cognitive task (Alter et al. 2007). If information processing is perceived as easy, system 1 is more likely to be activated. If, on the other hand, information processing is perceived as difficult, system 2 is more likely to be activated. Making learning materials harder to read (i.e., perceptual disfluency) can increase the perceived difficulty (Alter and Oppenheimer 2009), thus activating system 2 and stimulating deep rather than superficial processing. Deep processing in turn fosters learning performance (e.g., recall or transfer).

However, an opposite explanation is proposed by cognitive load theory (CLT, Sweller et al. 2011; Sweller et al. 1998). In the field of educational research, CLT is mainly used to explain the effects of various kinds of instructional design. When performing a cognitive task, learners will consume the cognitive resources in working memory, leading to cognitive load. Resources in working memory, however, are limited (cf. Baddeley 1992); only a small fraction of elements can be consciously handled per unit time. Therefore, one important objective of instructional design is to ensure that the cognitive load is within the learner’s working memory capacity. According to CLT, three kinds of cognitive load in working memory, namely intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane cognitive load (GCL), can be created when learning with instructional material (Sweller et al. 2011; Sweller et al. 1998). ICL is a kind of load that is not directly affected by instructional design, but is related to element interactivity (i.e., inherent complexity) in learning materials and learners’ prior knowledge. ECL is affected by the quality of instructional design. When the material is poorly presented (e.g., in a way that requires learners to split attention), high ECL will be consumed to cope with the poor design rather than the learning task per se, possibly resulting in poor learning performance. Reducing ECL would be beneficial to learning because it would allow adequate cognitive resources to be allocated to learning processes. GCL is a desirable working memory load which directly correlates with the learning task per se and contributes to the learning performance. For instance, this load can be imposed by prompting learners to generate self-explanations when learning from worked examples (Paas and Van Gog 2006; Atkinson et al. 2003).

Although not specifically proposed by CLT, perceived difficulty due to text disfluency might be assumed to be detrimental to learning because of an increase in ECL (Eitel et al. 2014). When learners are presented a specific material that is identical in different groups (i.e., without a change in ICL), making instructional texts perceptually harder to read is thought to create additional cognitive demands to cope with the low legibility of the information. This load is neither related to the learning task per se, nor helpful for learners to actively generate

information (i.e., the generation effect), leading to an increase in ECL with GCL unchanged (Eitel et al. 2014; Köhl et al. 2014a). From this perspective, learning performance (e.g., recall, transfer) will be hindered by perceptual disfluency.

In summary, these two theories make opposite predictions about the effect of perceptual disfluency on learning. Disfluency theory predicts that learners who receive disfluent materials will show better learning outcomes than learners who receive fluent materials, whereas CLT predicts that learners who receive fluent materials will show better learning outcomes than those who receive disfluent materials.

Effects of Perceptual Disfluency on Recall and Transfer

According to the empirical evidence, the effects of disfluency on recall and transfer are not consistent, and the generality of the disfluency effect with respect to text-based learning is an open question. These inconsistencies were the impetus for our meta-analytic investigation. The first important question we addressed was whether or not students who are learning with a disfluent text recall and transfer better than students learning with a fluent text. The answer to this question is key to testing the opposite predictions made by disfluency theory and CLT.

Dozens of studies have been conducted to check whether disfluency facilitates or impedes recall and transfer of learning in a text-based educational context. However, empirical evidence regarding this question is quite mixed (for overviews, see Köhl and Eitel 2016; Weissgerber and Reinhard 2017; Xie et al. 2016).

Both disfluency theory and CLT are supported only to a small extent. Only a handful of studies have found positive effects of disfluency on text-based learning outcomes, such as shallow level recall (Diemand-Yauman et al. 2011; French et al. 2013; Lee 2013; Weltman and Eakin 2014) and deep level transfer (Eitel et al. 2014, Experiment 1). For example, French et al. (2013) asked students to read a text describing facts about a fictional star at the beginning of the class. About 35 min later, students who had read the text in a disfluent font (i.e., Monotype Corsiva) were found to recall more facts than those who had read information in an easier to read font (i.e., Arial). Similarly, there are also limited findings showing that disfluency hinders educational performance in the form of recall and transfer (Carpenter et al. 2016, Experiment 3; Köhl et al. 2014a; Miele and Molden 2010, Experiment 3; Pieger et al. 2017). For example, Miele and Molden (2010) asked participants to read a brief expository text describing the ways in which television news affected its viewers and found that participants who had read disfluent text (italicized 12-point Juice ITC font) showed worse recall of explicitly stated information than those who had read the fluent text (12-point black Times New Roman font).

More importantly, it is more often the case that null (direct) effects of disfluency on educational performance are found (e.g., Ball et al. 2014; Carpenter et al. 2016, Experiments 1 and 2; Carpenter et al. 2013; Eitel and Köhl 2016; Eitel et al. 2014, Experiments 2, 3, and 4; Faber et al. 2017; Guenther 2012; Pieger et al. 2016; Sanchez and Khan 2016; Strukelj et al. 2016). Eitel et al. (2014) conducted a Bayesian analysis for conditions with disfluent text versus fluent text across their four experiments and found no overall effect of perceptual disfluency on recall and transfer. Even though some researchers used the same material as used in Experiment 1 of the original study (Diemand-Yauman et al. 2011), they did not detect differences in learning outcomes between disfluent and fluent groups (Haysom 2012; Rummer et al. 2016; Whitehouse 2011), resulting in no support for either disfluency theory or CLT.

Effects of Perceptual Disfluency on Judgments of Learning and Learning Time

Perceptual disfluency is largely assumed to reduce judgments of learning and to increase learning time. The second question we addressed was whether this pattern is apparent in text-based educational contexts. Experiments often manipulate perceptual disfluency in order to assess its effect on learning performance, but this type of manipulation is also used to investigate the effect of perceptual disfluency on metacognitive monitoring, usually inspected via JOLs. JOLs are learners' predictions about the likelihood of recalling recently studied knowledge in a subsequent recall test. For instance, participants might be asked to explicitly judge what percentage of questions about the studied text passages they would answer correctly on a continuous scale from 0 to 100 (Pieger et al. 2016, 2017).

Learners use a variety of cues to monitor a cognitive task (Koriat 1997). In dual-system processing, perceptual disfluency is an effective metacognitive cue that affects processing and judgments (Alter et al. 2007; Pieger et al. 2016). If information is processed fluently, intuitive processes in system 1 will be activated to guide judgment. However, if information is processed disfluently, this experience will function as a cue that the cognitive task is difficult and that one's intuitive judgment is likely to be wrong. In this case, more analytic and deliberate processes in system 2 will be activated and the processes will be judged with more caution.

A large number of word-learning studies indicate that when asked to make JOLs, students are likely to predict lower performance after learning disfluent items compared to fluent items (Besken and Mulligan 2013; Magreehan et al. 2016, Experiments 4 and 5; Mueller et al. 2014; Yue et al. 2013, Experiments 1a, 2a, 2b, and 3). Some studies with more complex instructional materials show the same results (Ball et al. 2014; Carpenter et al. 2013; Pieger et al. 2017; Weltman and Eakin 2014). Carpenter et al. (2013) asked learners to watch a video explaining a scientific concept and then to make a JOL about how they would perform 10 min later. They found that learners in the disfluent speaker group predicted that they would recall a smaller amount of information than those in the fluent speaker group, while the actual recall performance did not differ between the two groups. Analogously, immediately after reading text passages about social psychology, participants in Pieger et al.'s (2017) study predicted that they would perform worse for the disfluent text than for the fluent text. However, there is still a handful of evidence that based on the fluency-disfluency distinction, students do not make different predictions about the amount of information they would accurately recall (Carpenter et al. 2016).

One of the impacts of metacognitive monitoring is related to metacognitive control during learning, for example, the allocation of learning time. Students' metacognitive monitoring appears to influence decisions about how to regulate their learning time (Metcalf and Finn 2008). Under normal circumstances, students spend more time to learn hard-to-read items than to learn easy-to-read ones (for reviews, see Dunlosky and Ariel 2011; Son and Metcalfe 2000). In this sense, the time spent on learning disfluent instructional materials might be greater when compared with the time spent on learning fluent materials if learning is time-unlimited or controlled by learners themselves. This has been confirmed in several disfluency-related empirical studies (Ball et al. 2014; Eitel and Kühn 2016; Miele and Molden 2010, Experiment 3; Pieger et al. 2016, 2017), showing that learning with a disfluent text required significantly longer time than learning with a fluent text. For instance, Seufert et al. (2017, Experiment 2) set no time limits for participants and discovered that the time learners needed to read a scientific text increased with

increasing disfluency levels. Of course, it should be noted that a couple of studies using educationally relevant materials did not find an influence of disfluency on learning time (Kühl et al. 2014a; for an eye-tracking study, see Strukelj et al. 2016).

Previous and Present Meta-analyses

There has been only one published meta-analysis that has addressed questions related to those in the current study. To test whether disfluent fonts contributed to math problem-solving, Meyer et al. (2015) conducted a meta-analysis of 17 studies that attempted to replicate Alter et al.'s (2007) findings on reasoning. Overall, there was no effect of disfluency on solution rates ($d = -0.01$). Therefore, there was little evidence that disfluent fonts activated analytic reasoning in system 2.

The current meta-analysis differs from that of Meyer et al. (2015) in numerous respects. First, Meyer et al. (2015) included only studies that manipulated the disfluency of written questions on the Cognitive Reflection Test (CRT, Frederick 2005). The CRT consists of three misleading math problems (e.g., “A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? _____ cents”). In disfluency-related studies, the CRT items are usually manipulated in different fonts and directly presented to the participants (e.g., Alter et al. 2007; Thompson et al. 2013). That is, there is no learning phase before the test in these studies. However, in the field of educational psychology, learners are commonly required to learn a (visual or spoken) text first before being tested. It is the materials presented in the learning phase, rather than the test items presented in the test phase, that are manipulated to be perceptually disfluent or fluent (e.g., Eitel et al. 2014; Seufert et al. 2017). Second, what Meyer et al. (2015) were concerned with was the effect of disfluent fonts on reasoning, rather than on memory or learning outcomes (e.g., recall). Therefore, it is still an open question whether perceptual disfluency affects (text-based) memory or learning.

The current meta-analysis, unlike Meyer et al.'s (2015) work, was designed to test the influence of perceptual disfluency in a text-based educational context in which the task is analogous to a classroom learning task. It is necessary to note that we paid attention to learning with (visual or spoken) texts, rather than with words. Previous research often focused on word lists or word pairs when investigating disfluency and metacognition (for reviews, see Alter and Oppenheimer 2009; Dunlosky and Metcalfe 2009). However, text-based learning is considered to be more complex and closer to classroom learning than word-based learning (Pieger et al. 2016, 2017). When learning with words, learners only have to decipher the words at a surface level. When learning with texts, they not only have to process the words but also process information within and between sentences at a deep level to integrate this information into memory (De Bruin and Van Gog 2012) and ultimately to comprehend the text (Pieger et al. 2017). Before introducing disfluency-based manipulation to actual classrooms, it is worthwhile to understand the effect of disfluency on learning with texts.

Theory-Based Analyses Our general plan for the meta-analysis was first to test the different predictions about the effect of perceptual disfluency on learning performance, respectively from the perspectives of disfluency theory and CLT. According to disfluency theory, better recall and transfer would be expected in conditions with disfluent text compared to conditions with fluent text (Hypothesis 1a). According to CLT, better recall and transfer would be assumed in conditions with fluent text compared to disfluent text (Hypothesis 1b). We then

tested whether perceptual disfluency works as a cue for learners, both when they are predicting how well it would be for them to recall the recently studied knowledge and when they are allocating their learning time. According to the notion of dual-system processing, we expected that learners in the disfluent group would make a lower JOL (Hypothesis 2) and spend more time on learning with texts (Hypothesis 3) than learners in the fluent group.

Exploratory Analyses The mixed findings on the effects of disfluency have triggered a controversial discussion concerned with undetected moderators of these effects (Kühl et al. 2014b; Oppenheimer and Alter 2014). However, in most cases, there is no theoretical rationale or empirical evidence to justify hypotheses about moderators. On an exploratory basis, we examined three sets of study characteristics (i.e., participant, learning material, and experimental design characteristics) that might moderate the impact of disfluency; because the effect of disfluency has been most studied in terms of its effects on recall, we focused on that specific outcome. We examined the following moderators (a) prior knowledge level (none, low, medium), (b) learning material domain (science, technology, engineering, mathematics, social science), (c) pacing of presentation (self-paced vs. system-paced), (d) type of presentation (static vs. dynamic), (e) modality of presentation (only a visual format vs. both visual and auditory formats), (f) medium of presentation (screen vs. paper), (g) inclusion of images (yes vs. no), (h) fluency manipulation type (font-related vs. audio-related), (i) design of study (between-subjects vs. within-subjects), (j) learning duration (equal to or shorter than 10 min vs. longer than 10 min), (k) time interval between learning and test (equal to or shorter than 10 min vs. longer than 10 min), (l) use of distraction task (yes vs. no), and (m) expectation of testing (yes vs. no). These potential moderators should prove of interest to disfluency researchers (Kühl et al. 2014b; Oppenheimer and Alter 2014).

Method

Literature Search

To identify relevant studies on the effects of perceptual disfluency on learning educational materials, a systematic literature search was conducted by searching the electronic databases PsycINFO, Educational Resources Information Center (ERIC), Science Direct, PubMed, and ProQuest. The search keywords were “perceptual disfluency,” “perceptual fluency,” and “font” with different combinations of “judgment of learning,” “learning time,” “recall,” “transfer,” “learning outcome,” and “educational performance.” Search engines such as Google Scholar and the reference lists of the identified articles were also used. The literature search was conducted for studies published up to August 2017. Finally, we sent requests by email for researchers to provide the details of their unpublished studies.

Study Selection

Studies were selected from journal articles, dissertations, conference presentations, and unpublished research. The studies were included for analysis if they met all of these criteria: (a) they were based on an experimental design; (b) text-based rather than word-based learning materials were used; (c) a group with disfluent material was compared with a group with fluent material; (d) disfluency was manipulated by changing font-related attributes (i.e., font type,

font size, font grayscale, or font bolding) of visual texts or audio-related attributes of spoken texts (i.e., audio quality, stumble during the speech, or a foreign accent), because these intrinsic perceptual features are most often manipulated (Xie et al. 2016); (e) there was a specific learning task; (f) they measured the recall performance after the learning phase; and (g) sufficient quantitative data (e.g., means, standard deviations and n ; t test or F test values) were reported to calculate the effect size.

Studies were excluded if they did not meet the inclusion criteria mentioned above. For example, a study was not included when it used word pairs as materials (e.g., Magreehan et al. 2016) or when learners were required to complete a reasoning task, rather than a memory or learning task (e.g., Sidi et al. 2016).

Coding of Studies

Three types of information were collected from each study. First, basic information was recorded, including authors, year of publication, sample size, and experiments or conditions from which the effect sizes were computed. Second, quantitative information was collected to calculate effect sizes. All studies had information needed to calculate effect sizes associated with disfluency effects and recall; if the study also reported information on transfer, JOL, or learning time, that information was also collected. Third, characteristics related to participants, learning materials, and experimental design were coded in order to test for moderation effects. The following section provides information about how these characteristics were coded for moderation analyses.

Participant characteristic:

- (1) Prior knowledge level. From an instructional design perspective, domain-specific prior knowledge level (Kalyuga 2007; Kalyuga et al. 2003) might affect the influence of the perceptual disfluency manipulation. Participants with no domain-specific prior knowledge were classified as “none”; participants with some prior knowledge were categorized as “low” or “medium” according to the classification made by the authors.

Learning material characteristics:

- (2) Learning material domain. Research on disfluency has used instructional materials from various domains, such as science (e.g., Kühl et al. 2014a), technology (e.g., Sanchez and Khan 2016), engineering (e.g., Strukelj et al. 2016), mathematics (e.g., Weltman and Eakin 2014), and social science (e.g., Pieger et al. 2016). Coding studies into the categories of science, mathematics, and social science was straightforward. Studies using technology and engineering materials were combined into a single category (i.e., technology/engineering) to maximize the number of studies in this category.
- (3) Pacing of presentation. If participants were allowed to interactively control the presentation by themselves (e.g., start, play, pause, replay, and stop), the pacing of the presentation was categorized as self-paced. If learners had no option to interact with the presentation, it was categorized as system-paced.
- (4) Type of presentation. Educationally relevant materials can be designed to be either static or dynamic (Tversky et al. 2002). Static learning materials, in which the elements remain still, are very common in educational environments (e.g., electronic text or printed text with static pictures). With the development of newer technologies, dynamic visual

materials (e.g., animations, videos) are increasingly being used to provide richer and more vivid instructional messages that are changeable over time. The type of information presentation (static or dynamic) was extracted.

- (5) **Modality of presentation.** Humans possess dual modalities for respectively processing learning materials, namely the visual modality (e.g., for pictures, printed text) and the auditory modality (e.g., for narration) (Mayer 2009). A single modality was coded when the learning material was presented only in a visual form. If the material was presented in both formats (i.e., visual and auditory), the modality of presentation was classified as dual.
- (6) **Medium of presentation.** Different kinds of media, such as screen and paper, can be used to present learning materials. The presentation medium of the learning material (screen or paper) was extracted.
- (7) **Inclusion of images.** Many studies (e.g., Faber et al. 2017; French et al. 2013; Lehmann et al. 2016) used pure text to present instructional materials as in the original work (Diemand-Yauman et al. 2011, Study 1). Some studies, meanwhile, added accompanying images (e.g., static picture, animation, or video) in the materials. Studies were coded according to whether there were learning-related images or not.

Experimental design characteristics:

- (8) **Fluency manipulation type.** Because the perceptual fluency of materials was usually manipulated with features of font (e.g., font type, font size, grayscale) or quality of spoken text, the fluency manipulation type was classified as font-related or audio-related.
- (9) **Design of study.** Most studies examining disfluency effects on learning with texts have used a between-subjects design, such that a portion of participants are asked to learn a disfluent text, whereas the rest of the participants are asked to learn a fluent one. However, a within-subjects design has also been used in some studies (e.g., Ball et al. 2014; Katzir et al. 2013; Lee 2013). Thus, design type of disfluency was coded as a categorical variable with between-subjects or within-subjects levels.
- (10) **Learning duration.** Learning duration was defined as the amount of time spent learning the material. Learning duration can be set as either long or short, and it should be carefully considered together with the effects of instructional design (Kalyuga 2012). We considered learning duration as a categorical variable, rather than a continuous variable, because some studies did not explicitly report the actual value. Thus, to guarantee enough number of effect sizes for moderator analyses, studies were coded and analyzed through dichotomizing the duration. According to earlier research (Reinwein 2012), learning duration was categorized as either “equal to or shorter than 10 minutes (10–)” or “longer than 10 minutes (10+).”
- (11) **Time interval between learning and test.** This variable was also considered as a categorical one and the rationale was the same as learning duration. Again, 10-min duration was defined as the point of the classification (10– or 10+).
- (12) **Use of distraction task.** After the learning phase, learners in a number of studies (e.g., Diemand-Yauman et al. 2011, Study 1; Eitel and Kühl 2016; Rummer et al. 2016) were asked to complete distraction or filler tasks that were completely unrelated to the contents of the learning materials. Other studies used no distraction tasks (e.g., Eitel et al. 2014; Kühl et al. 2014a; Sanchez and Khan 2016). Studies were coded according to whether there were distraction tasks or not (yes or no).
- (13) **Expectation of testing.** Studies were coded according to whether or not learners knew they would be tested (yes or no).

Calculation of Effect Sizes and Analysis

Data were analyzed using the Comprehensive Meta-Analysis (CMA) 2.0 software (<https://www.meta-analysis.com/>). Effect sizes were weighted using the reciprocal of their variances so that effect sizes based on studies with larger sample sizes were more heavily weighted in the analysis. The random-effects model was preliminarily used for all analyses because studies included in the meta-analysis differed on a number of variables (e.g., characteristics of participants, research design, and procedures), conforming to the assumption of the random-effects model that the true effect sizes are not exactly the same in all studies (Borenstein et al. 2009).

Because the data included in the present study were continuous data with no consistent unit, we chose Cohen's d as the standardized estimate of effect size (Cohen 1988). Specifically, Cohen's d was calculated as the mean score difference in recall, transfer, JOL, or learning time between a disfluent group and a fluent group. It should be pointed out that most studies tested the effect of disfluency on multiple dependent variables. A basic methodological assumption of meta-analysis is the independence of effects, and the inclusion of multiple dependent variables in each article would not conform to this assumption (Lipsey and Wilson 2001). Thus, to abide by this assumption and avoid potential deviation due to dependencies among effect sizes introduced by multiple variates per study, we separately conducted analyses with regard to different dependent variables (i.e., recall test, transfer test, JOL, and learning time). When a study reported multiple experiments or multiple conditions which were not related to the moderators, the data were merged to compute one pooled study-level effect size in order to minimize the deviation of results caused by a large number of effect sizes and disproportionate weight if not pooled (Borenstein et al. 2009). The generated study-level effect sizes were then averaged to obtain an overall average effect size point estimate for quantifying the central tendency among the effect sizes. A forest plot with 95% confidence interval (95% CI) was created to detect patterns in the magnitude of the individual effect sizes. For Cohen's d , the direction of the effect size was negative if recall, transfer, JOL, or learning time of the disfluent group was lower or shorter than that of the fluent group. The magnitude of an effect size was interpreted using Cohen's (1992) standards of small ($d = \pm 0.20$), moderate ($d = \pm 0.50$), and large ($d = \pm 0.80$).

The homogeneity statistic Q , along with its p value, was used for two purposes. First, Q was used to test whether there was significant variance within the set of effect sizes for each of the dependent measures. A related statistic, τ^2 , tests whether the between-study variance component is significantly different from 0. Another related statistic, I^2 , is the estimated percentage of total variance that is caused by true between-study heterogeneity rather than random error. According to Higgins et al. (2003), I^2 values of around 25, 50, and 75% are generally interpreted to indicate low, medium, and high heterogeneity, respectively. A high degree of heterogeneity indicates that the random-effects model used in the meta-analysis is reasonable and that there is a call for tests of moderation. Second, Q was used in tests of moderation to determine whether there was a significant difference between two sets of effect sizes (e.g., effect sizes from screen presentation vs. paper presentation). In this case, a statistically significant Q test indicates non-negligible between-study heterogeneity in effect size distributions beyond that accounted for by sampling error alone.

Studies have empirically confirmed the existence of publication bias (Franco et al. 2014; Rosenthal 1979). Publication bias is considered to emerge in meta-analyses if there are systematic errors between articles that ought to be included and those actually included

(Borenstein et al. 2009). This potential problem may lead to an overestimation of the mean effect size. In the present work, three strategies were adopted to examine the potential existence of publication bias. First, the funnel plot, a simple scatter plot with the standard error (the measure of study size) plotted on the vertical axis and the effect estimates plotted on the horizontal axis, was visually examined. A funnel plot shows a symmetrical distribution around the weighted mean effect and an asymmetric funnel plot suggests the existence of publication bias. Second, we performed Egger's linear regression test (Egger et al. 1997) to further assess funnel plot asymmetry. Through this test, a regression equation can be created with the standard normal deviate of each study as the dependent variable and the estimate's precision in each study as the independent variable. The intercept of the regression equation provides a measure of publication bias. The smaller its deviation from zero, the less pronounced the bias. Additionally, we conducted a rank correlation test (Begg and Mazumdar 1994) to evaluate the degree of asymmetry in the funnel plot. The rank correlation test quantifies the association between the effect sizes and their sampling variance. The less significant the correlation is, the more independent the effect sizes are from the sample sizes of the studies, and the less possible publication bias is.

Results

Descriptive Analysis

A total of 25 empirical articles that met the inclusion criteria were finally included and analyzed. An overview of the 25 articles with basic information and coded moderators is presented in Table 1. There were 21 studies obtained from journals, 2 from dissertations, 1 from an academic conference, and 1 from unpublished work. Across the 25 studies, 39 independent effect sizes with respect to recall were computed, involving 3135 participants. We were also able to compute 8 study-level effect sizes regarding transfer test containing 939 participants, 8 JOL-related effect sizes containing 901 participants, and 9 study-level effect sizes regarding learning time containing 689 participants.

The distribution of the derived effect sizes related to recall performance is presented in Fig. 1. The graph shows a roughly skewed distribution, with most of the effect sizes (74.4%) clustering between -0.60 and 0.20 . There were 22 out of 39 negative effect sizes (56.4%). Figure 2 presents forest plots with the point estimate of each effect size with a 95% confidence interval. No study-level effect size went beyond three *SD* of the mean of all effect sizes; thus, no outliers representing extreme values needed to be deleted.

Overall Analysis

Table 2 presents the results regarding the effects of disfluency on recall, transfer, JOL, and learning time. Concerning tests of recall, the meta-analysis revealed that the overall pooled effect size was not significant from 0 and was small in magnitude ($d = -0.01$, $p > 0.05$). Similarly, the overall effect size for transfer performance was small and not statistically significant ($d = 0.03$, $p > 0.05$). With regard to JOL scores, the overall effect size was significant and was considered to have small-to-medium magnitude with a negative direction ($d = -0.43$, $p < 0.001$). This result suggested that learners in the disfluent group predicted lower performance than those in the fluent group. The overall analysis regarding learning time was

Table 1 A list of studies included in the meta-analysis

Study	Sample size	Experiment(s) or condition(s)	Prior knowledge level	Learning material domain	Pacing of presentation	Type of presentation	Modality of presentation	Medium of presentation
Ball et al. (2014)	39 93	Exp. 1a (timed condition) Exp. 1a (unlimited condition), exp. 1b, exp. 2	n/a n/a	Social science Social science	System-paced Self-paced	Static Static	Single Single	Screen Screen
Carpenter et al. (2016)	172 91	Exp. 1, exp. 2 Exp. 3	None None	Mathematics Mathematics	System-paced System-paced	Dynamic Dynamic	Dual Dual	Screen Screen
Carpenter et al. (2013)	42 70	Exp. 1 Exp. 2	None None	Science Science	System-paced System-paced and self-paced	Static Static	Single Dual	Screen Screen
Diemand-Yauman et al. (2011)	27 222	Exp. 1 Exp. 2	None n/a	Science Language, physics, history, and chemistry	System-paced n/a	Static n/a	Single n/a	Paper Screen and paper Screen
Eitel and Köhl (2016)	47 50	Low test expectancy condition High test expectancy condition	Low Low	Technology/engineering Technology/engineering	Self-paced Self-paced	Static Static	Single Single	Screen Screen
Eitel et al. (2014)	218	Exp. 1, exp. 2 (text and pictures condition), exp. 3 Exp. 4	n/a	Technology/engineering	System-paced	Static	Single	Paper
Faber et al. (2017)	43 47	Exp. 2 (text only condition)	n/a n/a	Technology/engineering Science	System-paced System-paced	Static Static	Single Single	Paper Paper
French (n.d.)	204	/	None	Technology/engineering	Self-paced	Static	Single	Screen
French et al. (2013)	20 275	/ /	n/a none	n/a Science	n/a System-paced	n/a Static	n/a Single	Paper Screen
Guenther (2012)	104 47	Exp. 1 Exp. 2	n/a n/a	Science and mathematics Science	System-paced System-paced	Static Static	Single Single	Paper Paper
Haysom (2012)	56	Exp. 1	None	Science	System-paced	Static	Single	Screen
Katzir et al. (2013)	90	Exp. 1, exp. 2	n/a	n/a	Self-paced	Static	Single	Paper
Kühl et al. (2014a)	41	System-paced condition	n/a	Science	System-paced	Static	Dual	Screen
Lee (2013)	38	Self-paced condition	n/a	Science	Self-paced	Static	Dual	Screen
Lehmann et al. (2016)	22 47	/ /	n/a Medium	Social science Science	n/a System-paced	Static Static	Single Single	Screen Paper

Table 1 (continued)

Study	Sample size	Experiment(s) or condition(s)	Prior knowledge level	Learning material domain	Pacing of presentation	Type of presentation	Modality of presentation	Medium of presentation
Miele and Molden (2010)	114	Exp. 3	n/a	Social science	Self-paced	Static	Single	Screen
Pieger et al. (2016)	83	/	None	Social science	Self-paced	Static	Single	Screen
Pieger et al. (2017)	75	/	None	Social science	Self-paced	Static	Single	Screen
Rummer et al. (2016)	63	Exp. 1	None	Science	Self-paced	Static	Single	Screen
	58	Exp. 2	None	Science	System-paced	Static	Single	Paper
	80	Exp. 3	None	Science	System-paced	Static	Single	Paper
Sanchez and Khan (2016)	101	/	None	Technology/engineering	Self-paced	Dynamic	Dual	Screen
Seufert et al. (2017)	68	Exp. 1	None	Social science	Self-paced	Static	Single	Paper
	73	Exp. 2	Medium	Science	Self-paced	Static	Single	Paper
Strukelj et al. (2016)	53	/	Low	Technology/engineering	Self-paced	Static	Single	Screen
Weisserger and Reinhard (2017)	91	Session 1 (hard-font vs. control)	n/a	Science	System-paced	Static	Single	Paper
	89	Session 2 (hard-font vs. control)	n/a	Science	System-paced	Static	Single	Paper
Weltman and Eakin (2014)	101	Unusual fonts vs. control	n/a	Mathematics	Self-paced	Static	Single	Screen
Whitehouse (2011)	40	Exp. 1 (Haettenschweiler vs. Corbel)	None	Science	System-paced	Static	Single	Screen
	30	Exp. 2 (Haettenschweiler vs. Corbel)	n/a	Science	System-paced	Static	Single	Screen

Study	Inclusion of images	Fluency manipulation type	Design of study	Learning duration	Time interval between learning and test	Use of distraction task	Expectation of testing	Dependent variable(s)
Ball et al. (2014)	No	Font-related	Within-subjects	10–	10–	Yes	Yes	Recall, JOL
	No	Font-related	Within-subjects	10–	10–	Yes	Yes	Recall, JOL, learning time
Carpenter et al. (2016)	Yes	Audio-related	Between-subjects	10+	10–	No	Yes	Recall, JOL
	Yes	Audio-related	Between-subjects	10+	10+	No	Yes	Recall, JOL
Carpenter et al. (2013)	No	Audio-related	Between-subjects	10–	10–	Yes	Yes	Recall, JOL
	No	Audio-related	Between-subjects	10–	10–	Yes	Yes	Recall, learning time

Table 1 (continued)

Study	Inclusion of images	Fluency manipulation type	Design of study	Learning duration	Time interval between learning and test	Use of distraction task	Expectation of testing	Dependent variable(s)
Diemand-Yauman et al. (2011)	No	Font-related	Between-subjects	10–	10+	Yes	n/a	Recall
	n/a	Font-related	Between-subjects	10+	n/a	n/a	Yes	Recall
Eitel and Kühn (2016)	Yes	Font-related	Between-subjects	10–	10–	Yes	No	Recall, transfer; learning time
	Yes	Font-related	Between-subjects	10–	10–	Yes	Yes	Recall, transfer; learning time
Eitel et al. (2014)	Yes	Font-related	Between-subjects	10–	10–	No	Yes	Recall, transfer
	No	Font-related	Between-subjects	10–	10–	No	Yes	Recall, transfer
Faber et al. (2017)	Yes	Font-related	Between-subjects	10–	10–	No	Yes	Recall, transfer
	No	Font-related	Between-subjects	10–	10–	No	n/a	Recall, transfer
French (n.d.)	n/a	Font-related	Within-subjects	n/a	10+	n/a	n/a	Recall
French et al. (2013)	No	Font-related	Between-subjects	10–	10+	Yes	No	Recall
Guenther (2012)	No	Font-related	Between-subjects	10–	10–	No	Yes	Recall
	No	Font-related	Between-subjects	10–	10–	No	Yes	Recall
Haysom (2012)	No	Font-related	Between-subjects	10–	10–	Yes	Yes	Recall
Katzir et al. (2013)	No	Font-related	Between-subjects	10–	10+	Yes	n/a	Recall
Kühl et al. (2014a)	No	Font-related	Within-subjects	n/a	n/a	n/a	Yes	Recall
	Yes	Audio-related	Between-subjects	10–	10–	No	Yes	Recall, transfer
	Yes	Audio-related	Between-subjects	10–	10–	No	Yes	Recall, transfer; learning time
Lee (2013)	No	Font-related	Within-subjects	n/a	n/a	Yes	n/a	Recall
Lehmann et al. (2016)	No	Font-related	Between-subjects	n/a	10–	No	n/a	Recall, transfer
Miele and Molden (2010)	No	Font-related	Between-subjects	10–	10–	No	Yes	Recall, JOL; learning time
Pieger et al. (2016)	No	Font-related	Between-subjects	10+	n/a	No	n/a	Recall, JOL; learning time
Pieger et al. (2017)	No	Font-related	Within-subjects	10+	10–	No	n/a	Recall, JOL; learning time
Rummer et al. (2016)	No	Font-related	Within-subjects	10–	10+	Yes	n/a	Recall, learning time
	No	Font-related	Within-subjects	10–	10+	Yes	n/a	Recall
	No	Font-related	Between-subjects	10–	10+	Yes	n/a	Recall
Sanchez and Khan (2016)	Yes	Audio-related	Between-subjects	10–	10–	No	n/a	Recall
Seufert et al. (2017)	No	Font-related	Between-subjects	10–	10–	No	n/a	Recall
	No	Font-related	Between-subjects	10+	10–	No	n/a	Recall, transfer; learning time
Strukelj et al. (2016)	No	Font-related	Between-subjects	10–	10+	Yes	Yes	Recall, learning time
	No	Font-related	Between-subjects	10+	10–	Yes	Yes	Recall, JOL

Table 1 (continued)

Study	Inclusion of images	Fluency manipulation type	Design of study	Learning duration	Time interval between learning and test	Use of distraction task	Expectation of testing	Dependent variable(s)
Weissergerber and Reinhard (2017)	No	Font-related	Between-subjects	10+	10+	n/a	No	Recall
Weltman and Eakin (2014)	No	Font-related	Between-subjects	n/a	10+	Yes	n/a	Recall, transfer, JOL
Whitehouse (2011)	No	Font-related	Between-subjects	10–	10+	Yes	n/a	Recall
	No	Font-related	Within-subjects	10+	10+	Yes	Yes	Recall, transfer

“Self-paced” means learners were allowed to interactively control the presentation by themselves (e.g., start, play, pause, replay, and stop), whereas “system-paced” means learners had no option to interact with the presentation, and accordingly, “system-paced and self-paced” means the presentation in this experiment was first system-paced, and then self-paced; “static” means the elements in a presentation remained still, whereas “dynamic” means the elements were mobile; “single” means the learning material was presented only in a visual form, whereas “dual” means the material was presented in both visual and auditory formats. 10+ longer than 10 min, 10– equal to or shorter than 10 min, JOL judgment of learning, / only one experiment in the study, n/a not applicable

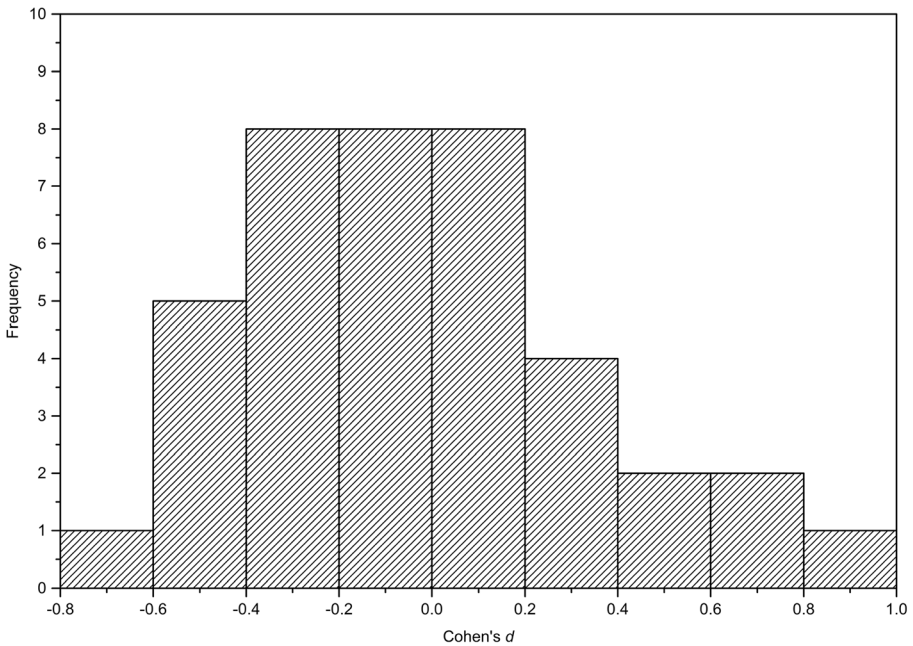


Fig. 1 Distribution of 39 effect sizes related to recall performance

limited to studies that used self-pacing (i.e., studies in which participants were allowed to work for as long as they wished, in contrast to studies that put limits on learning time). In this analysis, the overall effect size was statistically significant and medium in magnitude with a positive direction ($d = 0.52$, $p < 0.001$), indicating that (in a self-paced environment) learners spent longer time learning disfluent materials than learning fluent materials.

As shown in Table 2, the homogeneity test showed that effect sizes varied significantly across studies on each of the four dependent variables ($p_s < 0.05$), with a medium-to-high heterogeneity due to variance across studies (all $I^2 > 50$). These results warranted tests of moderation to identify sources of this heterogeneity. Only a small number of effect sizes were able to be computed for the dependent measures of transfer, JOL, and learning time (each being less than 10), possibly affecting the reliability of subgroup (moderation) analyses on these dependent variables. Therefore, the moderator analyses were conducted just on recall test scores (with 39 effect sizes to examine).

Moderator Analyses

The results are presented in Table 3. With respect to participant characteristic, we did not find a significant moderating effect regarding learners' prior knowledge; the difference between the disfluent group and the fluent group was not significant at any level of prior knowledge, and the three pooled effect sizes were not significantly different from each other ($p_s > 0.05$).

Regarding learning material characteristics as well as experimental design characteristics, no variable was found to be a moderator of the disfluency effect on recall performance (all $Q_B < 3.00$, $p_s > 0.05$). Furthermore, no subgroup showed a study-level effect size significantly different from 0 (see Table 3).

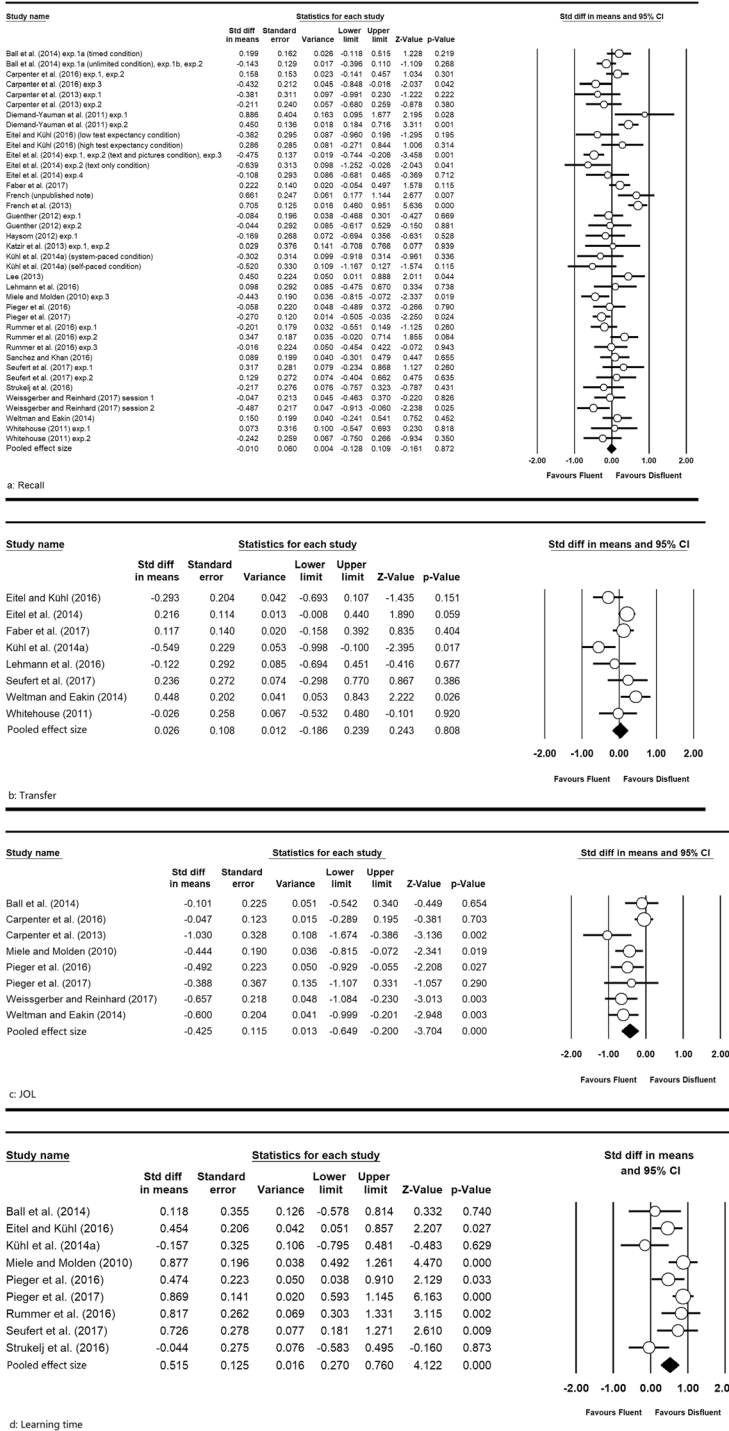


Fig. 2 Forest plots of effect sizes for **a** recall, **b** transfer, **c** JOL, **d** learning time. *Note.* the order of the effect sizes for recall is the same as in Table 1

Table 2 Main effects of disfluency on recall, transfer, JOL, and learning time, as well as results of the homogeneity test

Variables	<i>N</i>	<i>k</i>	Effect size			Homogeneity test			
			Cohen's <i>d</i>	95% CI	<i>z</i>	<i>Q</i>	<i>p</i>	<i>I</i> ²	tau ²
Recall	3135	39	−0.01	[−0.13, 0.11]	−0.16	118.49	<0.001	67.93	0.09
Transfer	939	8	0.03	[−0.19, 0.24]	0.24	16.65	<0.05	57.96	0.05
JOL	901	8	−0.43***	[−0.65, −0.20]	−3.70	15.70	<0.05	55.41	0.06
Learning time	689	9	0.52***	[0.27, 0.76]	4.12	20.09	<0.01	60.18	0.08

N total number of participants, *k* number of effect sizes, *CI* confidence interval, ****p* < 0.001

Further Moderator Analysis for Studies with Adults

The first moderation analyses showed that among participant characteristic, learning materials, and experimental designs, no significant moderator of the disfluency effect on recall test scores was found. However, most of the effect sizes were from adults (87.2%). Thus, next, a moderator analysis was conducted just for studies with adults.

The results are presented in Table 4. Again, no individual variable was a significant moderator of the disfluency effect. However, certain trends are of interest. These trends are apparent when the effect size of disfluency is significant at one level of the moderator but not the other level(s), or when the effect size is significant at both levels of the moderator. These trends are presented in italic in Table 4. First, the trends suggest that certain characteristic of the learning materials may affect the effect of disfluency. Specifically, learners in the disfluent group recalled less than learners in the fluent group when the presentation was system-paced. Second, the trends suggest that certain characteristics of the experimental design might influence the effect of disfluency. Learners in the disfluent group showed worse recall than learners in the fluent group when a between-subjects design was used; when the time interval between learning and test was ≤ 10 min; and when no distraction tasks were presented.

In summary, again no factor was found to be a moderator of the disfluency effect on recall performance for studies with adults. It should be pointed out that although there were not significant effects due to some potential moderators (i.e., pacing of presentation, design of study, time interval between learning and test, and distraction tasks), the pattern of results suggest that these factors should not be ignored in future research on the disfluency effect on adults' recall.

Publication Bias Analysis

We first examined the funnel plot (see Fig. 3) to determine if there appeared to be publication bias. The scatterplot was nearly symmetrical, indicating that no potential bias existed. Given the subjectivity in interpreting a funnel plot, we used Egger's linear regression test (Egger et al. 1997) and a rank correlation test (Begg and Mazumdar 1994) to further assess funnel plot asymmetry. Egger's linear regression test showed that publication bias was an unlikely influence on the findings of the present meta-analysis (intercept = −0.83, *p* > 0.05). This result was confirmed by the rank correlation test (tau = −0.06, *p* > 0.05).

Table 3 Moderator analysis on recall performance for all studies

Variables	<i>N</i>	<i>k</i>	Effect size			Homogeneity test between subgroups	
			Cohen's <i>d</i>	95% CI	<i>z</i>	<i>Q_B</i>	<i>p</i>
Prior knowledge level						0.68	> 0.05
None	1505	16	0.06	[-0.13, 0.25]	0.61		
Low	150	3	-0.10	[-0.49, 0.29]	-0.51		
Medium	120	2	0.11	[-0.28, 0.50]	0.58		
Learning material domain						0.52	> 0.05
Science	1125	18	-0.02	[-0.22, 0.18]	-0.21		
Technology/engineering	716	7	-0.14	[-0.43, 0.15]	-0.96		
Mathematics	364	3	-0.02	[-0.38, 0.34]	-0.13		
Social science	494	7	-0.03	[-0.25, 0.19]	-0.26		
Pacing of presentation						0.06	> 0.05
Self-paced	1253	15	-0.07	[-0.20, 0.06]	-1.13		
System-paced	1548	20	-0.05	[-0.23, 0.14]	-0.47		
Type of presentation						0.01	> 0.05
Static	2529	34	-0.04	[-0.17, 0.08]	-0.68		
Dynamic	364	3	-0.04	[-0.39, 0.31]	-0.24		
Modality of presentation						0.76	> 0.05
Single	2380	31	-0.02	[-0.15, 0.11]	-0.31		
Dual	513	6	-0.14	[-0.38, 0.10]	-1.17		
Medium of presentation						0.17	> 0.05
Screen	1900	23	-0.04	[-0.19, 0.10]	-0.58		
Paper	1013	15	0.01	[-0.20, 0.21]	-0.08		
Inclusion of images						1.80	> 0.05
Yes	805	9	-0.18	[-0.39, 0.04]	-1.59		
No	2088	28	-0.001	[-0.14, 0.13]	-0.02		
Fluency manipulation type						2.08	> 0.05
Font-related	2580	32	0.03	[-0.11, 0.16]	0.40		
Audio-related	555	7	-0.16	[-0.38, 0.06]	-1.45		
Design of study						0.69	> 0.05
Between-subjects	2645	30	-0.04	[-0.18, 0.10]	-0.54		
Within-subjects	490	9	0.07	[-0.15, 0.29]	0.64		
Learning duration						0.06	> 0.05
10+	837	9	-0.08	[-0.30, 0.15]	-0.66		
10-	2018	25	-0.04	[-0.20, 0.12]	-0.51		
Time interval between learning and test						1.44	> 0.05
10+	983	13	0.07	[-0.19, 0.33]	0.54		
10-	1824	22	-0.10	[-0.22, 0.02]	-1.69		
Use of distraction task						2.99	> 0.05
Yes	1237	18	0.07	[-0.11, 0.24]	0.74		
No	1566	17	-0.13	[-0.27, 0.01]	-1.82		
Expectation of testing						0.05	> 0.05
Yes	1695	20	-0.13	[-0.28, 0.01]	-1.79		
No	411	3	-0.03	[-0.92, 0.85]	-0.08		

Discussion

The present meta-analysis of 25 empirical studies, involving 3135 participants, theoretically examined the effects of experimentally manipulated perceptual disfluency on learners' recall, transfer, JOL, and learning time in a text-based educational context. In addition, this meta-analysis tentatively tested three sets of study characteristics (i.e., participant, learning material, and experimental design characteristics) to check whether they might moderate the impact of disfluency on the most frequently studied outcome, namely recall.

Table 4 Moderator analysis on recall performance for studies with adults

Variables	<i>N</i>	<i>k</i>	Effect size			Homogeneity test between subgroups	
			Cohen's <i>d</i>	95% CI	<i>z</i>	<i>Q_B</i>	<i>p</i>
Prior knowledge level						0.64	> 0.05
None	1026	14	-0.03	[-0.18, 0.12]	-0.38		
Low	150	3	-0.10	[-0.49, 0.29]	-0.51		
Medium	120	2	0.11	[-0.28, 0.50]	0.58		
Learning material domain						1.33	> 0.05
Science	850	17	-0.09	[-0.22, 0.05]	-1.21		
Technology/engineering	512	6	-0.23	[-0.51, 0.06]	-1.58		
Mathematics	364	3	-0.02	[-0.38, 0.34]	-0.13		
Social science	494	7	-0.03	[-0.25, 0.19]	-0.26		
Pacing of presentation						0.11	> 0.05
Self-paced	959	13	-0.09	[-0.25, 0.07]	-1.18		
System-paced	1273	19	-0.13*	[-0.25, -0.01]	-2.11		
Type of presentation						0.09	> 0.05
Static	1960	31	-0.10	[-0.20, 0.01]	-1.79		
Dynamic	364	3	-0.04	[-0.39, 0.31]	-0.24		
Modality of presentation						0.23	> 0.05
Single	1811	28	-0.08	[-0.19, 0.04]	-1.35		
Dual	513	6	-0.14	[-0.38, 0.10]	-1.17		
Medium of presentation						0.22	> 0.05
Screen	1421	21	-0.10	[-0.21, 0.01]	-1.81		
Paper	903	13	-0.05	[-0.25, 0.15]	-0.46		
Inclusion of images						0.87	> 0.05
Yes	805	9	-0.18	[-0.39, 0.04]	-1.59		
No	1519	25	-0.06	[-0.17, 0.05]	-1.05		
Fluency manipulation type						0.53	> 0.05
Font-related	1769	27	-0.07	[-0.18, 0.05]	-1.19		
Audio-related	555	7	-0.16	[-0.38, 0.06]	-1.45		
Design of study						1.08	> 0.05
Between-subjects	1944	27	-0.12*	[-0.24, -0.01]	-2.14		
Within-subjects	380	7	0.004	[-0.21, 0.22]	0.04		
Learning duration						0.23	> 0.05
10+	615	8	-0.16	[-0.33, 0.01]	-1.81		
10-	1539	23	-0.11	[-0.23, 0.02]	-1.63		
Time interval between learning and test						0.26	> 0.05
10+	688	11	-0.07	[-0.27, 0.13]	-0.67		
10-	1620	21	-0.13*	[-0.24, -0.01]	-2.14		
Use of distraction task						3.41	> 0.05
Yes	962	17	0.01	[-0.12, 0.14]	0.17		
No	1362	16	-0.17*	[-0.31, -0.03]	-2.40		
Expectation of testing						2.11	> 0.05
Yes	1383	18	-0.18**	[-0.30, -0.06]	-2.82		
No	136	2	-0.45*	[-0.79, -0.11]	-2.57		

***p* < 0.01; **p* < 0.05

Null Effects of Perceptual Disfluency on Text-Based Learning Performance

The most important theory-based test in the present study addressed whether perceptual disfluency could facilitate or impede recall and transfer performance when learning with texts. According to disfluency theory (Alter et al. 2007), a disfluent text should introduce desirable

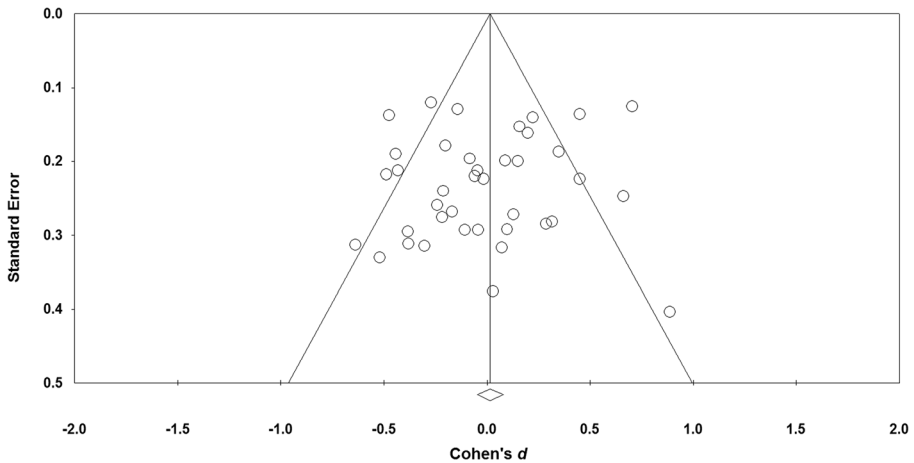


Fig. 3 Funnel plot of standard error by Cohen's d effect sizes

difficulty and lead to deeper processing of the instructional material in system 2, leading to better recall and transfer. By contrast, CLT (Sweller et al. 2011; Sweller et al. 1998) predicts that disfluent text would increase ECL, leading to worse recall and transfer.

However, contrary to Hypotheses 1a and 1b, the results showed that students who learned with a disfluent text recalled ($d = -0.01$) and transferred ($d = 0.03$) to the same degree as students learning with a fluent text. Thus, neither disfluency theory nor CLT was supported. These results are consistent with other empirical studies using texts as learning materials that have failed to find significant differences between disfluent and fluent groups on recall and transfer tests (for overviews, see Köhl and Eitel 2016; Weissgerber and Reinhard 2017; Xie et al. 2016).

Given the null effects of perceptual disfluency on recall and transfer in the present meta-analysis, one might argue that it is text comprehension that counts. Compared with word learning, learning with texts requires more cognitive resources to select and build connections between relevant information for text comprehension (Mayer 1984). Thus, even if disfluency is able to activate analytic processing as predicted by disfluency theory, its influence might not be strong enough to invoke the cognitive resources needed for deep text learning to occur. From the perspective of CLT, one might cautiously argue that perceptual disfluency increases ECL during text-based learning, but the ECL does not necessarily result in overload of total working memory capacity. Learners, for example, have the chance to compensate for the negative effect of ECL through monitoring or control of the learning process (e.g., through increasing learning time). Thus, even if disfluency is able to increase ECL as predicted by CLT, the total cognitive load may be still well within the limits of working memory (Sweller et al. 1998).

One important consideration is the boundary line of disfluency (Diemand-Yauman et al. 2011; Seufert et al. 2017). On a descriptive level, Seufert et al. (2017) found a reversed u-shape pattern when investigating the relationship between increasing disfluency levels and learning outcomes. Their results identified an optimal level of perceptual disfluency, that is, the level of perceptual disfluency associated with optimal learning performance on the study task. Therefore, it is vital to determine the exact point at which a text begins to be disfluent or fluent, as well as the exact point at which disfluency begins to (significantly) improve or hinder learning

outcome. Because very few studies in our meta-analysis took this into account, we are not sure whether the disfluency manipulation was within an optimal scope. A poor manipulation of disfluency might narrow the gap between the learning performance in conditions of disfluent text and fluent text, leading to a null effect.

After failing to find a disfluent font benefit for math problem reasoning, Meyer et al. (2015, e20) speculated that “disfluent fonts may aid memory but not reasoning—presumably because reading words more slowly benefits memory, but not reasoning.” However, the current meta-analysis did not confirm a disfluent text benefit for recall, and the exploratory moderator analyses did not discover any factor that might moderate the impact of disfluency on recall. In short, our meta-analysis is one more failed attempt to replicate the disfluency effect in the text-based learning domain.

With respect to the moderator analyses, although no significant moderators were found in adult samples, the pattern of results suggested that the negative effect of disfluency might be especially apparent under some circumstances (i.e., system-paced presentation, between-subjects design, shorter time interval between learning and test, without distraction task included). It should be pointed out that because there were very few studies represented in some subgroups (e.g., design of study), the results must be treated with some caution.

Perceptual Disfluency Influences JOL and Learning Time

Another theory-based test in the present study addressed whether students would use perceptual disfluency as a metacognitive cue for monitoring and controlling their learning with texts. According to models of dual-system processing, making a text disfluent would reduce learners’ judgments of learning and increase learning time.

Our findings revealed that perceptual disfluency reduced the judgment magnitudes of learning. Specifically, students predicted that they would perform worse after learning disfluent instructional materials than after learning regularly fluent materials ($d = -0.43$), which is in line with Hypothesis 2 and previous empirical studies (Ball et al. 2014; Carpenter et al. 2013; Pieger et al. 2017; Weltman and Eakin 2014). Meanwhile, we found that making instructional materials harder to read influenced the allocation of learning time when the learning process could be controlled by students themselves. Specifically, students spent longer time to learn disfluent information than to learn fluent information ($d = 0.52$), which is in line with Hypothesis 3 and prior qualitative reviews (Dunlosky and Ariel 2011; Son and Metcalfe 2000) as well as empirical studies (e.g., Ball et al. 2014; Eitel and Kühl 2016; Miele and Molden 2010, Experiment 3).

Compared with learning with simple words, learning with instructional materials is more complex (Pieger et al. 2017). Regulating the complex learning process requires monitoring, and control and typically, a variety of cues (Koriat 1997, 2012; Koriat et al. 2006) including metacognitive cues (Koriat 1997; Pieger et al. 2016). When the studied information is perceived as difficult, poor learning performance is likely to be predicted or assessed. This monitoring process (i.e., low JOL) will affect students’ control during learning, for example, by reading more slowly and using more time to study.

These results should also be interpreted in the context of the results on all four dependent variables (i.e., recall, transfer, JOL, and learning time). Based on the entire set of findings, one might conclude that a perceptually disfluent text may affect learning processes (monitoring or control), rather than learning outcomes (recall or transfer). It is possible that perceptual disfluency can function as a cue to justify the learning processes, rather than overturn the

final learning performance. That would explain both the lower JOLs, longer learning time, and the null effects on recall and transfer. However, we cannot overstate the importance of these results because several studies included in the meta-analysis had low sample sizes ($N < 30$, Diemand-Yauman et al. 2011, Study 1; French n.d.; Lee 2013), and as noted earlier, only a small number of effect sizes were computed for the dependent measures of transfer, JOL, and learning time.

Limitations and Future Directions

The current meta-analysis provides a reference for research on text disfluency and suggests that it is too early to widely introduce disfluency-based educational interventions to instructional design. The following limitations should be acknowledged. First, there are two issues related to the dependent measures. One of the four dependent measures we analyzed was the magnitudes of judgments of learning, but the impact of disfluency on judgment accuracy is not very clear. On the one hand, lower magnitudes of judgments of learning in the disfluency group may make learners become more cautious and rational (Carpenter et al. 2013); on the other hand, disfluency may lead to significant under-confidence (Pieger et al. 2016). The other dependent variable that needs further research attention is learning time. If a disfluent text produces longer learning time when compared with a fluent text under time-unlimited circumstances, then the ingredients of “learning time” are ambiguous. It is possible that this “learning time” contains the time for real learning and the time for deciphering the disfluent words.

Second, only a small number of studies used transfer, JOL, or learning time as a dependent variable. Therefore, the tests of moderation were limited to studies using the more commonly used dependent measure, namely recall. It was also not possible to test age as a moderator because almost all studies on text disfluency are conducted with adults. However, other research suggests that there might be a developmental change in the disfluency effect (e.g., Katzir et al. 2013). That is, the positive or negative influence of disfluency might emerge in certain stages of learners’ lives, but disappear or even go in the opposite direction during another period. Of course, this possibility needs more evidence. In addition, to ensure enough number of effect sizes for moderator analyses in the present study, both learning duration and time interval between learning and test were treated as categorical variables rather than continuous ones, which might influence the reliability of these corresponding results.

Finally, there are several questions that can be addressed in future studies. For example, learning materials, rather than test materials, were usually manipulated to be disfluent or fluent in previous text-based disfluency studies. Perhaps disfluency manipulation of test items would show additional unexpected but vital results. There could also be boundary conditions that deserve to be further examined, such as learning motivation and working memory capacity. In addition, closer examination of the specific nature of the disfluency manipulation (e.g., 8- vs. 20-point font in one study; 12- vs. 24-point font in another) may generate information about when exactly text becomes disfluent. As another example, because our focus was perceptual disfluency, we included studies where disfluency was manipulated on a perceptual level but not on a conceptual level (e.g., by altering text coherence; McNamara et al. 1996), and another interesting direction for future research would be to investigate the different effects of perceptual disfluency and conceptual disfluency on learning with texts.

Conclusions

This meta-analysis on the effect of disfluent vs. fluent texts on participants' recall, transfer, JOL, and learning time indicates that (a) perceptual disfluency can reduce judgments of learning and increase learning time, but appears to have no effect on recall or transfer, providing insufficient evidence that it either stimulates analytic processing or increases extraneous cognitive load; (b) characteristics of participants, learning materials, and experimental design do not appear to moderate the effect of disfluency on recall, but the pattern of results in samples of adults suggests that several factors should be examined in future research. This study has implications for future research as well as a warning function for interventions in educational settings.

Acknowledgements The authors would like to thank the researchers who provided details of their unpublished work.

Funding This study was funded by the Major Program of National Social Science Foundation of China (grant number 11&ZD151) and Fundamental Research Funds of Central China Normal University (grant number CCNU14Z02004).

Compliance with Ethical Standards This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235.
- 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*(4), 569–576.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*(4), 774–783.
- Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.
- Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied*, *20*(4), 336–348 This study is included in the meta-analysis.
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101.
- Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, *41*(6), 897–903.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 185–205). Cambridge: MIT Press.
- 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.
- Bjork, R. A., & Yue, C. L. (2016). Commentary: is disfluency desirable? *Metacognition and Learning*, *11*(1), 133–137.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

- 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*, 20(6), 1350–1356 This study is included in the meta-analysis.
- Carpenter, S. K., Mickes, L., Rahman, S., & Fernandez, C. (2016). The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. *Journal of Experimental Psychology: Applied*, 22(2), 161–172 This study is included in the meta-analysis.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- De Bruin, A. B., & Van Gog, T. (2012). Improving self-monitoring and self-regulation: from cognitive psychology to the classroom. *Learning and Instruction*, 22(4), 245–252.
- 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 This study is included in the meta-analysis.
- Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. Ross (Ed.), *The psychology of learning and motivation: advances in research and theory* (pp. 103–140). San Diego: Academic Press.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Los Angeles: SAGE.
- Dunlosky, J., & Mueller, M. L. (2016). Recommendations for exploring the disfluency hypothesis for establishing whether perceptually degrading materials impacts performance. *Metacognition and Learning*, 11(1), 123–131.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–634.
- Eitel, A., & Kühl, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, 11(1), 107–121 This study is included in the meta-analysis.
- 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 This study is included in the meta-analysis.
- Faber, M., Mills, C., Kopp, K., & D'Mello, S. (2017). The effect of disfluency on mind wandering during text comprehension. *Psychonomic Bulletin & Review*, 24(3), 914–919 This study is included in the meta-analysis.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: unlocking the file drawer. *Science*, 345(6203), 1502–1505.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- French, M. M. J. (n.d.). Changing fonts in education: the time dependence of the benefits. Retrieved from <http://matthewfrench.net/pubs/Changing%20Fonts%20in%20Education%20Time%20Dependence.pdf>. This study is included in the meta-analysis.
- French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, A., Heritage, J., Poland, R. L., Reece, S., & Tabor, J. (2013). Changing fonts in education: how the benefits vary with ability and dyslexia. *The Journal of Educational Research*, 106(4), 301–304 This study is included in the meta-analysis.
- Guenther, R. K. (2012). Does the processing fluency of a syllabus affect the forecasted grade and course difficulty? *Psychological Reports*, 110(3), 946–954 This study is included in the meta-analysis.
- Haysom, H. J. (2012). *Could Comic Sans make you smarter? An exploration of the effects of disfluency on learning outcomes (Honours thesis)*. Brisbane: University of Queensland This study is included in the meta-analysis.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557–560.
- Hirshman, E. L., & Bjork, R. A. (1988). The generation effect: support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 14(3), 484–494.
- James, W. (1950). *The principles of psychology*. New York: Dover (Original work published 1890).
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4), 509–539.
- Kalyuga, S. (2012). Instructional benefits of spoken words: a review of cognitive load factors. *Educational Research Review*, 7(2), 145–159.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.
- Katzir, T., Hershko, S., & Halamish, V. (2013). The effect of font size on reading comprehension on second and fifth grade children: bigger is not always better. *PLoS One*, 8(9), e74061 This study is included in the meta-analysis.

- Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370.
- Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction*, *22*(4), 296–298.
- Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69.
- Kühl, T., & Eitel, A. (2016). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning*, *11*(1), 1–13.
- Kühl, T., Eitel, A., Dammik, G., & Körndle, H. (2014a). The impact of disfluency, pacing, and students' need for cognition on learning with multimedia. *Computers in Human Behavior*, *35*, 189–198 This study is included in the meta-analysis.
- Kühl, T., Eitel, A., Scheiter, K., & Gerjets, P. (2014b). A call for an unbiased search for moderators in disfluency research: reply to Oppenheimer and Alter (2014). *Applied Cognitive Psychology*, *28*(5), 805–806.
- Lee, M., H. (2013). *Effects of disfluent Kanji fonts on reading retention with e-book*. Paper presented at the 2013 IEEE 13th International Conference on Advanced Learning Technologies (pp. 481–482). Beijing, China: IEEE. This study is included in the meta-analysis.
- Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: an aptitude-treatment-interaction study. *Metacognition and Learning*, *11*(1), 89–105 This study is included in the meta-analysis.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Magrehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning*, *11*(1), 35–56.
- Mayer, R. E. (1984). Aids to text comprehension. *Educational Psychologist*, *19*(1), 30–42.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179.
- Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., et al. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16–e30.
- Miele, D. B., & Molden, D. C. (2010). Naive theories of intelligence and the role of processing fluency in perceived comprehension. *Journal of Experimental Psychology: General*, *139*(3), 535–557 This study is included in the meta-analysis.
- Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12.
- Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology*, *28*(4), 502–504.
- Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: different ways to increase germane cognitive load. *Learning and Instruction*, *16*(2), 87–91.
- Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency—does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*, *44*, 31–40 This study is included in the meta-analysis.
- Pieger, E., Mengelkamp, C., & Bannert, M. (2017). Fostering analytic metacognitive processes and reducing overconfidence by disfluency: the role of contrast effects. *Applied Cognitive Psychology*, *31*(3), 291–301 This study is included in the meta-analysis.
- Reber, R., & Greifeneder, R. (2017). Processing fluency in education: how metacognitive feelings shape learning, belief formation, and affect. *Educational Psychologist*, *52*(2), 84–103.
- Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring “fringe” consciousness: the subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, *13*(1), 47–60.
- Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research*, *41*(1), 1–32.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.
- Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: null-effects of disfluency on learning outcomes. *Metacognition and Learning*, *11*(1), 57–70 This study is included in the meta-analysis.

- Sanchez, C. A., & Khan, S. (2016). Instructor accents in online education and their effect on learning and attitudes. *Journal of Computer Assisted Learning, 32*(5), 494–502 This study is included in the meta-analysis.
- Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science, 45*(2), 221–238 This study is included in the meta-analysis.
- Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority—does the medium, screen versus paper, affect performance even with brief tasks? *Metacognition and Learning, 11*(1), 15–33.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning Memory, and Cognition, 26*(1), 204–221.
- Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: evidence from eye movements. *Metacognition and Learning, 11*(1), 71–88 This study is included in the meta-analysis.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Thompson, V. A., Prowse Turner, J. A., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., et al. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237–251.
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247–262.
- Weissgerber, S. C., & Reinhard, M. A. (2017). Is disfluency desirable for learning? *Learning and Instruction, 49*, 199–217 This study is included in the meta-analysis.
- Weltman, D., & Eakin, M. (2014). Incorporating unusual fonts and planned mistakes in study materials to increase business student focus and retention. *INFORMS Transactions on Education, 15*(1), 156–165 This study is included in the meta-analysis.
- Whitehouse, E. (2011). *An investigation into disfluency effects: depth of learning and affective outcomes* (Honours thesis). Brisbane: University of Queensland. This study is included in the meta-analysis.
- deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: implications for making a better reader. *Memory & Cognition, 32*(6), 945–955.
- Xie, H., Wang, F., Wang, Y., & An, J. (2016). Does harder-to-read mean better-to-learn? Effects of disfluency on learning outcomes. *Advances in Psychological Science, 24*(7), 1077–1090.
- 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*(2), 229–241.