frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 08 January 2026
DOI 10.3389/feduc.2025.1731080

OPEN ACCESS

EDITED BY
Juan Cristobal Castro-Alonso,
University of Birmingham, United Kingdom

REVIEWED BY
Anna Gorbunova,
National Research University Higher School of Economics, Russia
Elina Tsigeman,
National Research University Higher School of Economics, Russia

*CORRESPONDENCE
Maik Beege
✉ maik.beege@ph-freiburg.de

RECEIVED 23 October 2025
REVISED 08 December 2025
ACCEPTED 09 December 2025
PUBLISHED 08 January 2026

CITATION
Beege M, Mengelkamp C and Nebel S (2026) Challenging the notion of a general disfluency effect: the moderating effect of element-interactivity on perceptual disfluency.
Front. Educ. 10:1731080.
doi: 10.3389/feduc.2025.1731080

Challenging the notion of a general disfluency effect: the moderating effect of element-interactivity on perceptual disfluency

Maik Beege[1,2]*, Christoph Mengelkamp[3] and Steve Nebel[4]

[1]Digital Media in Education, Department of Psychology, University of Education Freiburg, Freiburg, Germany, [2]Center for Interdisciplinary Research on Digital Education, University of Education, Freiburg, Germany, [3]Institute Human-Computer-Media, University of Würzburg, Würzburg, Germany, [4]Media Education, Structural Unit Education Science, Faculty of Human Sciences, University of Potsdam, Potsdam, Germany

Introduction: The disfluency effect proposes that deliberately introducing challenges or difficulties into the learning process can be advantageous. Particularly, perceptual disfluency (e.g., harder to read fonts) might affect learning positively. However, an ongoing debate persists regardxing the robustness of this effect, as some studies have failed to replicate it or have uncovered opposing outcomes.

Methods: To investigate potential moderators of the disfluency effect, two experiments were conducted using different types of instructional materials (instructional texts: N_1 = 76; concept maps: N_2 = 74). In both experiments, the fluency was manipulated by using either a legible font or an illegible font, while element interactivity (high vs. low) was manipulated as a moderator. Learning outcomes, cognitive load, accuracy of metacognitive judgments, learning time, and efficiency were assessed in both experiments.

Results: Results indicated that disfluency did not have a general impact on the dependent variables, except for a detrimental effect on extraneous load. Notably, disfluency increased learning outcomes and germane load for low element interactivity. Contrary to common explanations of the disfluency effect, the use of a disfluent font did not yield metacognitive benefits.

KEYWORDS

cognitive load, disfluency effect, element interactivity, metacognition, perceptual disfluency

1 Introduction

We often reminisce about our university education. Students either hastily transcribed notes from the blackboard or received worksheets from teachers containing information that was copied with varying degrees of legibility. During exam preparation, we would revisit these materials, despite their frequent illegibility. Today, most students have access to digital presentation slides featuring excellent readability due to computerized fonts, or printed materials that are easy to read. If you had a choice between less legible materials on one hand and highly legible ones on the other, which would you use for learning? Although the straightforward choice might be the legible materials, some studies suggest otherwise. Current studies have investigated the effects of purposefully designing educational materials to be more challenging to perceive. One such approach involves presenting texts in difficult-to-read fonts. What may sound paradoxical has actually been defined as the disfluency effect (Yue et al., 2013).

1.1 The disfluency effect

The exploration of perceptual disfluency's impact in learning environments began with Diemand-Yauman et al. (2011; whereas older work referring to perceptual inference goes back further; e.g., Mulligan, 1999), who found that hard-to-read fonts improved learning outcomes in university and high school students. In this vein, decreasing text-font legibility, as demonstrated by Alter and Oppenheimer (2009), Beege et al. (2021), and Yue et al. (2013), is a means of adding difficulties to learning material. A potential explanation originates from research conducted by James, 1950 (1890/1950) and Kahneman and Frederick (2002), who proposed a dual-system model of human processing, i.e., a quick, intuitive System One and a slow, analytical System Two (Eitel and Kühl, 2016). The activation of these systems depends on perceived task difficulty, with System One operating for easy tasks and System Two for difficult ones (Alter et al., 2007). Introducing desirable difficulties deliberately, such as illegible learning materials, can enhance learning by engaging learners in deeper analytical System Two processing (Diemand-Yauman et al., 2011; Pashler et al., 2007).

1.1.1 Disfluency and cognitive load

Considering Cognitive Load Theory (CLT) (Sweller, 1988; Sweller et al., 2019), the postulated positive effects of perceptual disfluency on learning processes should at least be questioned. One objective of the CLT is to derive design principles to adjust learning materials to the limited capacity of working memory and to facilitate learning (Paas et al., 2003; Sweller et al., 2019). The cognitive load imposed on the working memory can be classified into three distinct categories, namely intrinsic, extraneous, and germane cognitive load (Paas et al., 2003; Sweller et al., 1998). Firstly, intrinsic cognitive load (ICL) can be defined as the inherent intricacy of the educational material (Kalyuga, 2011). It is dependent upon both the complexity of the task and the learner's domain-specific prior knowledge (Sweller et al., 2019). The complexity of the task increases the intrinsic load, whereas prior knowledge decreases the intrinsic load for example by the ability of experts to chunk smaller units of information into larger units (e.g., Thalmann et al., 2019). The load on working memory arises not solely from the intricacy of the task but also from the design of the instructional material. Inappropriately designed instructional materials introduce extraneous cognitive load (ECL), which hinders the learning process (Paas and Sweller, 2014). Consequently, extraneous processing should be minimized to foster the construction of mental schemas and the automation of knowledge (de Jong, 2010). The third component, germane cognitive load (GCL), is characterized as a redistributive function (Sweller et al., 2019) and as an active processing mechanism (i.e., mental effort; Jiang and Kalyuga, 2020). To be more precise, GCL does not constitute the entire cognitive load but instead allocates working memory capacities to activities that are germane to the learning process (Kalyuga, 2011).

Regarding the CLT, the role of perceptual disfluency in learning is ambiguous, as indicated by several studies (e.g., Eitel et al., 2014; Lehmann et al., 2016). On the one hand, disfluent instructional materials contribute to ECL due to cognitive resources that are needed to contend with suboptimal instructional design (Seufert et al., 2017). Illegible or difficult-to-process fonts necessitate deciphering efforts before effective learning can commence. Conversely, disfluent text may increase GCL as learners become more actively engaged in the activation of elaborative strategies (Alter et al., 2007) and of mental effort (Jiang and Kalyuga, 2020). Thus, according to explanations of the disfluency effect, GCL was of particular interest and explicitly measured.

1.1.2 Disfluency and metacognitive processes

Disfluency may trigger metacognitive processes that are central for monitoring comprehension (Dunlosky and Lipko, 2007; Yang et al., 2023), with ease of learning (EOL; pre-assessment of how easy information will be to learn), judgments of learning (JOL; predictions made during or after learning about recall ability), and retrospective confidence (RC; post-task confidence in the accuracy of recalled information) being key concepts (Nelson and Narens, 1990). Metacognitive accuracy scores, derived from these variables, gauge learners' ability to assess their comprehension accurately and adjust strategies during learning. Potentially, perceptual disfluency improves the accuracy of metacognitive judgments, leading to more appropriate control strategies (Metcalfe, 2002). Pieger et al. (2017) revealed that perceptual disfluency serves as a cue that alerts learners to potential challenges and prompts them to reassess their comprehension and abilities more critically, reducing the likelihood of overestimating their knowledge or skills. By fostering an environment where learners must actively question and evaluate their learning processes, disfluency can effectively diminish overconfidence, leading to more realistic self-evaluations, an enhanced metacognitive accuracy and thus, and adaptive learning strategies. Disfluency's effects on metacognition, particularly metacognitive judgments, have been explored in various studies (e.g., Ball et al., 2014; Ebersbach et al., 2023; Pieger et al., 2017).

According to the Effort Monitoring and Regulation Framework (de Bruin et al., 2020), metacognitive processes require working memory resources. Indeed, working memory capacity is correlated with the accuracy of metacognitive judgments when reading a text for the first time (e.g., Griffin et al., 2008, Experiment 2). A recent metamemory study (Bryce et al., 2023) found experimental support in two out of three experiments, using three different working memory tasks, that the accuracy of retrospective confidence (RC) judgments decreased if working memory demands were increased. In the third experiment, in addition to working memory demands, the perceptual fluency of the stimuli was manipulated. The accuracy of RC judgments decreased as working memory demands were increased, whereas the (dis)fluency of the stimuli did not affect the accuracy of the RC judgments. Possibly, the fluency manipulation was not strong enough in this experiment as no effects on learning were found either. To sum up, it seems reasonable that the accuracy of metacognitive judgments suffers when cognitive load is high. Xie et al. (2018) pointed out that disfluent texts reduced judgments of learning (JOL) in contrast to fluent texts, indicating that learners tend not to overestimate their learning performance. Consequently, this may enhance the accuracy of metacognitive judgments in relation to further learning outcomes. However, the effects of disfluency on metacognitive judgments and their accuracy will be smaller for highly demanding tasks (i.e., high cognitive load) compared to less demanding tasks.

1.1.3 Moderating effects

These arguments emphasize that both perspectives (cognitive load as well as metacognitive processes) should be considered simultaneously. Consequently, boundary conditions or moderators of the disfluency effect come into focus, under which unfavorable effects associated with ECL are prevalent or favorable effects pertaining to GCL or metacognitive accuracy come to the fore (e.g., Geller et al., 2018).

One example is that signaling (visual cues to highlight important information, aiding focus and understanding in learning materials) moderates the effect of disfluency on mental effort and transfer (Lai and Zhang, 2021). Specifically, the disfluent text led to better learning outcomes with or without signaling. In the fluent text condition, only signaling facilitated learning. Further, Seufert et al. (2017) explored whether the level of perceptual disfluency (low, medium, or high) impacted learning outcomes. Their findings indicated that a high level of disfluency, where the text is almost unreadable, hinders learning success. Conversely, moderate levels of disfluency were beneficial for learning, leading to lower perceptions of extraneous load, higher engagement, and improved recall of information. However, the authors also stressed the uncertainty surrounding the optimal degree of perceptual disfluency for learning, especially in relation to extraneous load. In this vein, studies with regard to word recognition revealed that the level of blurring seems to have an effect; the effect was in the opposite direction for 10% blurring compared to 15% blurring. Subsequent studies (Rummer et al., 2016; Eitel et al., 2014) revealed mixed results as well, with disfluency effects not consistently replicated. Geller et al. (2018) investigated the impact of font text on recognition in a word list learning task, finding cursive font to be a desirable difficulty, although the degree of perceptual disfluency significantly affected the results. The authors found that easy-to-read cursive words tended to be better remembered than hard-to-read cursive words.

In summary, empirical evidence on the learning benefits of hard-to-process instructional materials is inconclusive, with some studies reporting no benefit for disfluency (Faber et al., 2017; İlic and Akbulut, 2019; Rummer et al., 2016; Yue et al., 2013). A meta-analysis by Xie et al. (2018) questions the robustness of the disfluency effect in text-based educational settings, suggesting no significant effects on recall (d = –0.01) and transfer outcomes (d = 0.03) but a negative impact on judgments of learning (d = –0.43). However, from a methodical perspective, the meta-analysis and the reporting of the results have been questioned (Weissgerber et al., 2021).

1.2 The moderating role of element-interactivity

Beege et al. (2021) highlighted another significant moderator, which is also the focus of this research paper. According to them, the element interactivity of the learning material (dependency or independency of single elements within the material) plays a crucial role in determining the effectiveness of disfluency. In cognitive load research, increased element interactivity is referred to as increased complexity. Across three experiments, the authors increased the element interactivity of the learning material by making the elements more dependent on each other and thus increased the complexity of the learning content. In one study, rather basic information had to be remembered; in a second study, the learning topic was related to biochemical information; and in a third study, a mathematical scheme was taught that systematically built on each other. Results revealed that if learning materials have low element interactivity, disfluency will have positive effects on learning outcomes and meaningful processing.

To get an insight into the learning relevant mechanisms, disfluency can basically be viewed as additional ECL but has the potential to trigger additional, learning-promotion processes as discussed above. To specify whether it is the cognitive load that hinders learning or the mechanisms that promote learning that are at work, one can consider element interactivity as a moderator variable. That is, in learning scenarios demanding a low ICL, learners have sufficient cognitive resources available to cope with the additional induced ECL. Therefore, the disfluency triggers the investment of effort in processing the learning material, leading to beneficial effects on learning (Alter et al., 2007). However, if element interactivity is high, disfluency will have a negative impact on learning and learning efficiency (e.g., Paas and Van Merriënboer, 1993), as learners are already heavily burdened by the material's complexity and have no resources left for processing the additional ECL induced by the disfluent material. Thus, there is no capacity left for learning-promotion processes. Nevertheless, Beege et al. (2021) did not manipulate element interactivity within an experiment, but they investigated materials of varying complexity across three experiments using different learning materials. Thus, element interactivity may have been confounded by the materials used. Furthermore, results were ambiguous even within individual experiments. Nevertheless, related studies might support the results from Beege et al. (2021). Lehmann et al. (2016) observed the disfluency effect only when learners had a high working memory capacity. However, this effect did not replicate in a more recent study (Weissgerber et al., 2023). To further corroborate the findings by Beege et al. (2021), Lehmann et al. (2016), and Weissgerber et al. (2023), the current study investigates the moderating effect of element interactivity on the disfluency effect through manipulating element interactivity within experiments.

As mentioned previously, element interactivity contributes to ICL (Sweller, 1994). An element refers to any piece of information that must be processed to comprehend the learning content. In learning materials of low element interactivity, each element is processed independently or with minimal reference to other elements, such as learning vocabulary in a foreign language. In contrast, high element interactivity materials consist of elements that are strongly interrelated, e.g., learning the grammar of a language or understanding how an engine works. To comprehend the learning content, these elements must be processed simultaneously in the working memory. Consequently, the more elements interact with one another, the greater the cognitive load on working memory (Tindall-Ford et al., 1997). According to the element interactivity effect, in high element interactivity materials, instructional support should be provided to comply with the load on working memory. In low-element interactivity materials, this support is not needed because learners have cognitive resources available to process materials that are suboptimal designed (Chen et al., 2015, 2017). Consequently, in high-element interactivity materials, additional cognitive load should be avoided, whereas in low-element interactivity materials, additional load (e.g., imposed by disfluent material) may have a smaller or negligible negative impact on learning. For the present study, we take the work of Sweller (2010) into account, extending the concept of element interactivity: "element interactivity is the major source of working memory load underlying both extraneous and intrinsic cognitive load" (Sweller, 2010, p. 125). Thus, design features in learning environments can be additional elements that interact with the core content and must be processed unavoidably (Beckmann, 2010). For example, a text that is convoluted with non-linear structure or excessive length can transform elements that would otherwise be simple into complex interactive units requiring simultaneous consideration and integration by the learner and thus, enhance element interactivity. Therefore, optimizing instructional design by simplifying text complexity is not only associated with reduced ECL but also affects element interactivity.

1.3 The present study

Two experiments were carried out to (1) replicate the null-effect of disfluency when element interactivity is high and (2) test if disfluency is beneficial for learning when element interactivity is low. In this vein, the focus lies on one prominent form of disfluency: perceptual disfluency. In both of our experiments, we manipulated perceptual disfluency and element interactivity, but the experiments used different instructional media (instructional texts vs. concept maps), allowing us to generalize the findings to verbal and pictorial learning materials. The learning content was biology in both experiments to minimize the effects of the learning content on the results and to recruit the same participants for both experiments. In line with Beege et al.'s results (2021) and considering the effects of element interactivity discussed above, we hypothesized that disfluency enhances learning when element interactivity is low. Conversely, when element interactivity is high, disfluency should hinder learning.

> *H1a*: There is no main effect for disfluency regarding learning outcomes.

> *H1b:* Disfluency fosters learning outcomes when the element interactivity of the learning material is low but impairs learning outcomes when the element interactivity is high.

Considering CLT (e.g., Sweller et al., 2019), disfluency should increase GCL if learners have sufficient resources for effortful processing (Alter et al., 2007). Consequently, disfluency will increase GCL if element interactivity is low. Furthermore, if the manipulation of element interactivity works, ICL will be increased. This hypothesis was tested as a manipulation check. A further manipulation check was carried out with respect to ECL. According to the CLT, an illegible font should increase perceptions of ECL (Seufert et al., 2017).

> *H2:* Disfluency increases germane processes when the element interactivity of the learning material is low but impairs germane processes when the element interactivity is high.

> *H3:* Increasing the element interactivity leads to increases in ICL.

> *H4:* Increasing the disfluency leads to increases in ECL.

Since metacognitive variables are essential to explain the disfluency effect (Seufert et al., 2017), metacognitive variables were investigated as well. In line with Pieger et al. (2017), disfluency should reduce overconfidence and, thus, enhance absolute metacognitive accuracy.

> *H5:* Increasing the disfluency leads to increases in absolute metacognitive accuracy.

Finally, learning time was tracked and explored. Time measures and learning outcomes are further used to calculate instructional efficiency which will be included in the statistical analyses.

2 Experiment 1

2.1 Methods

2.1.1 Participants and design

Since some studies and a meta-analysis did not detect a disfluency effect, no clear effect size could be determined for a power analysis for main effects. An orientation could be provided by Beege et al. (2021). When element-interactivity was low, disfluency fostered learning outcomes with a large effect size ($\eta_p{}^2$ = 0.30). When element-interactivity was high, disfluency impaired learning outcomes with a large effect size ($\eta_p{}^2$ = 0.16). Thus, a large effect size was assumed for the current analysis with respect to a potential interaction effect. An a-priori power analysis ($1-\beta$ = 0.90, f = 0.40, α = 0.05) revealed that 68 participants should be suitable for the current investigation. Overall, 76 students (77.3% female, age: M = 22.46 years, SD = 3.04) from Chemnitz University of Technology and Freiburg University of Education were included in the statistical analyses. Students were enrolled in media communication/psychology (77.6%), teacher education (18.5%), and other fields of study (3.9%). Each participant received either 5€ or course credit. As expected, prior knowledge of the participants in biocenoses (M = 0.46, SD = 0.7, with a maximum of six points) was low.

The participants were randomly assigned to one condition in a 2 (perceptual disfluency: disfluent text vs. fluent text) × two (element interactivity: high vs. low) factorial between-subjects design using the online software LimeSurvey (LimeSurvey GmbH, 2020). No significant differences were found between conditions in terms of prior knowledge, students' semester, or age, $F(3, 72)$ =

(0.59, 1.33), p = (0.27, 0.63) as well as gender, or participant's field of study, χ^2 = (0.55, 2.12), p = (0.55, 0.76).

2.1.2 Materials

The learning material consisted of an instructional text dealing with the basic definition of biocenoses, the subdivision for the study of species, and succession and disturbances of biocenoses. The text was divided into five subsections, and each section was presented on a separate webpage. The text was presented in full screen. The text was self-paced; participants decided how long they wanted to stay on each webpage before moving on to the next section. However, participants were not allowed to move back. Learning time was tracked.

2.1.2.1 Perceptual disfluency

In the fluent condition, the text was presented in the font Arial. The text was written in black on a neutral white background. To create the material for the disfluent condition, the text from the fluent condition was printed, scanned, and repeatedly printed and scanned again. Consequently, readability was reduced in the disfluent condition. We aimed for a medium disfluency since Seufert et al. (2017) outlined that a slight or excessive disfluency does not lead to positive effects on learning processes. Readability was measured using one item as an additional measure. The item was: "How would you rate the readability of the text?" Test subjects could rate this on a Likert scale from 1 (very poor readability) to 6 (ideal readability). The fluent text was rated as easy to read (M = 4.50; SD = 1.52) and the disfluent text was difficult to read, but not very difficult (M = 2.45; SD = 0.68). The difference was statistically significant, t = 7.45, p < 0.001, d = 1.74 (Welch corrected because of violation of variance homogeneity and normal distribution). This procedure ensured that the text remained the same. Examples of a text from the fluent vs. disfluent condition are displayed in Figure 1.

2.1.2.2 Element interactivity

It was a challenge to manipulate element interactivity without changing the content of the learning material and, thus, allowing for the same learning test in both conditions. Element interactivity refers to the number of elements that must be processed simultaneously to comprehend the learning content (Sweller et al., 2019). Consequently, element interactivity was manipulated by altering the sentence structure of the text. To measure element interactivity within our material, we based our approach on a recent publication by Chen et al. (2024) and counted the elements within a coherent sentence. In the low element interactivity condition, the text was presented with a simple sentence structure, presenting short main sentences one after the other (1.48 elements per sentence). Consequently, only a few elements must be processed simultaneously to understand each sentence. In the high element interactivity condition, the text was made up of a combination of main clauses and several subordinate clauses (3.68 elements per sentence). Consequently, multiple elements of information had to be maintained in working memory simultaneously to understand each sentence. An example of the manipulation of element interactivity is displayed in Figure 2. A translated example can be found in Supplementary Appendix A.

2.1.3 Measures

2.1.3.1 Cognitive load

Cognitive load was assessed using the self-report scale from Klepsch et al. (2017), which provides sub-scales for the three types of cognitive load. The questionnaire was created and validated to measure cognitive load facets in multimedia learning studies, independently from the learning media as well as the investigated design principle. Thus, this questionnaire is used to determine whether the experimental variations have an influence on the general manifestation of cognitive load, without the items being explicitly formulated in the context of disfluency or element interactivity. Two items measured ICL (Cronbach's α = 0.80, Spearman's ϱ = 0.67, e.g., "This task was very complex"). Three items measured ECL (α = 0.80, e.g., "During this task, it was exhausting to find the important information"). Two items measured GCL (α = 0.60, ϱ = 0.46, e.g., "My point while dealing with the task was to understand everything correctly"). The participants rated the items on a 7-point scale ranging from 1 (absolutely wrong) to 6 (absolutely correct). However, Sweller et al. (2011) pointed out that subjective reports can be influenced by learners' perceptions and metacognitive awareness. Participants may not accurately gauge their cognitive state or might conflate effort with actual cognitive load (Paas et al., 2003). Thus, our scales are referred to as perceived cognitive load.

2.1.3.2 Metacognitive judgments

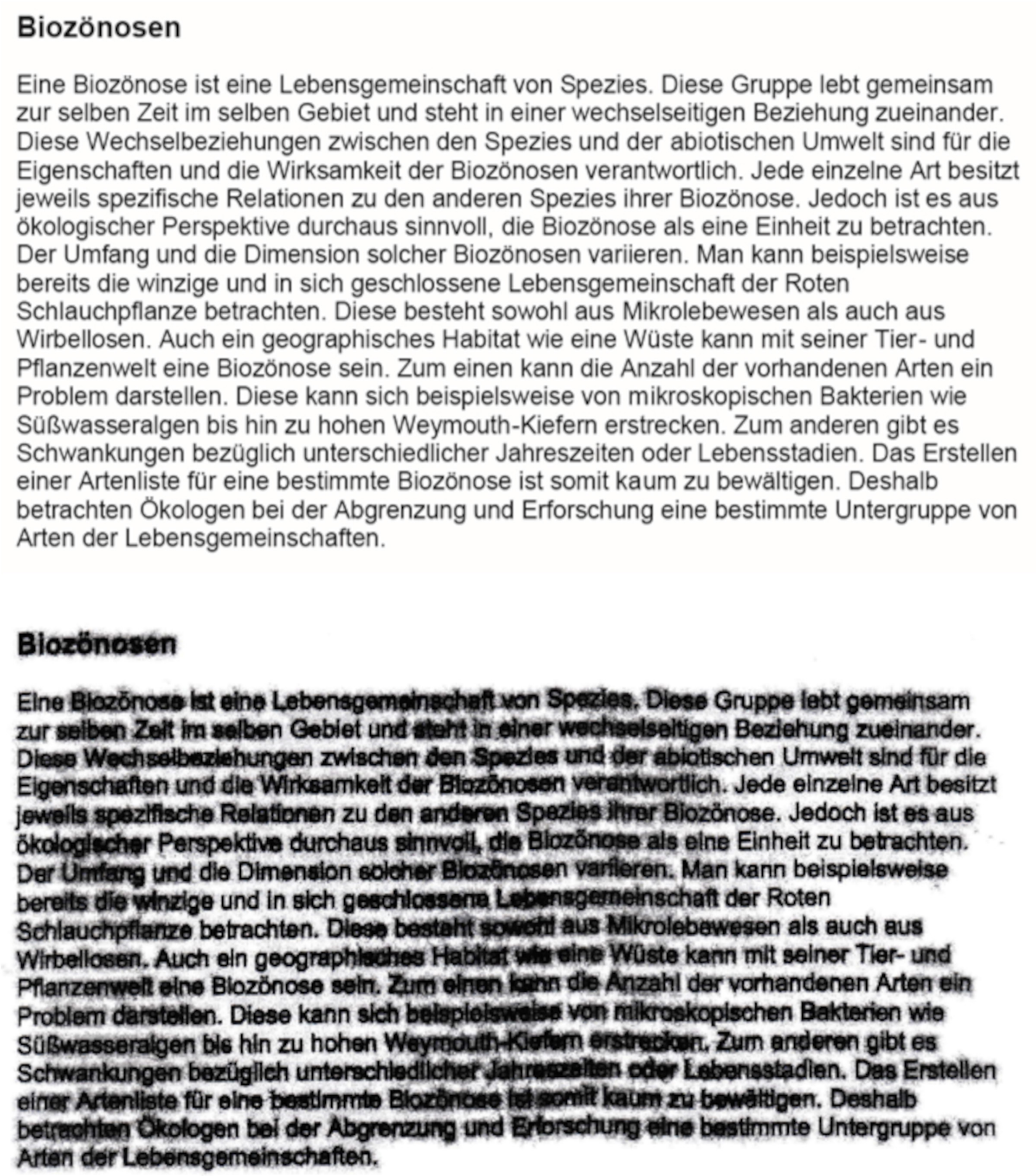
The procedure for assessing metacognitive judgments and metacognitive accuracy was based on Pieger et al. (2017). Ease of Learning (EOL) was measured asking the question "How easy or difficult will it be to learn the text?" and participants indicated their answer on a visual analog scale from 0 (very easy) to 100 (very difficult). A judgment of learning (JOL) was obtained after the learning phase ("Imagine you had to answer study questions about the text you read. What percentage of the questions do you estimate you will be able to answer correctly?") on a visual analog scale from 0 (no questions) to 100 (all questions). Retrospective confidence (RC) was assed asking the question, "What percentage of the study questions you just answered do you estimate you answered correctly?" to be answered on a visual analog scale from 0 (no correct answers) to 100 (all answers correct). One RC judgment was made after a retention test and one RC judgment after a transfer test.

In addition, we examined the absolute accuracy of the metacognitive judgments. We calculated the absolute difference between the estimated proportion of correct responses (i.e., dividing the judgment by 100) and the actual proportion of correct responses for both retention and transfer items (cf. Schraw, 2009, for a similar measure). We utilized the proportion of correct responses derived from both retention and transfer scores, as participants provided an overall judgment of their correct responses. For the RC judgments, the average of the two judgments was calculated and subtracted from the average learning scores in the test.

$$Accuracy_{EOL} = |proportion_{EOL} - proportion_{correct}|$$

$$Accuracy_{JOL} = |proportion_{JOL} - proportion_{correct}|$$

$$Accuracy_{RC} = |proportion_{RC} - proportion_{correct}|$$



Biozönosen

Eine Biozönose ist eine Lebensgemeinschaft von Spezies. Diese Gruppe lebt gemeinsam zur selben Zeit im selben Gebiet und steht in einer wechselseitigen Beziehung zueinander. Diese Wechselbeziehungen zwischen den Spezies und der abiotischen Umwelt sind für die Eigenschaften und die Wirksamkeit der Biozönosen verantwortlich. Jede einzelne Art besitzt jeweils spezifische Relationen zu den anderen Spezies ihrer Biozönose. Jedoch ist es aus ökologischer Perspektive durchaus sinnvoll, die Biozönose als eine Einheit zu betrachten. Der Umfang und die Dimension solcher Biozönosen variieren. Man kann beispielsweise bereits die winzige und in sich geschlossene Lebensgemeinschaft der Roten Schlauchpflanze betrachten. Diese besteht sowohl aus Mikrolebewesen als auch aus Wirbellosen. Auch ein geographisches Habitat wie eine Wüste kann mit seiner Tier- und Pflanzenwelt eine Biozönose sein. Zum einen kann die Anzahl der vorhandenen Arten ein Problem darstellen. Diese kann sich beispielsweise von mikroskopischen Bakterien wie Süßwasseralgen bis hin zu hohen Weymouth-Kiefern erstrecken. Zum anderen gibt es Schwankungen bezüglich unterschiedlicher Jahreszeiten oder Lebensstadien. Das Erstellen einer Artenliste für eine bestimmte Biozönose ist somit kaum zu bewältigen. Deshalb betrachten Ökologen bei der Abgrenzung und Erforschung eine bestimmte Untergruppe von Arten der Lebensgemeinschaften.

Biozönosen

Eine Biozönose ist eine Lebensgemeinschaft von Spezies. Diese Gruppe lebt gemeinsam zur selben Zeit im selben Gebiet und steht in einer wechselseitigen Beziehung zueinander. Diese Wechselbeziehungen zwischen den Spezies und der abiotischen Umwelt sind für die Eigenschaften und die Wirksamkeit der Biozönosen verantwortlich. Jede einzelne Art besitzt jeweils spezifische Relationen zu den anderen Spezies ihrer Biozönose. Jedoch ist es aus ökologischer Perspektive durchaus sinnvoll, die Biozönose als eine Einheit zu betrachten. Der Umfang und die Dimension solcher Biozönosen variieren. Man kann beispielsweise bereits die winzige und in sich geschlossene Lebensgemeinschaft der Roten Schlauchpflanze betrachten. Diese besteht sowohl aus Mikrolebewesen als auch aus Wirbellosen. Auch ein geographisches Habitat wie eine Wüste kann mit seiner Tier- und Pflanzenwelt eine Biozönose sein. Zum einen kann die Anzahl der vorhandenen Arten ein Problem darstellen. Diese kann sich beispielsweise von mikroskopischen Bakterien wie Süßwasseralgen bis hin zu hohen Weymouth-Kiefern erstrecken. Zum anderen gibt es Schwankungen bezüglich unterschiedlicher Jahreszeiten oder Lebensstadien. Das Erstellen einer Artenliste für eine bestimmte Biozönose ist somit kaum zu bewältigen. Deshalb betrachten Ökologen bei der Abgrenzung und Erforschung eine bestimmte Untergruppe von Arten der Lebensgemeinschaften.

FIGURE 1
Manipulation of disfluency in Experiment 1 (upper panel: fluent, lower panel: disfluent).

Please note, that these indices indicate a discrepancy between a confidence judgment and a learning score. Higher values indicate a reduced accuracy. Therefore, these indices are further referred as inaccuracy indices.

2.1.3.3 Knowledge measures

A prior knowledge test assessed prior knowledge about biocenoses. Three open questions were included (α = 0.84). Thus, participants gained up a maximum of six points in total. The questions covered different sub-topics of the learning content (e.g., "What are biocenoses?").

Furthermore, a retention test and a transfer test were used to assess the learning outcomes. Retention refers to recall of learning content, whereas transfer is defined as solving novel problems that were not presented explicitly within the learning content (Mayer, 2014). The retention test consisted of nine multiple-choice questions and six open questions (α = 0.83, e.g., "What species belong to the biocenosis of the red hose plant?"). One open question had to be excluded from the analyses because of low inter-rater reliability (ICC = 0.42). The multiple-choice questions consisted of four possible answers of which one, two, three, or all answers could be correct. Participants gained one point when they marked a correct answer or correctly rejected a wrong answer. This

Biozönosen

Eine Biozönose ist eine Lebensgemeinschaft von Spezies. Diese Gruppe lebt gemeinsam zur selben Zeit im selben Gebiet und steht in einer wechselseitigen Beziehung zueinander. Diese Wechselbeziehungen zwischen den Spezies und der abiotischen Umwelt sind für die Eigenschaften und die Wirksamkeit der Biozönosen verantwortlich. Jede einzelne Art besitzt jeweils spezifische Relationen zu den anderen Spezies ihrer Biozönose. Jedoch ist es aus ökologischer Perspektive durchaus sinnvoll, die Biozönose als eine Einheit zu betrachten. Der Umfang und die Dimension solcher Biozönosen variieren. Man kann beispielsweise bereits die winzige und in sich geschlossene Lebensgemeinschaft der Roten Schlauchpflanze betrachten. Diese besteht sowohl aus Mikrolebewesen als auch aus Wirbellosen. Auch ein geographisches Habitat wie eine Wüste kann mit seiner Tier- und Pflanzenwelt eine Biozönose sein. Zum einen kann die Anzahl der vorhandenen Arten ein Problem darstellen. Diese kann sich beispielsweise von mikroskopischen Bakterien wie Süßwasseralgen bis hin zu hohen Weymouth-Kiefern erstrecken. Zum anderen gibt es Schwankungen bezüglich unterschiedlicher Jahreszeiten oder Lebensstadien. Das Erstellen einer Artenliste für eine bestimmte Biozönose ist somit kaum zu bewältigen. Deshalb betrachten Ökologen bei der Abgrenzung und Erforschung eine bestimmte Untergruppe von Arten der Lebensgemeinschaften.

Biozönosen

Eine Biozönose ist eine Lebensgemeinschaft von Spezies, die gemeinsam zur selben Zeit im selben Gebiet leben und in einer wechselseitigen Beziehung zueinanderstehen, wobei diese Wechselbeziehungen zwischen den Spezies und der abiotischen Umwelt für die Eigenschaften und die Wirksamkeit der Biozönosen verantwortlich sind. Jede einzelne Art besitzt jeweils spezifische Relationen zu den anderen Spezies ihrer Biozönose, jedoch ist es aus ökologischer Perspektive durchaus sinnvoll, die Biozönose als eine Einheit zu betrachten. Der Umfang und die Dimension solcher Biozönosen variieren: man kann beispielsweise bereits die winzige und in sich geschlossene Lebensgemeinschaft der Roten Schlauchpflanze betrachten, welche sowohl aus Mikrolebewesen als auch aus Wirbellosen besteht, und auch ein geographisches Habitat wie eine Wüste kann mit seiner Tier- und Pflanzenwelt eine Biozönose sein. Zum einen kann die Anzahl der vorhandenen Arten ein Problem darstellen, da diese sich von beispielsweise mikroskopischen Bakterien wie Süßwasseralgen bis hin zu hohen Weymouth-Kiefern erstrecken kann, zum anderen gibt es Schwankungen bezüglich unterschiedlicher Jahreszeiten oder Lebensstadien. Das Erstellen einer Artenliste für eine bestimmte Biozönose ist somit kaum zu bewältigen, weshalb Ökologen bei der Abgrenzung und Erforschung, eine bestimmte Untergruppe von Arten der Lebensgemeinschaften betrachten.

FIGURE 2
Manipulation of element interactivity in Experiment 1 (upper panel: low element interactivity, lower panel: high element interactivity).

approach was chosen because giving points for correctly rejecting false answers reduces guessing and leads to higher reliability of the knowledge test (Burton, 2005). Therefore, participants gained a maximum of 51 points in the retention test.

The transfer test consisted of two multiple-choice questions and seven open questions (α = 0.83, e.g., "What impacts can beavers cutting trees have on the biocoenosis?"). Overall, students gained up to 24 points on the transfer test. For the open questions (prior knowledge, retention, and transfer), intra-class correlation coefficients (ICCs) were satisfactory, ICC (2, k) = (0.65, 0.97), F(75, 75) = (5.19, 59.51), ps < 0.001 (Koo and Li, 2016). All items per test were presented on the same page.

Additionally, learning efficiency was calculated. Since learning material was self-paced, it could be assumed that learning time increases when learning with disfluent material (Xie et al., 2018). Thus, a potential confounding might occur (disfluency and learning time are varied between the conditions). Efficiency was used to assess the learning gain in dependence of the time invested in the learning phase to resolve this issue. Further, in line with Gorbunova et al. (2025), discussion of efficiency emphasizes the importance of well-designed instructional strategies to address persistent challenges such as learner engagement and learning outcomes together (enhancing outcomes by minimizing temporal cost). According to Gorbunova et al. (2025), efficiency is be a multi-facet construct (for examples, efficiency in terms of time and cognitive effort). However, we focus on the temporal facet to control for learning time. Therefore, the formula from van Gog and Paas (2008) was used. *Z*-standardized learning outcomes were used for zL and z-standardized time (time in which participants worked on the learning text) was used for zT. Efficiency was calculated for retention and transfer separately.

$$Efficiency = \frac{zL - zT}{\sqrt{2}}$$

2.1.4 Procedure

The study was conducted in a supervised online setting using the educational platform *BigBlueButton*, where up to four students participated simultaneously. Each student was allocated to a separate breakout room and provided with a link for study participation. Students shared their screens during the experiment. First, the participants took the prior knowledge test, followed by receiving a link to the learning material. They were instructed to have a preliminary look at the text for 2 s. Then, the participants were automatically redirected to a webpage and made the EOL judgment. Subsequently, the learning phase began. Participants were instructed that a learning test was implemented after studying. Students were asked to read the text at their own pace, navigating from one webpage to the next, but not allowed to navigate backwards. Upon completion of the learning phase, students were asked to make the JOL. Next, dependent variables were measured after the learning phase as follows: cognitive load, retention, and transfer. The learning scales were presented in another font (Noto Sans) to prevent retrieval cue effects. One RC judgment had to be made after the retention test, and another RC judgment after the transfer test. Finally, students were asked to fill in demographic questions before exiting the *BigBlueButton* platform. On average, the entire experiment lasted for a total of 45 min.

2.1.5 Analysis strategy

All analyses were conducted using JASP (JASP Team, 2025). Hypotheses were tested by conducting analyses of variance (ANOVAs) for learning outcomes (i.e., retention and transfer), cognitive load (i.e., ICL, ECL, and GCL), and metacognitive judgments (i.e., EOL judgment, JOL, RC). Disfluency and element interactivity were used as independent variables, and dependent variables were chosen in line with the hypotheses. Following the argumentation from Huang (2020), no omnibus-MANOVAs were calculated. To investigate significant interactions, Bonferroni-Holm-corrected post hoc tests were conducted between all conditions. Since no other variable (i.e., gender, participant's field of study, prior knowledge, age, semester) differed significantly between conditions, no covariates were included. To provide evidence for the null hypothesis regarding the disfluency effect (H1a), Bayes Factors [BF_{10} and log (BF_{10}); directed hypothesis in favor of the disfluency condition] were calculated for retention and transfer as dependent variables. Bayes factors were conducted using a Bayesian ANOVAs. Since a recent meta-analysis revealed no significant effect for disfluency (Xie et al., 2018), priors were adjusted in favor to the null hypothesis. Thus, a beta binomial model prior with $\alpha = 1$ and $\beta = 3$ was chosen. However, since results strongly differ between single studies, the prior coefficient was set to $r_{fixedeffects} = 1.00$. Descriptive statistics are displayed in Table 1. Bar charts for all dependent variables can be found in Supplementary Appendices B–D. Since we conducted several analyses concerning multiple measures, we included a summary of all results in Supplementary Appendix I.

2.2 Results

2.2.1 Learning outcomes and efficiency

An ANOVA with retention as dependent variable revealed no effect for disfluency, $F(1, 72) = 2.64$, $p = 0.109$, $\eta_p^2 = 0.04$ but for element interactivity, $F(1, 72) = 5.66$, $p = 0.020$, $\eta_p^2 = 0.07$; and a significant interaction, $F(1, 72) = 4.38$, $p = 0.040$, $\eta_p^2 = 0.06$. *Post-hoc* tests revealed that disfluency increased retention scores in the condition with low element interactivity ($p = 0.043$). If element interactivity is high, no significant difference was found regarding disfluency ($p = 1.00$) (see Supplementary Appendix B). Regarding transfer, no effect for disfluency was found, $F(1, 72) = 0.34$, $p = 0.564$, $\eta_p^2 = 0.01$, but an effect for element interactivity, $F(1, 72) = 8.261$, $p = 0.005$, $\eta_p^2 = 0.10$. Low element interactivity fostered learning outcomes. No interaction was found, $F(1, 72) = 1.46$, $p = 0.230$, $\eta_p^2 = 0.02$. Bayes factors did not support the null hypotheses regarding disfluency for retention, $BF_{10} = 0.46$, log (BF_{10}) = −0.77, but for transfer, $BF_{10} = 0.16$, log (BF_{10}) = −1.84. Descriptively, the group with low element interactivity and disfluent learning material showed the strongest learning success in terms of both learning measures (see Supplementary Appendix B).

Additionally, an ANOVA revealed a main effect on learning time with a large effect size for disfluency, $F(1, 72) = 16.97$, $p < 0.001$, $\eta_p^2 = 0.19$, but no effect for element interactivity, $F(1, 72) = 2.31$, $p = 0.133$, $\eta_p^2 = 0.03$, and no interaction effect, $F(1, 72) = 0.68$, $p = 0.411$, $\eta_p^2 = 0.01$. Therefore, disfluency increased learning time.

An ANOVA with retention efficiency as dependent variable revealed no effect for disfluency, $F(1, 72) = 2.62$, $p = 0.110$, $\eta_p^2 = 0.04$ but for element interactivity, $F(1, 72) = 6.93$, $p = 0.010$, $\eta_p^2 = 0.09$. Efficiency was enhanced in the low element interactivity conditions. There was no significant interaction, $F(1, 72) = 0.78$, $p = 0.380$, $\eta_p^2 = 0.01$. Regarding transfer efficiency, a significant effect was found for disfluency, $F(1, 72) = 5.08$, $p = 0.027$, $\eta_p^2 = 0.07$ and element interactivity, $F(1, 72) = 8.50$, $p = 0.005$, $\eta_p^2 = 0.11$. Efficiency was enhanced in the low element interactivity conditions. Further, disfluency impaired efficiency. There was no significant interaction, $F(1, 72) = 0.08$, $p = 0.776$, $\eta_p^2 = 0.001$. According to the descriptive data, the condition with high element interactivity and disfluent learning material showed reduced efficiency in contrast to the other conditions (see Supplementary Appendix B).

2.2.2 Perceived cognitive load

An ANOVA with ICL as dependent variable revealed no effect for disfluency, $F(1, 72) = 0.30$, $p = 0.58$, $\eta_p^2 = 0.004$ but for element interactivity, $F(1, 72) = 5.59$, $p = 0.021$, $\eta_p^2 = 0.07$. ICL was reduced in the low element interactivity conditions. There was no significant interaction, $F(1, 72) = 1.98$, $p = 0.164$, $\eta_p^2 = 0.03$. An ANOVA with ECL as dependent variable revealed a significant effect for disfluency, $F(1, 72) = 31.13$, $p < 0.001$, $\eta_p^2 = 0.30$ but no effect for element interactivity, $F(1, 72) = 0.26$, $p = 0.615$, $\eta_p^2 = 0.004$. ECL was enhanced in the disfluent conditions. There was

TABLE 1 Means and standard deviations of all dependent variables from Experiment 1.

Dependend varables	Experimental groups							
	Disfluent material				Fluent material			
	High element interactivity (N = 19)		Low element interactivity (N = 21)		High element interactivity (N = 19)		Low element interactivity (N = 17)	
	M	SD	M	SD	M	SD	M	SD
Retention	27.18	9.64	34.20	3.82	27.92	5.92	28.37	6.87
Transfer	11.29	4.96	15.38	3.95	11.92	4.41	13.59	4.04
ICL	4.32	1.02	4.07	1.05	4.82	0.99	3.85	1.37
ECL	5.11	0.63	5.00	0.71	3.95	1.14	3.84	1.05
GCL	3.87	0.70	4.95	0.59	3.79	0.75	4.09	0.87
EOL	37.47	25.19	19.38	19.34	42.74	27.84	44.82	22.65
JOL	36.74	25.93	31.62	20.39	38.37	20.65	41.12	19.28
RC	32.71	13.62	27.24	16.35	32.45	20.88	39.62	17.72
EOL accuracy	0.31	0.17	0.47	0.16	0.26	0.18	0.20	0.16
JOL accuracy	0.24	0.14	0.36	0.17	0.18	0.13	0.18	0.15
RC accuracy	0.21	0.13	0.39	0.14	0.25	0.14	0.21	0.15
Learning time	329.36	58.70	319.78	56.90	283.99	57.83	251.64	66.78
Learning efficiency (retention)	−0.57	1.03	0.22	0.66	−0.01	1.01	0.38	1.16
Learning efficiency (transfer)	−0.61	1.00	0.12	0.74	−0.03	1.10	0.12	0.74

For accuracy higher values indicate an increased discrepancy between confidence judgments and performance. Retention scores ranged from 0 to 51. Transfer scores ranged from 0 to 24. Cognitive load ratings ranged from 1 to 6. EOL, JOL, and RC ratings ranged from 0 to 100.

no significant interaction, $F(1, 72) < 0.001$, $p = 0.998$, $\eta_p^2 < 0.001$. Regarding GCL, there was a significant effect for disfluency, $F(1, 72) = 7.93$, $p = 0.006$, $\eta_p^2 = 0.10$ and for element interactivity, $F(1, 72) = 17.05$, $p < 0.001$, $\eta_p^2 = 0.19$. There was a significant interaction, $F(1, 72) = 5.50$, $p = 0.022$, $\eta_p^2 = 0.07$. *Post-hoc* tests revealed that disfluency increased GCL in the condition with low element interactivity ($p < 0.001$). If element interactivity is high, no significant difference was found regarding disfluency ($p = 0.739$) (see Supplementary Appendix C).

2.2.3 Metacognitive judgments and accuracy

An ANOVA with EOL judgment as dependent variable revealed a significant effect for disfluency, $F(1, 72) = 7.76$, $p = 0.007$, $\eta_p^2 = 0.10$ but not for element interactivity, $F(1, 72) = 2.11$, $p = 0.151$, $\eta_p^2 = 0.03$. Disfluency reduced EOL. There was no significant interaction, $F(1, 72) = 3.35$, $p = 0.071$, $\eta_p^2 = 0.04$. Regarding the JOL, there was no effect for disfluency, $F(1, 72) = 1.24$, $p = 0.270$, $\eta_p^2 = 0.02$, element interactivity, $F(1, 72) = 0.06$, $p = 0.814$, $\eta_p^2 < 0.001$, and no interaction, $F(1, 72) = 0.62$, $p = 0.434$, $\eta_p^2 = 0.01$. Regarding RC judgments (averaged for retention and transfer), there was no effect for disfluency, $F(1, 72) = 2.32$, $p = 0.132$, $\eta_p^2 = 0.03$, element interactivity, $F(1, 72) = 0.05$, $p = 0.832$, $\eta_p^2 < 0.001$, and no interaction, $F(1, 72) = 2.52$, $p = 0.117$, $\eta_p^2 = 0.03$.

Regarding EOL-accuracy, there was a significant effect regarding disfluency, $F(1, 72) = 16.33$, $p < 0.001$, $\eta_p^2 = 0.19$, but no effect for element interactivity, $F(1, 72) = 1.61$, $p = 0.209$, $\eta_p^2 = 0.02$. Overall, the inaccuracy indices were enhanced in the disfluency conditions indicating reduced EOL-accuracy. There was a significant interaction, $F(1, 72) = 8.21$, $p = 0.005$, $\eta_p^2 = 0.10$. *Post-hoc* tests revealed that disfluency reduced EOL-accuracy in the condition with low element interactivity ($p < 0.001$). If element interactivity is high, no significant difference was found regarding disfluency ($p = 0.549$) (see Supplementary Appendix D). Regarding JOL-accuracy, there was a significant effect regarding disfluency, $F(1, 72) = 11.74$, $p = 0.001$, $\eta_p^2 = 0.14$, but no effect for element interactivity, $F(1, 72) = 2.88$, $p = 0.094$, $\eta_p^2 = 0.04$. Overall, the inaccuracy indices were enhanced in the disfluency conditions indicating reduced JOL-accuracy. There was no significant interaction, $F(1, 72) = 3.32$, $p = 0.073$, $\eta_p^2 = 0.04$. Regarding RC-accuracy, there was a significant effect regarding disfluency, $F(1, 72) = 4.59$, $p = 0.035$, $\eta_p^2 = 0.06$, and for element interactivity, $F(1, 72) = 4.59$, $p = 0.036$, $\eta_p^2 = 0.06$. Overall, the inaccuracy indices were enhanced in the disfluency as well as low element interactivity conditions indicating reduced RC-accuracy. There was a significant interaction, $F(1, 72) = 11.60$, $p = 0.001$, $\eta_p^2 = 0.14$. Disfluency reduced RC-accuracy in the condition with low element interactivity ($p = 0.001$). If element interactivity is high, no significant difference was found regarding disfluency ($p = 1.00$). With respect to the descriptive data, particularly participants in the disfluent—low element interactivity condition was inaccurate in the judgments (see Supplementary Appendix D).

2.3 Discussion

H1a (no main effect for disfluency regarding learning outcomes) was partially supported since Bayes factors indicated evidence for the null-hypothesis, regarding transfer but not retention. Results further indicated no interaction effects of perceptual disfluency and element interactivity for learning outcomes. However, descriptively, disfluency had a positive impact on learning outcomes when element interactivity was low, but it did not affect learning outcomes when element interactivity was high. Therefore, we cannot support H1b (disfluency fosters learning outcomes when the element interactivity is low but impairs learning outcomes when the element interactivity is high). Consistent with H2 (disfluency increases germane processes when the element interactivity is low but impairs germane processes when the element interactivity is high), disfluency increased germane cognitive load (GCL) in the condition of low element interactivity. However, disfluency did not decrease GCL when element interactivity was high, i.e., H2 was partly supported. H3 (increasing the element interactivity leads to increases in ICL) and H4 (increasing the disfluency leads to increases in ECL) were supported, that is, high element interactivity increased ICL, and the use of a disfluent font increased ECL. Contrary to H5 (increasing the disfluency leads to increases in absolute metacognitive accuracy), disfluency reduced absolute metacognitive accuracy as indicted by increased inaccuracy indices. Particularly in the low element interactivity condition, accuracy decreased. The results from Experiment 1 showed that cognitive load plays a significant role in understanding the effects of disfluent learning material. Especially under conditions of significant working memory load, characterized by both high element interactivity and the presence of a disfluent font, efficiency scores decreased as the efficiency was descriptively lowest in this group. While the theoretical implications were partially supported, it is essential to note that further discussion is warranted. First, disfluency did not impair variables relevant to learning in the high-element interactivity condition. Maybe the manipulation of element interactivity was not robust enough in Experiment 1, or the fluency manipulation did not increase the ECL sufficiently to overwhelm learners in the high element interactivity condition. Further, recent research outlined that the disfluency effect might origin from a novelty effect regarding disfluent material (Sung et al., 2022). However, in our opinion, students might be familiar with scanned and printed material (as used in this experiment) since this is still a common learning material during secondary and tertiary education. Consequently, the disfluent material could no foster learning due to the experience of a novel learning scenario.

Additionally, the results concerning metacognitive variables did not align with prior research's theoretical implications. Metacognitive judgments were not significantly affected by the experimental manipulations overall, and the effects on judgments' accuracy were contrary to our expectations. Contrary to prior studies (Pieger et al., 2017), disfluency did not increase the accuracy of judgments; instead, participants judged their learning outcomes fairly accurately for fluent texts, but participants who read disfluent texts under- and overestimated their learning outcomes as indicated by increased inaccuracy indies. One explanation is that the disfluent design of the instructional text suggested that the learning task might be challenging, and thus participants assumed that learning the relevant information would be more difficult than it actually was. According to the CLT (Sweller et al., 2019), cognitive resources are disproportionately allocated to decoding and processing the format rather than understanding the content. This shift in resource allocation could impair the self-monitoring processes necessary for accurate metacognitive judgments, leading participants to underestimate their learning potential despite retaining accurate comprehension. Furthermore, affective processes might play a role despite not being investigated in this study. The feeling of difficulty or frustration when processing disfluent text might influence affective states, which in turn can affect metacognitive judgments. Research by Efklides (2006) underscores that affective responses during learning tasks interact with metacognitive monitoring and control. Participants who experience negative emotions due to text difficulty might be less confident in their learning evaluation, hence skewing metacognitive accuracy.

To verify these results and to gain deeper insights into the processes and outcomes of learning with disfluent materials, we changed the manipulation of disfluency to a different instructional format. Thus, in Experiment 2, the effect of disfluency with concept maps was the focus of research, with the hypotheses remaining the same.

3 Experiment 2

3.1 Methods

3.1.1 Participants and design

The a-priori power analysis was identical to study 1. Overall, 74 students (79.5% female, age: M = 22.50 years, SD = 3.08) from Chemnitz University of Technology and Freiburg University of Education were included in the analyses. Students were enrolled in media communication/psychology (75.3%), teacher education (20.5%), and other programs of study (4.1%). Each participant received either 5€ or a course credit. As expected, prior knowledge of the participants in terms of genetics (M = 3.50, SD = 4.46, possible maximum of 27 points) was low.

The participants were randomly assigned to one condition of a two (disfluency: disfluent concept map vs. fluent concept map) × two (element interactivity: high vs. low) factorial between-subjects design by the online software LimeSurvey (LimeSurvey GmbH, 2020). For the experimental conditions, no significant differences were found for prior knowledge, students' semester, or age, $F(3, 65) = (0.06, 0.99)$, $p = (0.404, 0.981)$ or gender, $\chi^2 = 5.46$, $p = 0.141$. The degrees of freedom of the ANOVA were below 70 since five participants did not report their semester. However, there was a significant difference regarding program of study, $\chi^2 = 12.63$, $p = 0.049$. Program of study was not a significant covariate concerning retention, $F(1, 68) = 0.40$, $p = 0.530$, and transfer, $F(1, 68) = 2.20$, $p = 0.142$. Thus, the program of study was not included as a covariate.

3.1.2 Materials

The learning material consisted of a concept map dealing with genetics. Genetics was chosen because the prior knowledge of participants was considered low, and the content was from the same domain (i.e., biology) as the learning material of Experiment 1. The concept map consisted of multiple subtopics, including the structure of DNA, mutation mechanisms, genetic engineering, protein biosynthesis, and inheritance. In all conditions, the entire concept map was presented to the participants in a non-animated

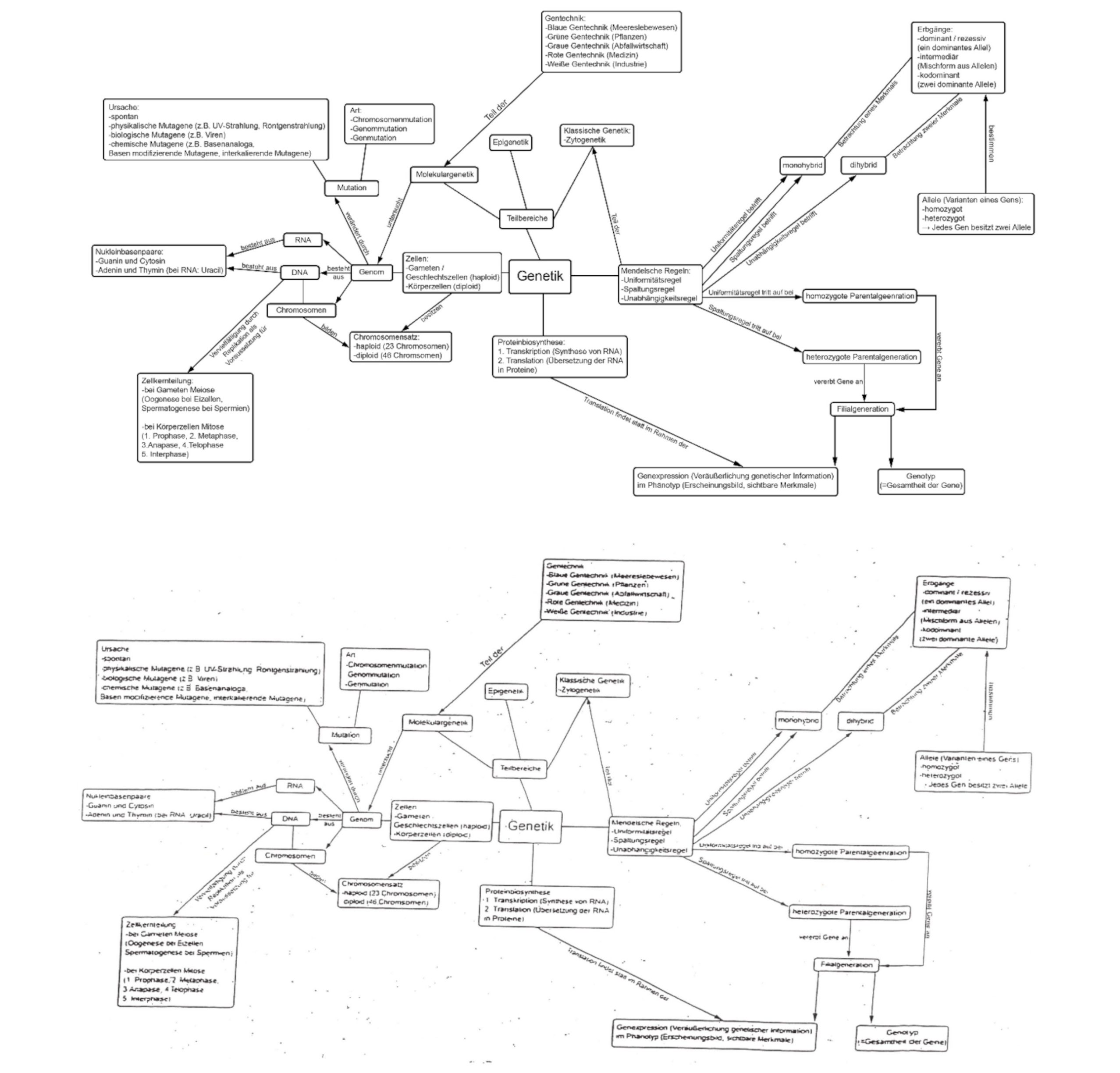


FIGURE 3
Manipulation of disfluency in Experiment 2 (upper panel: fluent, bottom panel: disfluent).

fashion. Learning with the concept map was learner-paced, and the learning time was tracked.

3.1.2.1 Perceptual disfluency

The manipulation of disfluency was similar to Experiment 1. In the fluent condition, the text was presented in the computer-written font Arial. The text was written in black on a neutral white background. To fluent condition into the disfluent condition, the text from the fluent condition was printed, scanned, and repeatedly printed and scanned again. Consequently, readability was low in the disfluent condition. Again, one item was used to assess readability (Likert scale from 1 to 7). The fluent concept map was rated as easy to read (M = 6.35; SD = 2.02) and the disfluent map was rather difficult to read, (M = 4.38; SD = 0.98). The difference was statistically significant, t = 5.35, p < 0.001, d = 1.24 (Welch corrected because of violation of variance homogeneity and normal distribution). All other aspects of the learning content (i.e., number and arrangement of concepts) were kept the same. The fluent vs. disfluent learning material is displayed in Figure 3.

3.1.2.2 Element interactivity

The manipulation was similar to Experiment 1. Whereas we manipulated the amount of information per sentence in Experiment 1, we manipulated the amount of information per node in Experiment 2. In the low-element interactivity condition, subtopics were presented within a few nodes (2.11 elements per nod). Consequently, learners had to focus only on a few nodes to understand each subtopic. In the condition with high element interactivity, more nodes and relations between these nodes were presented (one element per nod), leading to a higher number of

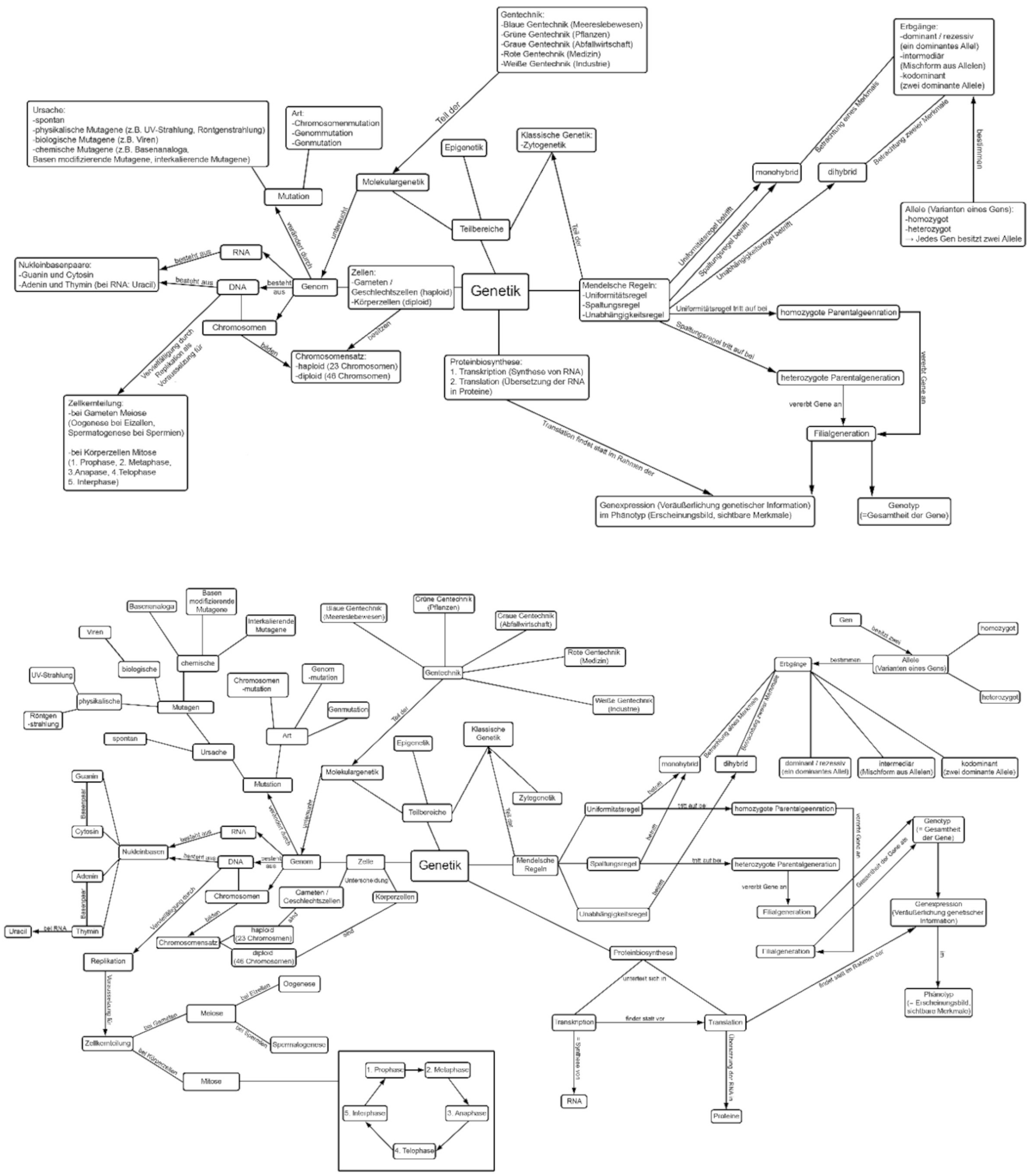


FIGURE 4
Manipulation of element interactivity in Experiment 2 (upper panel: low element interactivity, bottom panel: high element interactivity).

nodes that had to be processed simultaneously to comprehend each subtopic. An example of the manipulation of the node structure is displayed in Figure 4. A translated example can be found in Supplementary Appendix E.

3.1.3 Measures

3.1.3.1 Cognitive load

The scales were the same as in Experiment 1 (ICL: α = 0.75; ϱ = 0.65, ECL: α = 0.76, GCL: α = 0.52; ϱ = 0.32). Reliabilities were good, except for GCL.

3.1.3.2 Metacognitive judgments

The procedure of obtaining metacognitive judgments and metacognitive accuracy was the same as in Experiment 1.

3.1.3.3 Knowledge measures

Prior knowledge, retention, and transfer tests were assessed. Six open questions were asked to assess prior knowledge (α = 0.89). Participants gained up to 27 points. The questions covered different sub-topics (e.g., "What types of genetic engineering are there? How are they characterized?").

The retention test (α = 0.78) consisted of five multiple-choice questions, five open questions (e.g., "Name two different causes for the occurrence of mutagens"), and one mapping task ("Fill in the blanks in this text. Use the scientific terms provided below"). Participants gained a maximum of 43 points in the retention test.

The transfer test (α = 0.84) consisted of four open questions (e.g., With regard to genetics, "why is it important to wear

TABLE 2 Means and standard deviations of dependent variables from Experiment 2.

Dependend varables	Experimental groups							
	Disfluent material				Fluent material			
	High element interactivity (N = 16)		Low element interactivity (N = 21)		High element interactivity (N = 19)		Low element interactivity (N = 18)	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Retention	27.19	8.44	38.29	11.23	29.74	12.16	29.08	10.04
Transfer	15.25	5.64	21.19	8.14	17.72	8.66	15.39	8.24
ICL	5.19	1.26	5.24	1.37	5.87	1.10	5.14	0.68
ECL	5.54	0.97	4.97	1.38	4.32	1.21	3.50	1.28
GCL	4.81	1.14	5.90	0.62	5.16	1.09	4.83	1.10
EOL	67.88	25.24	43.71	31.59	63.00	27.76	55.85	26.26
JOL	34.13	25.13	34.95	23.99	33.00	22.07	44.61	18.74
RC	22.19	14.80	27.24	23.88	29.82	24.93	34.64	20.22
EOL accuracy	0.25	0.19	0.42	0.27	0.26	0.18	0.26	0.17
JOL accuracy	0.28	0.16	0.40	0.17	0.26	0.14	0.20	0.14
RC accuracy	0.32	0.14	0.49	0.20	0.30	0.15	0.22	0.14
Learning time	287.33	65.46	299.65	79.03	222.74	74.32	202.36	42.06
Learning efficiency (retention)	-0.57	0.73	0.01	1.13	0.17	0.81	0.32	0.58
Learning efficiency (transfer)	-0.51	0.88	-0.10	1.21	0.29	0.80	0.27	0.84

For accuracy, higher values indicate an increased discrepancy between confidence judgments and performance. Retention scores ranged from 0 to 43. Transfer scores ranged from 0 to 37. Cognitive Load ratings ranged from 1 to 7. EOL, JOL, and RC ratings ranged from 0 to 100.

a lead apron when taking an X-ray?"), two ranking questions (e.g., "Arrange the following terms in order of size, starting with the smallest unit! "Genome," "nucleic acids," . . ."), and two visualization tasks ("Create a brief visualization of some basic concepts in genetics. Use the concepts displayed below"; "Complete the picture of a DNA strand by adding the complementary nucleic acids"). In total, students gained up to 37 points in the transfer test. For the open questions (prior knowledge, retention, and transfer), intra-class correlation coefficients (ICCs) were satisfactory, ICC (2, k) = (0.92, 0.998), F(73, 73) = (26.13, 952.89), $ps < 0.001$ or perfect (ICC = 1). Again, learning efficiency was calculated.

3.1.4 Procedure

The procedure was the same as in Experiment 1.

3.1.5 Analysis strategy

The analysis strategy was the same as in Experiment 1.

3.2 Results

Means and standard deviations for the dependent variables are presented in Table 2.

3.2.1 Learning outcomes and efficiency

An ANOVA with retention as dependent variable revealed no effect for disfluency, $F(1, 70) = 1.78$, $p = 0.186$, $\eta_p^2 = 0.03$ but for element interactivity, $F(1, 70) = 4.39$, $p = 0.040$, $\eta_p^2 = 0.06$. Low element interactivity led to higher retention scores. There was a significant interaction, $F(1, 70) = 5.56$, $p = 0.021$, $\eta_p^2 = 0.07$. *Post-hoc* tests revealed that disfluency increased retention in the condition with low element interactivity ($p = 0.045$). If element interactivity is high, no significant difference was found regarding disfluency ($p = 1.00$) (see Supplementary Appendix F). Regarding transfer, no effect for disfluency, $F(1, 70) = 0.82$, $p = 0.367$, $\eta_p^2 = 0.01$ and element interactivity was found, $F(1, 70) = 0.97$, $p = 0.329$, $\eta_p^2 = 0.01$. There was a significant interaction, $F(1, 70) = 5.10$, $p = 0.027$, $\eta_p^2 = 0.07$. However, *post-hoc* tests revealed no significant effects between the conditions ($p \geq 0.146$). Bayes factors did not support the null hypotheses regarding disfluency for retention, $BF_{10} = 0.41$, log (BF_{10}) = –0.91, but for transfer, $BF_{10} = 0.23$, log (BF_{10}) = –1.49.

Additionally, an ANOVA of learning time was calculated. ANOVA revealed no effect for element interactivity, $F(1, 70) = 0.07$, $p = 0.799$, $\eta_p^2 < 0.001$, and no interaction effect $F(1, 70) = 1.08$, $p = 0.303$, $\eta_p^2 = 0.02$. However, a significant disfluency effect with a large effect size was found, $F(1, 70) = 26.37$, $p < 0.001$, $\eta_p^2 = 0.27$, i.e., disfluency increased learning time.

An ANOVA with retention efficiency as dependent variable revealed a significant effect for disfluency, $F(1, 70) = 6.91$, $p = 0.011$, $\eta_p^2 = 0.09$. Disfluency reduced efficiency. There was no effect for element interactivity, $F(1, 70) = 3.30$, $p = 0.074$, $\eta_p^2 = 0.05$ and no significant interaction, $F(1, 70) = 1.20$, $p = 0.277$, $\eta_p^2 = 0.02$. Regarding transfer efficiency, a significant effect was found for disfluency, $F(1, 70) = 6.80$, $p = 0.011$, $\eta_p^2 = 0.09$. Disfluency reduced efficiency. There was no effect for element interactivity, $F(1, 70) = 0.75$, $p = 0.390$, $\eta_p^2 = 0.01$ and no significant interaction, $F(1, 70) = 0.91$, $p = 0.342$, $\eta_p^2 = 0.01$. According to the descriptive data, disfluency reduced instructional efficiency. Particularly the group with high element interactivity and disfluent learning material showed reduced efficiency in contrast to the other conditions (see Supplementary Appendix F).

3.2.2 Perceived cognitive load

An ANOVA with ICL as dependent variable revealed no effect for disfluency, $F(1, 70) = 1.19$, $p = 0.279$, $\eta_p^2 = 0.02$, no effect for element interactivity, $F(1, 70) = 1.62$, $p = 0.207$, $\eta_p^2 = 0.02$, and no significant interaction, $F(1, 70) = 2.14$, $p = 0.148$, $\eta_p^2 = 0.03$. An ANOVA with ECL as dependent variable revealed a significant effect for disfluency, $F(1, 70) = 21.84$, $p < 0.001$, $\eta_p^2 = 0.24$ and a significant effect for element interactivity, $F(1, 70) = 5.81$, $p = 0.019$, $\eta_p^2 = 0.08$. ECL was enhanced in the disfluent conditions as well as in the conditions with high element interactivity. There was no significant interaction, $F(1, 70) = 0.18$, $p = 0.675$, $\eta_p^2 = 0.003$. Regarding GCL, there was no significant effect for disfluency, $F(1, 70) = 2.44$, $p = 0.123$, $\eta_p^2 = 0.03$ and no effect for element interactivity, $F(1, 70) = 2.73$, $p = 0.103$, $\eta_p^2 = 0.04$. There was a significant interaction, $F(1, 70) = 9.29$, $p = 0.003$, $\eta_p^2 = 0.12$. *Post-hoc* tests revealed that disfluency increased GCL in the condition with low element interactivity ($p = 0.008$). If element interactivity is high, no significant difference was found regarding disfluency ($p = 0.929$) (see Supplementary Appendix G).

3.2.3 Metacognitive judgments and accuracy

An ANOVA with EOL judgment as dependent variable revealed no significant effect for disfluency, $F(1, 70) = 0.31$, $p = 0.58$, $\eta_p^2 = 0.004$ but an effect for element interactivity, $F(1, 70) = 5.71$, $p = 0.020$, $\eta_p^2 = 0.08$. EOL was reduced in the low element interactivity conditions. There was no significant interaction, $F(1, 70) = 1.68$, $p = 0.199$, $\eta_p^2 = 0.02$. Regarding the JOL, there was no effect for disfluency, $F(1, 70) = 0.65$, $p = 0.422$, $\eta_p^2 = 0.01$, element interactivity, $F(1, 70) = 1.39$, $p = 0.243$, $\eta_p^2 = 0.02$, and no interaction, $F(1, 70) = 1.04$, $p = 0.310$, $\eta_p^2 = 0.02$. Regarding RC judgments (averaged for retention and transfer), there was no effect for disfluency, $F(1, 70) = 2.21$, $p = 0.142$, $\eta_p^2 = 0.03$, element interactivity, $F(1, 70) = 0.95$, $p = 0.333$, $\eta_p^2 = 0.01$, and no interaction, $F(1, 70) < 0.001$, $p = 0.982$, $\eta_p^2 < 0.001$.

Regarding EOL-accuracy, there was no significant effect for disfluency, $F(1, 70) = 2.16$, $p = 0.146$, $\eta_p^2 = 0.03$, no effect for element interactivity, $F(1, 70) = 3.28$, $p = 0.075$, $\eta_p^2 = 0.05$, and no significant interaction, $F(1, 70) = 3.06$, $p = 0.085$, $\eta_p^2 = 0.04$. Regarding JOL-accuracy, there was a significant effect regarding disfluency, $F(1, 70) = 8.75$, $p = 0.004$, $\eta_p^2 = 0.11$, but no effect for element interactivity, $F(1, 70) = 0.68$, $p = 0.411$, $\eta_p^2 = 0.01$. Overall, the inaccuracy indices were enhanced in the disfluency conditions indicating reduced JOL-accuracy. There was a significant interaction, $F(1, 70) = 6.46$, $p = 0.013$, $\eta_p^2 = 0.09$. *Post-hoc* test revealed that disfluency reduced JOL-accuracy in the condition with low element interactivity ($p < 0.001$). If element interactivity is high, no significant difference was found regarding disfluency ($p = 0.78$) (see Supplementary Appendix H). Regarding RC-accuracy, there was a significant effect regarding disfluency, $F(1, 70) = 15.09$, $p < 0.001$, $\eta_p^2 = 0.18$, but not for element interactivity, $F(1, 70) = 1.56$, $p = 0.216$, $\eta_p^2 = 0.02$. The inaccuracy indices were enhanced in the disfluency conditions indicating reduced RC-accuracy. There was a significant interaction, $F(1, 70) = 11.19$, $p = 0.001$, $\eta_p^2 = 0.14$. Disfluency reduced RC-accuracy in the condition with low element interactivity ($p = 0.001$). If element interactivity is high, no significant difference was found regarding disfluency ($p = 0.712$). Looking at the descriptive data, particularly participants in the disfluent—low element interactivity condition were inaccurate in their judgments (see Supplementary Appendix H).

3.3 Discussion

Results of Experiment 1 and Experiment 2 were largely similar. Again, Bayes factors supported H1a (no main effect for disfluency regarding learning outcomes) in terms of transfer but not retention. Disfluency had a positive effect on learning outcomes when element interactivity was low. In contrast, when element interactivity was high, disfluency did not significantly affect learners' test outcomes. Consequently, H1b (disfluency fosters learning outcomes when the element interactivity is low but impairs learning outcomes when the element interactivity is high) received partial support in Experiment 2. A similar pattern emerged for GCL, thereby partially supporting H2 (disfluency increases germane processes when the element interactivity is low but impairs germane processes when the element interactivity is high) as well, i.e., disfluency increased GCL in the low element interactivity condition but not in the high element interactivity condition. Furthermore, the results supported H4 (increasing the disfluency leads to increases in ECL) since using a disfluent font elevated ECL. However, H5 (increasing the disfluency leads to increases in absolute metacognitive accuracy) was not supported. Disfluency lowered accuracy, particularly in the low element interactivity condition: Contrary to our hypothesis, for low element interactivity, disfluency decreased the accuracy of the JOL and the accuracy of the RC judgments. Furthermore, the findings related to efficiency were consistent with those of Experiment 1. Consequently, the same explanations as those applied in Experiment 1 can also be applied here. Concerning H3 (increasing the element interactivity leads to increases in ICL), the two experiments yielded different results. An effect of element interactivity on ICL was found in Experiment 1 but not in Experiment 2. Increasing element interactivity by breaking down individual nodes in a concept map into multiple concept maps may not have been sufficient to increase the subjective perception of content complexity. The two ICL items asked for the complexity of the learning material and for the number of concepts that have to be simultaneously held in memory. Thus, breaking down complex nodes into several nodes, each focusing on only one concept, seems less associated to ICL to the participants. Further, the variation of element interactivity affected ECL in Experiment 2 but bot in Experiment 1. A potential explanation can be derived considering the work from Costley et al. (2025). According to the authors, connections between nodes in a concept map reduced ECL but did not affect ICL or GCL. The interconnections acted as structuring elements that facilitated learning by organizing knowledge. Similarly, a concept map with information consolidated into fewer nodes, containing a larger amount of information, reduced ECL since related information is already structured meaningfully. This might lead to a clearer organization of the material, allowing learners to grasp overarching concepts. In the high element interactivity condition, the distribution was too fine-grained; learners might have to navigate excessively between numerous nodes, which increased ECL. Consequently, varying text structure was rather associated with ICL and nod structure was rather associated with ECL. Nevertheless, since most other results closely resembled those of Experiment 1, it is reasonable to assume that the manipulation was still successful to a certain extent.

4 Mini meta-analysis

4.1 Methods

Despite sample planning based on power analysis, the sample size is of course quite small overall. Thus, a mini meta-analysis (Goh et al., 2016) was conducted to summarize the results of both studies. The disfluency effects and the interaction effect regarding retention and transfer scores were the focus of the analyses. The effect size of the group difference between fluency and disfluency for learning outcomes is included in the calculation separately for high and low element interactivity. This is done for both experiments, so that a total of four effect sizes are included in the calculation. For the main effect of disfluency, the effect sizes are included in the calculation in such a way that higher effect sizes generally mean better learning results for disfluency. For the interaction effect, positive effect sizes mean that disfluency improves the learning outcome when element interactivity is low and worsens it when element interactivity is high.

A small sample-adjusted standardized mean difference (Hedges' g^* for aggregated effect sizes, e.g., Hedges and Olkin, 1985) was chosen as the standard effect size. The effect sizes of all pairwise comparisons were computed using the means and standard deviations reported in the two experiments. A random-effects model (restricted maximum likelihood estimation) was used. This approach is based on Field and Gillett (2010), who recommended a random-effects model in social sciences. Each computed effect size was standardized by the inverse squared standard error (e.g., Cooper et al., 2009).

4.2 Results

Table 3 outlines the effects of disfluency and interaction effects on retention and transfer.

Results revealed no disfluency effect on retention, $g+ = 0.43$, $z = 1.29$, $p = 0.198$ ($I^2 = 52.29\%$, $Q = 6.35$, $p = 0.096$) as well as transfer, $g+ = 0.18$, $z = 0.81$, $p = 0.416$ ($I^2 = 24.61\%$, $Q = 4.16$, $p = 0.244$). However, a significant moderation was observed for retention, $g+ = 0.61$, $z = 2.32$, $p = 0.020$ ($I^2 = 22.35\%$, $Q = 3.51$, $p = 0.320$) and transfer, $g+ = 0.38$, $z = 2.01$, $p = 0.045$ ($Q = 1.09$, $p = 0.780$). As indicated by the heterogeneity statistics, heterogeneity between the two Experiments was low (transfer, interaction effects) up to moderate (retention).

TABLE 3 Aggregated effect sizes and confidence intervals for learning outcomes across Experiment 1 and 2.

Outcome measure	Effect size $g+$	Standard error	95% CI for $g+$
Disfluency effect			
Retention	0.43	0.34	(−0.23, 1.09)
Transfer	0.18	0.22	(−0.25, 0.61)
Moderating effect of element interactivity			
Retention	0.61*	0.26	(0.09, 1.12)
Transfer	0.38*	0.19	(0.01, 0.75)

*$p < 0.05$.

5 General discussion

5.1 Implications

As expected in H1a, we did not find any evidence for a generalized disfluency effect with regard to deeper processing, i.e., transfer performance. We further did not find any significant main effects of perceptual disfluency on learning outcomes according to the small-scale meta-analysis. These findings align with previous studies and meta-analyses that were also skeptical about the disfluency effect or failed to replicate it (e.g., İlic and Akbulut, 2019; Rummer et al., 2016; Xie et al., 2018). Note that we are not challenging the arguments by Kahneman and Frederick (2002) and the elaborations by Alter et al. (2007), Alter and Oppenheimer (2009). The fundamental concept of two processing systems that are activated under different conditions is supported by extensive empirical evidence, and specific triggers or conscious effort are required to initiate System Two processing (e.g., the use of heuristics vs. conscious processing; Strack and Deutsch, 2002).

One possible trigger is perceptual disfluency. However, our results show that a disfluent font may not necessarily serve as an effective trigger for more elaborate processing in general. Our results indicate that an illegible font increased ECL and thus may potentially hinder the learning process. But how can we explain the disfluency effects found in some studies (e.g., Eitel and Kühl, 2016; Seufert et al., 2017)? We argue that two processes may occur simultaneously: an inhibiting effect due to increased ECL (Sweller et al., 2019) and a facilitating effect due to more elaborate processing via metacognitive processes (Alter et al., 2007). Among other moderators, element interactivity (i.e., ICL) may determine which of the two processes prevails, i.e., GCL induced by disfluency may be detrimental to learning processes if ICL is high but beneficial if ICL is low.

Geller et al. (2018) and Seufert et al. (2017) previously showed that the degree of perceptual disfluency (i.e., the degree of ECL) is crucial. Inducing a small amount of ECL can trigger deeper processing, but if ECL reaches a certain threshold, it will inhibit learning. This argument is supported by findings that increasing the load on working memory decreases the accuracy of metacognitive monitoring when solving working memory capacity tasks (Bryce et al., 2023); that is, if the load on working memory (e.g., induced by disfluency) gets too high, metacognitive accuracy will suffer, and the detrimental effect of disfluency will prevail.

The results from our experiments supported the twofold effect of disfluency, as the complexity of the learning content was found to be a moderator in Experiment 2 and in the mini meta-analysis. In line with our reasoning, disfluency increased ECL in both experiments, i.e., disfluency imposed a load on working memory. When the element interactivity is high, learners may struggle to process the content deeply because substantial cognitive resources are required merely to handle the content itself. Therefore, any additional ECL induced by disfluency counteracts potential positive effects and hinders learning. However, when the element interactivity is low, students may have the cognitive capacity to engage in metacognitive processing and deeper processing of the learning content even in the presence of ECL caused by disfluency. This empirical finding aligns with results previously reported by Beege et al. (2021).

This leads to another theoretical and methodological contribution of this paper, namely the manipulation of element interactivity. In part, we have followed Chen et al. (2024) and counted elements per unit of information to determine element interactivity. However, varying this without changing the content of the material, but rather through the structure of the information presentation, is at least a new approach that could be pursued in future research. Of course, this depends heavily on the theoretical conception of element interactivity, and current definitions of ICL and element interactivity also give the impression that Sweller's (2010) arguments have been received or discussed to a lesser extent to date.

From a practical perspective, we cannot generally recommend the use of disfluent fonts in instructional settings. While disfluency can be employed in instructional settings with simple and undemanding material to encourage learners to invest more effort and avoid superficial skimming, such situations are infrequent in educational settings. There might be situations where learners have high prior knowledge or strong cognitive abilities (resulting in lower element interactivity), but more often, the learning content is new to the learners and perceived to be complex. Therefore, using disfluent fonts risks demanding too much from learners and overburdening them. Especially when learners revisit learning materials after a certain period, there is a risk of disorientation and confusion since they have to review the disfluent material again, and a quick overview might not be possible. Processing disfluent materials can be time-consuming (i.e., disfluency increased learning time significantly in both experiments) and may be perceived as aversive by learners. Therefore, poor printouts, low-quality scans, or multiple copies can lead to disfluent materials and pose limitations for effective learning, at least for learning materials of high complexity. Alternatively, interventions like summary writing, keyword taking, diagram completion, concept mapping, rereading, and the announcement of a comprehension test can be recommended to increase the accuracy of metacognitive monitoring (Gutierrez and de Blume, 2022; Prinz et al., 2020).

5.2 Limitations and future directions

The first limitation is the rather small sample size of the studies (N_1 = 76; N_2 = 74). As a result, non-significant findings with lower effect sizes can be attributed to low power. However, the non-significant effect sizes were very small (with the exception of possible interactions), so non-significant effects have little practical relevance even with larger samples. Nonetheless, future studies could benefit from larger participant numbers.

A second limitation pertains to the manipulation of element interactivity. We based our manipulation on Sweller's (2010) definition, but element interactivity was manipulated without altering the actual learning content. This manipulation maintained equivalent learning content across experimental conditions; however, the manipulation deviates from the original definition of element interactivity as an inherent aspect of ICL. Modifying ICL by changes in the learning content (as discussed by Beege et al., 2021) would confound ICL with changes in the information provided to the learners. Thus, our approach provides a practical means to investigate the moderating effect of element interactivity on the efficacy of disfluency. Further, manipulating element interactivity by altering the sentence structure might have effects on perceived text cohesion. Text cohesion refers to the linguistic elements that link sentences and paragraphs together to create a coherent whole, facilitating the reader's understanding and processing of information. According to McNamara and Graesser (2005), cohesive texts use devices such as referential ties, conjunctions, and lexical repetition to establish connections within the text. However, presenting instructional text as a series of simple sentences, while seemingly straightforward, can potentially undermine text coherence. This approach may lead to fragmented information and disrupt the reader's mental model of the subject, as coherence relies on the ability to integrate information across sentences (Graesser et al., 2003).

Lastly, research on disfluency may hold diminishing potential for new insights, particularly following the somewhat discouraging results of the meta-analysis by Xie et al. (2018), which led to a decline in research interest. Nevertheless, there are still new findings about specific forms of disfluency (Ebersbach et al., 2023), effects of learner characteristics (Astley et al., 2023), or conditions for disfluency effects (Lai and Zhang, 2021). Thus, there remain some open questions to be answered by future research, e.g., exploring additional factors that may moderate the disfluency effect. Nevertheless, it will be difficult to convince teachers or media designers of the rather limited practical implications.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://doi.org/10.17605/OSF.IO/PNW5E.

Ethics statement

Ethical approval was not required for the studies involving humans because all procedures were performed in full accordance with the ethical guidelines of the German Psychological Society and the APA Ethics code. Since the experiments constitute non-medical low risk research, no special permission by an ethics committee is required for psychological research in the Institute for Psychology of the Freiburg University of Education as well as in Germany. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MB: Validation, Investigation, Supervision, Writing – review & editing, Conceptualization, Formal analysis, Writing – original draft, Project administration, Data curation, Methodology. CM: Writing – review & editing, Writing – original draft, Formal analysis. SN: Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. Open Access funding enabled and organized by Projekt DEAL. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2025.1731080/full#supplementary-material

References

Alter, A. L., and Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personal. Soc. Psychol. Rev.* 13, 219–235. doi: 10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., and Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *J. Exp. Psychol. General* 136, 569–576. doi: 10.1037/0096-3445.136.4.569

Astley, J., Keage, H. A., Kelson, E., Callahan, R., Hofmann, J., Thiessen, M., et al. (2023). Font disfluency and reading performance in children: An event-related potential study. *Brain Cogn.* 169:105986. doi: 10.1016/j.bandc.2023.105986

Ball, B. H., Klein, K. N., and Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *J. Exp. Psychol. Appl.* 20, 336–348. doi: 10.1037/xap0000023

Beckmann, J. (2010). Taming a beast of burden—On some issues with the conceptualisation and operationalisation of cognitive load. *Learn. Instruct.* 20, 250–264. doi: 10.1016/j.learninstruc.2009.02.024

Beege, M., Krieglstein, F., Schneider, S., Nebel, S., and Rey, G. D. (2021). Is There a (Dis-) fluency effect in learning with handwritten instructional texts? Evidence from three studies. *Front. Educ.* 6:678798. doi: 10.3389/feduc.2021.678798

Bryce, D., Kattner, F., Birngruber, T., and Wellingerhof, P. (2023). Monitoring accuracy suffers when working memory demands increase: Evidence of a dependent relationship. *J. Exp. Psychol. Learn. Mem. Cogn.* 49, 1909–1922. doi: 10.1037/xlm0001262

Burton, R. F. (2005). Multiple-choice and true/false tests: Myths and misapprehensions. *Assess. Eval. Higher Educ.* 30, 65–72. doi: 10.1080/0260293042003243904

Chen, O., Chan, B. K. Y., Anderson, E., O'sullivan, R., Jay, T., Ouwehand, K., et al. (2024). The effect of element interactivity and mental rehearsal on working memory resource depletion and the spacing effect. *Contemp. Educ. Psychol.* 77:102281. doi: 10.1016/j.cedpsych.2024.102281

Chen, O., Kalyuga, S., and Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *J. Educ. Psychol.* 107, 689–704. doi: 10.1037/edu0000018

Chen, O., Kalyuga, S., and Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educ. Psychol. Rev.* 29, 393–405. doi: 10.1007/s10648-016-9359-1

Cooper, H., Hedges, L. V., and Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage Foundation.

Costley, J., Kapuza, A., Gorbunova, A., and Shcheglova, I. (2025). How adding structure reduces complexity: More interconnections in concept maps do not increase cognitive load. *Instruct. Sci.* 53, 1243–1262. doi: 10.1007/s11251-025-09736-5

de Bruin, A. B. H., Roelle, J., Carpenter, S. K., Baars, M., and Efg, M. R. E. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educ. Psychol. Rev.* 32, 903–915. doi: 10.1007/s10648-020-09576-4

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instruct. Sci.* 38, 105–134. doi: 10.1007/s11251-009-9110-0

Diemand-Yauman, C., Oppenheimer, D. M., and Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition* 118, 111–115. doi: 10.1016/j.cognition.2010.09.012

Dunlosky, J., and Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Curr. Direct. Psychol. Sci.* 16, 228–232. doi: 10.1111/j.1467-8721.2007.00509.x

Ebersbach, M., Guschlbauer, J. A., and Rummer, R. (2023). Flickering presentations do affect the judgment of learning but not the learning outcome. *Appl. Cogn. Psychol.* 37, 889–898. doi: 10.1002/acp.4086

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educ. Res. Rev.* 1, 3–14. doi: 10.1016/j.edurev.2005.11.001

Eitel, A., and Kühl, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacogn. Learn.* 11, 107–121. doi: 10.1007/s11409-015-9145-3

Eitel, A., Kühl, T., Scheiter, K., and Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: Does harder-to-read mean better-to-understand? *Appl. Cogn. Psychol.* 28, 488–501. doi: 10.1002/acp.3004

Faber, M., Mills, C., Kopp, K., and D'Mello, S. (2017). The effect of disfluency on mind wandering during text comprehension. *Psychon. Bull. Rev.* 24, 914–919. doi: 10.3758/s13423-016-1153-z

Field, A. P., and Gillett, R. (2010). How to do a meta-analysis. *Br. J. Mathemat. Statist. Psychol.* 63, 665–694. doi: 10.1348/000711010X502733

Geller, J., Still, M. L., Dark, V. J., and Carpenter, S. K. (2018). Would disfluency by any other name still be disfluent? Examining the disfluency effect with cursive handwriting. *Mem. Cogn.* 46, 1109–1126. doi: 10.31274/etd-180810-5137

Goh, J. X., Hall, J. A., and Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Soc. Personal. Psychol. Comp.* 10, 535–549. doi: 10.1111/spc3.12267

Gorbunova, A., Adamovich, K., Savelyev, A., and Costley, J. (2025). Balancing autonomy and efficiency: How learner control shapes cognitive load and online learning outcomes. *Interact. Learn. Environ.* 1–22. doi: 10.1080/10494820.2025.2547273

Graesser, A. C., McNamara, D. S., and Louwerse, M. M. (2003). "What do readers need to learn in order to process coherence relations in narrative and expository text?," in *Rethinking reading comprehension*, eds A. P. Sweet and C. E. Snow (New York, NY: Guilford Press), 82–98.

Griffin, T. D., Wiley, J., and Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Mem. Cogn.* 36, 93–103. doi: 10.3758/mc.36.1.93

Gutierrez, and de Blume, A. P. (2022). Calibrating calibration: A meta-analysis of learning strategy instruction interventions to improve metacognitive monitoring accuracy. *J. Educ. Psychol.* 114, 681–700. doi: 10.1037/edu0000674

Hedges, L. V., and Olkin, I. (1985). *Statistical methods for meta-analysis*. Cambridge, MA: Academic Press.

Huang, F. L. (2020). MANOVA: A procedure whose time has passed? *Gifted Child Quar.* 64, 56–60. doi: 10.1177/0016986219887200

İlic, U., and Akbulut, Y. (2019). Effect of disfluency on learning outcomes, metacognitive judgments and cognitive load in computer assisted learning environments. *Comp. Hum. Behav.* 99, 310–321. doi: 10.1016/j.chb.2019.06.001

James, W. D. (1950). *The principles of psychology*. Downers Grove, IL: Dover.

JASP Team (2025). *JASP (Version 0.95.3)[Computer software]*. Netherlands: JASP Team.

Jiang, D., and Kalyuga, S. (2020). Confirmatory factor analysis of cognitive load ratings supports a two-factor model. *Tutor. Quan. Methods Psychol.* 16, 216–225. doi: 10.20982/tqmp.16.3.p216

Kahneman, D., and Frederick, S. (2002). "Representativeness revisited: Attribute substitution in intuitive judgment," in *Heuristics and biases*, eds T. Gilovich, D. Griffin, and D. Kahneman (Cambridge, MA: Cambridge University Press), 49–81.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educ. Psychol. Rev.* 23, 1–19. doi: 10.1007/s10648-010-9150-7

Klepsch, M., Schmitz, F., and Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* 8:1997. doi: 10.3389/fpsyg.2017.01997

Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropractic Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012

Lai, T., and Zhang, J. (2021). The influence of signaling on the disfluency effect in multimedia learning. *Front. Psychol.* 12:755804. doi: 10.3389/fpsyg.2021.755804

Lehmann, J., Goussios, C., and Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacogn. Learn.* 11, 89–105. doi: 10.1007/s11409-015-9149-z

LimeSurvey GmbH (2020). LimeSurvey: An open source survey tool. *LimeSurvey GmbH*. Available online at: https://www.limesurvey.org

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning*, 2nd Edn. Cambridge, MA: Cambridge University Press.

McNamara, D. S., and Graesser, A. C. (2005). "Coh-Metrix: An automated tool for theoretical and applied natural language processing," in *Handbook of latent semantic analysis*, eds T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch (Mahwah, NJ: Lawrence Erlbaum Associates), 219–241.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *J. Exp. Psychol. General* 131, 349–363. doi: 10.1037/0096-3445.131.3.349

Mulligan, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 54–69. doi: 10.1037/0278-7393.25.1.54

Nelson, T. O., and Narens, L. (1990). "Metamemory: A theoretical framework and new findings," in *The psychology of learning and motivation*, ed. G. H. Bower (Academic Press), 125–173.

Paas, F. G. W. C., and Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Hum. Fact.* 35, 737–743. doi: 10.1177/001872089303500412

Paas, F., and Sweller, J. (2014). "Implications of cognitive load theory," in *The Cambridge handbook of multimedia learning*, ed. R. E. Mayer (Cambridge, MA: Cambridge University Press), 27–43.

Paas, F., Tuovinen, J. E., Tabbers, H., and Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* 38, 63–71. doi: 10.1207/s15326985ep3801_8

Pashler, H., Rohrer, D., Cepeda, N. J., and Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychon. Bull. Rev.* 14, 187–193. doi: 10.3758/BF03194050

Pieger, E., Mengelkamp, C., and Bannert, M. (2017). Fostering analytic metacognitive processes and reducing overconfidence by disfluency: The role of contrast effects. *Appl. Cogn. Psychol.* 31, 291–301. doi: 10.1002/acp.3326

Prinz, A., Golke, S., and Wittwer, J. (2020). To what extent do situation-model-approach interventions improve relative metacomprehension accuracy? Meta-analytic insights. *Educ. Psychol. Rev.* 32, 917–949. doi: 10.1007/s10648-020-09558-6

Rummer, R., Schweppe, J., and Schwede, A. (2016). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacogn. Learn.* 11, 57–70. doi: 10.1007/s11409-015-9151-5

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacogn. Learn.* 4, 33–45. doi: 10.1007/s11409-008-9031-3

Seufert, T., Wagner, F., and Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instruct. Sci.* 45, 221–238. doi: 10.1007/s11251-016-9387-8

Strack, F., and Deutsch, R. (2002). "Urteilsheuristiken," in *Theorien der sozialpsychologie: Band III: Motivations-, Selbst- und informationsverarbeitungstheorien*, eds D. Frey and M. Irle (Atlanta, GA: Huber).

Sung, B., Vanman, E. J., and Hartley, N. (2022). Revisiting (dis) fluency: Metacognitive difficulty as a novelty cue that evokes feeling-of-interest. *Psychol. Market.* 39, 1451–1466. doi: 10.1002/mar.21664

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cogn. Sci.* 12, 257–285. doi: 10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learn. Instruct.* 4, 295–312. doi: 10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* 22, 123–138. doi: 10.1007/s10648-010-9128-5

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive load theory*. Berlin: Springer.

Sweller, J., van Merriënboer, J. J., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 1–32. doi: 10.1007/s10648-019-09465-5

Sweller, J., Van Merrienboer, J. J., and Paas, F. G. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Thalmann, M., Souza, A. S., and Oberauer, K. (2019). How does chunking help working memory? *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 37–55. doi: 10.1037/xlm0000578

Tindall-Ford, S., Chandler, P., and Sweller, J. (1997). When two sensory modes are better than one. *J. Exp. Psychol. Appl.* 3, 257–287. doi: 10.1037/1076-898x.3.4.257

van Gog, T., and Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educ. Psychol.* 43, 16–26. doi: 10.1080/00461520701756248

Weissgerber, S. C., Brunmair, M., and Rummer, R. (2021). Null and void? Errors in meta-analysis on perceptual disfluency and recommendations to improve meta-analytical reproducibility. *Educ. Psychol. Rev.* 33, 1221–1247. doi: 10.1007/s10648-020-09579-1

Weissgerber, S. C., Terhorst, D. I. P., and Rummer, R. (2023). Is the perceptual disfluency effect moderated by working memory capacity? Direct replication of Lehmann et al. (2016). *Metacogn. Learn.* 19, 293–318. doi: 10.1007/s11409-023-09366-7

Xie, H., Zhou, Z., and Liu, Q. (2018). Null effects of perceptual disfluency on learning outcomes in a text-based educational context: A meta-analysis. *Educ. Psychol. Rev.* 30, 745–771. doi: 10.1007/s10648-018-9442-x

Yang, C., Zhao, W., Yuan, B., Luo, L., and Shanks, D. R. (2023). Mind the gap between comprehension and metacomprehension: Meta-analysis of metacomprehension accuracy and intervention effectiveness. *Rev. Educ. Res.* 93, 143–194. doi: 10.3102/00346543221094083

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