Computers in Human Behavior 99 (2019) 310–321

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Effect of disfluency on learning outcomes, metacognitive judgments and cognitive load in computer assisted learning environments

Ulaş İlic[a,*], Yavuz Akbulut[b]

[a] Department of Computer Education & Instructional Technology, Faculty of Education, Pamukkale University, Kınıklı Campus, Denizli, Turkey
[b] Department of Educational Sciences, Faculty of Education, Anadolu University,, Yunusemre Campus, 26470, Eskisehir, Turkey

ARTICLE INFO

Keywords:
Disfluency effect
Fluency
Learning outcomes
Metacognitive judgments
Cognitive load

ABSTRACT

Cognitive challenges that are presented through the modification of established design principles may contribute to learning. One such challenge to the promotion of deeper processing is the Disfluency Effect. Specifically, disfluency manipulations in learning materials interfere with the perceptional fluency, which may in turn lead to better learning outcomes. This likelihood of reaching better learning outcomes through minor instructional modifications has led scholars to investigate the construct further. Accordingly, the effect of fluency modifications on learning outcomes, metacognitive judgments and cognitive load is investigated in the current study with a true experiment conducted with 292 undergraduate students who were assigned randomly to one of the six disfluency scenarios in a computer-based learning environment. Additional variables were also considered such as the working memory capacity, prior knowledge and cognitive load. Significant differences were observed with regard to the extraneous cognitive load, while there was no variation in the learning outcomes and metacognitive judgments in the experimental groups. Moreover, significant relationships were observed between cognitive load, the number of animations watched by the students and the learning outcomes. The results were interpreted in accordance with the approach in contemporary studies into the Disfluency Effect, and theoretical and practical implications are discussed.

1. Introduction

The contemporary opportunities provided by online connectivity facilitate individuals in carrying out many activities through digital devices. This has resulted in a situation in which individuals overlook their own cognitive abilities when carrying out simple operations, and to lose the ability to apply skills that cannot be delegated to machines (Alter, 2013). In order to overcome such a situation, it is important to face with cognitive difficulties for an active engagement. Such cognitive difficulties can be helpful also in instructional settings (Bjork, 1994), that is, they can be useful in breaking the study routines of the individual, varying the predictable nature of the instructional material and creating an unusual instructional pattern for effective learning (Bjork & Bjork, 2011).

Challenges that slow down the learning process can revive long-term memory and improve the transfer of learning (Bjork & Bjork, 2011). One of the cognitive challenges is the Disfluency Effect, which is created through the interruption of fluency. According to Oppenheimer (2008) fluency is “the subjective experience of ease or difficulty associated with completing a mental task” (p.237). Fluency is not regarded as a separate cognitive process, but rather a facilitator of the cognitive processes. For instance, a frequently or recently viewed object is regarded as fluent, although there are several fluency types, such as perceptual, linguistic and memory-based (Alter & Oppenheimer, 2009b). Disfluency, on the other hand, can be defined as the subjective and metacognitive difficulties associated with cognitive tasks (Diemand-Yauman, Oppenheimer, & Vaughan, 2011). Disfluency is associated primarily with perceptual fluency, which is implemented through font manipulations (Alter & Oppenheimer, 2008b; 2009a). For instance, some scholars changed font type (French et al., 2013; Song & Schwarz, 2008) whereas some manipulated the font size (Rhodes & Castel, 2008; Strukelj, Scheiter, Nyström, & Holmqvist, 2016).

The theoretical basis of the Disfluency Effect is derived from James (1950) assumption that people have two different processing systems: System 1 and System 2. While System 1 is effortless, quick, associative and intuitive, System 2 is slow, analytical and deliberate, which requires effort. More specifically, learners monitor challenges before initiating the relevant control mechanisms (Serra & Metcalfe, 2009). If the learning process appears to be easy, learners feel self-confident and devote little effort, as they believe they will quickly reach the desired

---

[*] Corresponding author.
E-mail addresses: uilic@pau.edu.tr (U. İlic), yavuzakbulut@anadolu.edu.tr (Y. Akbulut).

https://doi.org/10.1016/j.chb.2019.06.001
Received 1 March 2019; Received in revised form 10 May 2019; Accepted 5 June 2019

outcome. On the other hand, when they face difficulties, they feel less confident of reaching the desired outcome, and thus activate their controlling activities to make more effort. According to the Disfluency Effect, high perceptual fluency activates System 1, whereas low perceptual fluency (i.e., disfluency) leads to the operation of System 2 (Kühl & Eitel, 2016). Accordingly, effortless activities that trigger System 1 are regarded as fluent, whereas cognitively demanding activities that trigger System 2 are regarded as disfluent (Alter & Oppenheimer, 2009b).

Cognitive interventions appear to have potential in improving the effectiveness of instruction (Diemand-Yauman et al., 2011). For instance, minor modifications associated with disfluency (e.g., font changes) can contribute to learning outcomes (Diemand-Yauman et al., 2011; Eitel, Kühl, Scheiter, & Gerjets, 2014; Faber, Mills, Kopp, & D'mello, 2017; French et al., 2013; Lehmann, Goussios, & Seufert, 2015; Sungkhasettee, Friedman, & Castel, 2011; Weissgerber & Reinhard, 2017). In addition, none of these perceivably useful manipulations require major changes to either the curriculum or the method of instruction, and therefore, even prominent metacognition scholars as Dunlosky and Mueller (2016), who are skeptical of the effect, agree on the significant potential pertaining to disfluency. That said, empirical studies are yet to come up with convincing evidence on the affordances of the effect, and so it is important to investigate the role of disfluency in contemporary computer-assisted learning environments that are commonly used in educational practices. Accordingly, theoretical justifications and critical variables to consider are discussed in the next section.

2. Theoretical framework

2.1. Computer assisted learning and animations

The employment of computers during educational activities is referred to as computer-assisted learning (CAL) (Ornstein & Levine, 1993; Owusu, Monney, Appiah, & Wilmot, 2010), and their prevalence in educational settings is increasing due to many affordances, such as individual learning, instant feedback, interactivity, and freedom of time and space (Eggen & Kauchak, 2007). However, creating relevant instructional designs in CAL environments requires professional competence along with a high labor and hardware cost (Liao, 2007), making the proposal of empirically-supported manipulations to increase instructional effectiveness important.

Several forms of media can be used in CAL, such as text, visuals, sound and animation. For instance, animations can be quite useful for conveying abstract concepts (Tversky, Morrison, & Betrancourt, 2002) and can have a higher instructional potential than static images (Castro-Alonso, Ayres, & Paas, 2015; Höffler & Leutner, 2007; Ryoo & Linn, 2012). Through animations, learners can generate relevant cognitive schemas easily (Schnotz & Bannert, 2003). However, presenting a large amount of visual information increases the extraneous cognitive load and interferes with learning (Ayres & Paas, 2007; Sweller, 2005), although segmentation and user control can be implemented to overcome such limitations (Ayres & Paas, 2007; Mayer & Chandler, 2001). In addition to these principles, animations in the current study were presented in accordance with the cognitive theory of multimedia learning (Mayer & Moreno, 2002) to increase effectiveness.

2.2. Learning outcomes

Several scholars have to date investigated the implications of the Disfluency Effect pertaining to learning outcomes, and have come up with controversial findings. In some scholarly works, it was observed that disfluency had a positive and significant effect on retention (Diemand-Yauman et al., 2011; French et al., 2013; Sungkhasettee et al., 2011), comprehension (Faber et al., 2017; Lehmann et al., 2015) and the transfer of learning (Eitel et al., 2014; Weissgerber & Reinhard, 2017), while others found that the disfluency manipulations failed to facilitate learning (e.g., Carpenter, Wilford, Kornell, & Mullaney, 2013; Eitel et al., 2014; Eitel & Kühl, 2015; Magreehan, Serra, Schwartz, & Narciss, 2015; Miele, Son, & Metcalfe, 2013; Rummer, Schweppe, & Schwede, 2016; Strukelj et al., 2016), and so maintained that the Disfluency Effect could not be proved empirically (Eitel et al., 2014). Recognizing a lack of empirically sound experiments, several scholars emphasized the need to investigate the effect further (Alter, Oppenheimer, Epley, & Eyre, 2007; Eitel & Kühl, 2015; French et al., 2013; Lee, 2013; Pieger, Mengelkamp, & Bannert, 2016; Strukelj et al., 2016). Moreover, it is important to clarify when/how/where disfluency is a desirable difficulty (Alter et al., 2007; Yue, Castel, & Bjork, 2013), and there is a further need for empirical studies revealing the limitations of disfluency manipulations (Kühl, Eitel, Scheiter, & Gerjets, 2014; Oppenheimer & Alter, 2014). We thus believe that investigating the effect of disfluency on learning outcomes through a large scale and true experiment could contribute to existing literature.

2.3. Metacognitive judgments

Metacognition is defined as knowing and knowing how to know (Brown, 1975). The methods used to address metacognition vary according to the individual and to contextual differences. Metacognitive judgments require individuals to provide their judgments through monitoring and reflecting on their cognitive activities (Nelson & Narens, 1990). Judgments of Learning (JOL) and Ease of Learning (EOL) are among these metacognitive judgments. EOL refers to the expected difficulty of a forthcoming task, whereas JOL is defined as the expected probability of recalling items in a recently studied task (Dunlosky & Bjork, 2008). More specifically, EOL judgments are made before learning, and can be influenced by the Disfluency Effect due to the lack of further information about the topic being studied (Pieger et al., 2016). On the other hand, JOLs are made after studying the material, and are based on the personal estimates of the learners (Dunlosky & Bjork, 2008).

It has been suggested that such metacognitive decisions can be influenced by the Disfluency Effect (Dunlosky & Mueller, 2016; Koriat, Ma'ayan, & Nussinson, 2006; Mueller, Tauber, & Dunlosky, 2013; Pieger et al., 2016; Song & Schwarz, 2008; Yue et al., 2013). For instance, JOLs were slightly lower in the disfluent groups in several studies (Besken & Mulligan, 2014; Magreehan et al., 2015; Yue et al., 2013), whereas the effect was medium to large in other works (Mueller et al., 2013; Pieger et al., 2016; Yue et al., 2013). Disfluency affect also EOL judgments (Bannert, 2017; Pieger et al., 2016; Pieger, Mengelkamp & Bannert, 2017; Song & Schwarz, 2008). In addition to aforementioned literature, the activation potential of System 2 through the disfluency manipulations makes it a metacognitive regulation (Alter et al., 2007). Besides, considering the implications of disfluency through a multi-method assessment (i.e., both EOL and JOLs) could be quite fruitful (Desoete, 2008).

2.4. Cognitive Load Theory of multimedia learning and measuring cognitive load

Multimedia is defined as the use of text and pictures together (Mayer, 2009), and like other instructional implementations, it is likely to foster learning. Learning from multimedia materials with high information density may impose a cognitive load on the learner due to the limited capacity of the human cognitive architecture. This load can be defined as a kind of pressure or tension on the working memory as a result of the cognitive processes required for a learning task (Driscoll, 2005). That is, retaining instructional content in a limited capacity memory and trying to make a schema about the content can be quite challenging, depending on the complexity of the task (i.e., intrinsic cognitive load) or the material design (i.e., extraneous cognitive load) (Chandler & Sweller, 1991). While scholars have limited control over the intrinsic cognitive load, it is easy to manipulate the extrinsic one. Accordingly, there is not satisfactory evidence to show whether the disfluency manipulations stimulate information processing or trigger the extraneous cognitive load (Xie, Zhou, & Liu, 2018). Thus, the cognitive load was also considered in the current study in which the material was manipulated.

The Cognitive Theory of Multimedia Learning (CTML) focuses on the effective use of the limited working memory capacity through the development of relevant instructional materials (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). The theory is supported by three assumptions, being dual channel, limited capacity and active processing (Mayer & Moreno, 2003). According to the dual channel assumption, visual and pictorial representations are processed in the visual channel whereas audio and verbal information is processed in the verbal channel. Similarly, according to Paivio's (1991) Dual Coding Theory, visual materials (e.g. pictures & graphics) are processed in the visual channel, whereas written or verbal materials are processed in the verbal channel. According to the limited capacity assumption, only a few pieces of the new information can be processed in each channel simultaneously. Finally, according to the active processing assumption, meaningful learning occurs when individuals select the relevant information from the two channels, and organize and integrate it with their prior knowledge. In this regard, both texts (i.e., subtitles) and animations were considered in the current study, in accordance with the dual channel assumption, and the texts (i.e., subtitles) were designed in line with European subtitle standards to comply with the limited capacity assumption (Karamitroglou, 1998).

It is crucial to keep the cognitive load at an appropriate level for effective learning, but this requires robust measurement of the cognitive load first. Cognitive load measurements can be either subjective or objective. Self-reported cognitive load scales are generally subjective measures (Hart & Staveland, 1988; Paas, 1992; Van Gog & Paas, 2008), while learning outputs (Mayer, 2009), time-on-task (Van Gog & Paas, 2008), task complexity (Brünken, Seufert, & Paas, 2010), heart rate (Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004), eye tracking (Strukelj et al., 2016) and secondary task analysis can be used as objective measures (Brünken, Steinbacher, Plass, & Leutner, 2002). Contemporary disfluency literature often resorts to the subjective measurement of cognitive load (e.g., Sanchez & Jaeger, 2015; Strukelj et al., 2016), whereas the implementation of objective measures is rare (Eitel et al., 2014; Kühl et al., 2014b; Lehmann et al., 2015; Pieger, Mengelkamp, & Bannert, 2017). Subjective measures have been used to address the perceived difficulty (Eitel et al., 2014; Eitel & Kühl, 2015; Pieger et al., 2017) and the difficulty of the material (Diemand-Yauman et al., 2011; Eitel et al., 2014; Eitel & Kühl, 2015; Sanchez & Jaeger, 2015; Song & Schwarz, 2008; Thompson & Ince, 2013), there are a few studies investigating the relationship between the perceived difficulty and the material difficulty in literature. Finally, the importance of objective measures is often emphasized (Dunlosky & Mueller, 2016; Lehmann et al., 2015). Thus, we investigated the relationship between the perceived difficulty and the objective cognitive load in a disfluency experiment. More specifically, cognitive load is addressed through both subjective and objective means, including the total number of animations watched, secondary task performances and learning outcomes.

2.5. Disfluency manipulations

Several manipulation types are observed in disfluency studies. In some scholarly works, only the font type was changed (French et al., 2013; Song & Schwarz, 2008), while some scholars have compared the use of different fonts (Lee, 2013). The disfluency can be created through changes in font size (Rhodes & Castel, 2008; Strukelj et al., 2016) or using gray or italicized font types as well (Alter et al., 2007; Alter & Oppenheimer, 2008a). Some scholars utilized a gray tint along with a change in the font type (Diemand-Yauman et al., 2011; Lehmann et al., 2015). Unlike these, Yue et al. (2013) blurred manipulations. These changes were generally applied to the entire text or not applied at all (Alter & Oppenheimer, 2008b, 2009a; Diemand-Yauman et al., 2011; Eitel & Kühl, 2015; Oppenheimer & Frank, 2008; Pieger et al., 2017, 2016; Rhodes & Castel, 2008; Song & Schwarz, 2008; Strukelj et al., 2016).

Disfluent and fluent words were used together in some studies (Besken & Mulligan, 2013; Pieger et al., 2016; Rosner, Davis, & Milliken, 2015; Rummer et al., 2016). These scholars resorted to short word lists, but they did not examine different fluency types with a strong experimental design. For instance, the use of short word lists could be the reason why the effects of disfluency manipulations were not observed (Rummer et al., 2016), and so the effect needs to be tested with longer and more authentic materials (Pieger et al., 2017; Strukelj et al., 2016). It is also necessary for learners to focus on the important parts of an authentic material if they are to use their limited mental resources effectively (Alter, 2013). Accordingly, the current material was not formed into word pairs, but rather half of the original material was manipulated for the Disfluency Effect.

There are few studies in literature involving pictures and text together (Eitel & Kühl, 2015; Eitel et al., 2014; Kühl et al., 2014b), with static pictures included in these studies, with a single exception (i.e., Eitel et al., 2014). The only study to blur pictures was conducted by Eitel et al. (2014) who maintained that disfluent pictures were undesirable difficulties for better learning outcomes. That said, another study revealed that even a short presentation of pictures in such a multimedia presentation supported learners cognitively (Eitel, Scheiter, & Schüler, 2013).

Animations were used in the current study in addition to words, given the potential of the approach to improve comprehension when used in accordance with the CTML (Mayer & Moreno, 2002). The animations in the material were blurred as implemented by Eitel et al. (2014). Besides, they were provided as fuzzy images as Leder (2003) applied. The purpose was to force learners to focus on the relevant parts of the material and generate more information, rather than to create perceptual difficulties (Kühl & Eitel, 2016; Leder, 2003). We believed that such an approach would help us investigate the disfluency effect better.

2.6. Potential confounds

There are several contaminating variables in literature pertaining to learning outcomes that are generally associated with individual learning characteristics. Considering them as control variables may help scholars to design robust experiments. Prior knowledge and working memory capacity were control variables in related studies (Eitel et al., 2014; French et al., 2013; Kühl, Eitel, Damnik, & Körndle, 2014; Lehmann et al., 2015; Strukelj et al., 2016; Thompson et al., 2013; Xie et al., 2018). For instance, Lehmann et al. (2015) maintained that working memory capacity plays a crucial role in the Disfluency Effect, being positively related with both retention and comprehension. In contrast, disfluency was found not to be moderated by working memory capacity in another study (Strukelj et al., 2016). Such controversial findings may stem from the complexity of the material, the nature of the text manipulation and the length of the material (Kühl & Eitel, 2016; Weissgerber & Reinhard, 2017). Thus, we attempted to control the learners' prior knowledge and the working memory capacity to eliminate such potential confounds.

2.7. Current research and hypotheses

The current study involved text manipulation (fluent, mixed, disfluent), combined with animation manipulations (fluent vs. disfluent), with the goal of testing the implications of these disfluency manipulations on learning outcomes, metacognitive judgments and cognitive load. We hypothesized that learners in the disfluent groups would be required to invest more mental effort, which would lead to better retention, comprehension and achievement scores. In particular, we expected disfluent conditions to lead to higher levels of cognitive difficulty and to better learning outcomes, and thus the objective cognitive load scores under fluent conditions were expected to be lower than the cognitive load scores under other conditions. We further expected the correlation between learning outcomes and cognitive load to be significant and high. Finally, we hypothesized that disfluency should increase the perceived difficulty of the material, resulting in lower scores for the EOL and JOLs.

3. Method

3.1. Participants and design

A convenience sample of undergraduate students constituted the sample. The context was a high-rank state university in Eastern Europe where researchers worked. Of 317 voluntary participants, the proportion of valid cases was 92% (*n* = 292; 181 females [62%]; Mean age = 20.06 years; *SD* = 2.44). Students were given a voucher for a discount on coffee after the experiment. Two students who did not know the language of the instructional material (i.e., Turkish) and one who were visually impaired were excluded from the experiment at the inception. Students were randomly assigned to the levels of the treatment and their distribution with regard to demographics (e.g., gender, department, grade level) were similar in each condition summarized in Table 1: 1. A control group (FF) exposed to fluent text and animation; 2. fluent text and disfluent animation (FD); 3. mixed text (i.e. fluent & disfluent) and fluent animation (MF); 4. mixed text (i.e. fluent & disfluent) and disfluent animation (MD); 5. disfluent text and animation (DD); and 6. disfluent text and fluent animation (DF). Thus, the fluency manipulations on texts and animations were independent variables. Finally, three different conditions were used in texts, whereas two conditions were used in the animations.

3.2. Materials

A computer-assisted learning material on baseball and a paper-based Visual Working Memory (VWM) Test response sheet were prepared for the experiment in Turkish. The expectation of low prior knowledge due to the sports culture of the country was influential in the selection of the topic. The material was prepared to be compatible with all computer operating systems. Each action of the learners was recorded as a log file. The navigation menu was located on the left side and the navigation buttons were located below the menu. The titles of the materials were presented in the menu in addition to relevant explanations. Contents were presented on the right side after clicking the relevant titles.

Animations with texts were used since providing pictures and sound together is not recommended in the literature. The material contained 10 2D and 3D instructional animations of a total 261-s duration. As it was found useful to divide complex topics into sections for inexperienced learners, the participants were given the opportunity to process the material step-by-step (Mayer, 2009). The length of each animation varied from between 15 and 44 s across subtopics. The headings of the subtopics were determined by a former national baseball team coach, and the opinions of further field experts were considered while developing animations and texts (i.e., subtitles). Subtopics' were decided as "What is baseball", "Diamond (Game field)", "Teams and players", "Equipments", "Game length" and "Offence and Defence" respectively. The subtitles were prepared in line with the European Subtitle Standards to ensure compliance with the limited capacity assumption in material design (Karamitroglou, 1998). That is, the subtitles appeared at the bottom of the screen in a centered format. Each word in the subtitles could be read in 1/3 of a second, and the duration of each subtitle was set depending on the number of words. Overall, the material consisted of 267 words.

It was decided that the texts should be displayed on the same screen in a mixed text (i.e., fluent & disfluent) condition in the 3rd (i.e. MF) and 4th scenarios (i.e. MD), based on the idea that it could be useful to

Table 1
Assignment of participants to experimental conditions.

			FF (Fluent text & fluent animation) *n* = 46	FD (Fluent text & disfluent animation) *n* = 51	MF (Mixed text & fluent animation) *n* = 46	MD (Mixed text & disfluent animation) *n* = 50	DD (Disfluent text & disfluent animation) *n* = 52	DF (Disfluent text & fluent animation) *n* = 47
Gender	Female	*f*	33	35	30	34	26	23
		%	71.7	68.6	65.2	68.0	50.0	48.9
	Male	*f*	13	16	16	16	26	24
		%	28.3	31.4	34.8	32.0	50.0	51.1
Grade levels	1	*f*	21	22	27	28	30	26
		%	45.7	43.1	58.7	56.0	57.7	55.3
	2	*f*	9	9	6	7	8	5
		%	19.6	17.6	13.0	14.0	15.4	10.6
	3	*f*	11	17	9	12	9	9
		%	23.9	33.3	19.6	24.0	17.3	19.1
	4	*f*	5	3	4	3	5	7
		%	10.9	5.9	8.7	6.0	9.6	14.9
Departments	French Teaching	*f*	1	1	2	1	2	1
		%	2.2	2.0	4.3	2.0	3.8	2.1
	English Teaching	*f*	12	10	10	11	10	5
		%	26.1	19.6	21.7	22.0	19.2	10.6
	Special Education	*f*	5	6	4	7	8	8
		%	10.9	11.8	8.7	14.0	15.4	17.0
	Social Studies	*f*	8	6	9	9	6	8
		%	17.4	11.8	19.6	18.0	11.5	17.0
	Primary School Mathematics	*f*	10	8	9	7	8	10
		%	21.7	15.7	19.6	14.0	15.4	21.3
	Classroom Education	*f*	2	10	2	6	8	4
		%	4.3	19.6	4.3	12.0	15.4	8.5
	Arts and Crafts	*f*	0	1	0	0	0	0
		%	0	2.0	0	0	0	0
	Computer Education	*f*	8	9	10	9	10	11
		%	17.4	17.6	21.7	18.0	19.2	23.4


Konu Anlatımı - 1 / 5

Bu ekranda beyzbolun tanımına ve oyun alanına yönelik 2 animasyon yer almaktadır. Sonraki ekrana geçebilmek için her animasyonu en az 1 defa izlemeniz gerekmektedir. Her animasyonu istediğiniz kadar izleyebilirsiniz.

Beyzbol belirli bir mesafeden atılan topa sopayla vurulup, yine belirli bir mesafenin kat edildiği bir spor dalıdır.

Beyzbol pasta dilimine benzer bir alanda oynanır.

Sayfa : 13/29

Fig. 1. A screenshot from the 3rd scenario.

focus learners' attention on the important parts of the content (Alter, 2013). It was also maintained that presenting fluent and disfluent texts on one screen would be useful (Rummer et al., 2016), and so fluent and disfluent texts were presented on the same screen. As illustrated in Fig. 1, the text in one animation was fluent, while the other was disfluent. The conditions under which the fluent texts were given first were included in the experiment as this sequence had been found useful in previous studies (Pieger et al., 2017).

The instruction was designed as self-paced, since multimedia messages in a self-paced system have been shown to lead to better learning outcomes (Kühl et al., 2014a; Mayer, 2009). Finally, the total number of animations watched were considered as appropriate to address objective cognitive load under different scenarios effectively.

3.3. Measurements

The measurements involved metacognitive judgments (EOL, JOL), learning outcomes (retention, comprehension and achievement), objective and subjective ratings of the cognitive load, a Physical Vision Test, a Demographic Questionnaire and evaluations of the text and animation manipulations. In addition, the participants' VVM capacity and prior knowledge were measured to control individual differences. All measures (except VVM capacity) were assessed in a computer-assisted learning environment.

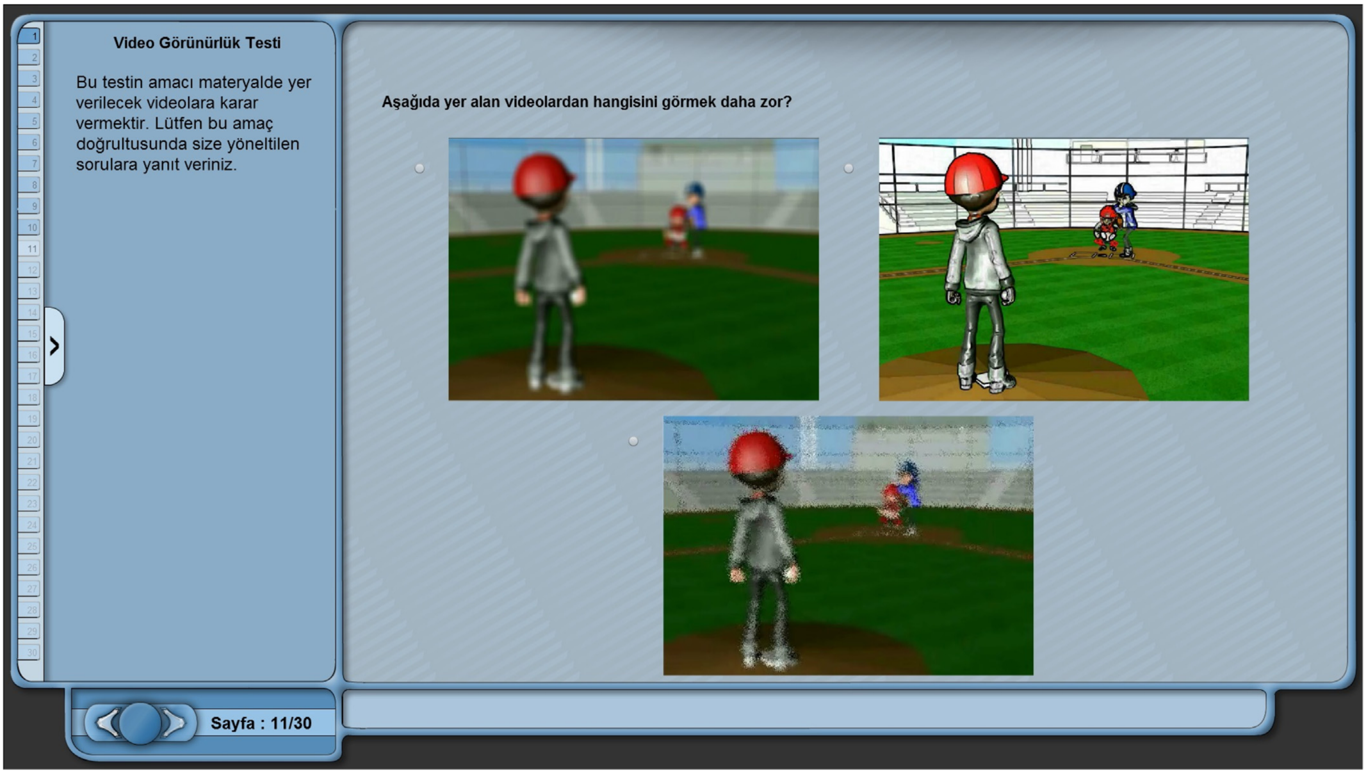
The Physical Vision Test was used to determine the normal-vision participants (Rosner et al., 2015). In this test, volunteers were asked to answer the following question: “Do you have any problems with eyesight?” The participants who selected “No” continued studying the material, whereas those who selected “Yes” were asked if they had tools to correct their vision problems. Those with the relevant tools passed to the next screen and those who didn't were excluded. Demographic information was gathered through the Demographic Questionnaire.

Disfluent font types were chosen based on the Text Manipulation Check. In this test, students selected the text manipulation to be used in the experimental condition, or more specifically, they decided which of the three fonts was more difficult to read (Sidi, Ophir, & Ackerman, 2016; Thompson et al., 2013). In this regard, the same content was given in Haettenschweiler, Monotype Corsiva and Comic Sans MS font types, in accordance with previous studies in literature in which Haettenschweiler (Diemand-Yauman et al., 2011; Eitel et al., 2014; Lehmann et al., 2015; Seufert, Wagner, & Westphal, 2017), Monotype Corsiva (French et al., 2013; Seufert et al., 2017) and Comic Sans MS (Diemand-Yauman et al., 2011; Rummer et al., 2016) had positive effects on learning outcomes. The Arial font type was chosen as the fluent text type, in line with literature (Eitel et al., 2014; Eitel & Kühl, 2015; Faber et al., 2017; Sidi, Ophir, & Ackerman, 2016), and the selected font type was used for the subtitles in the animations of the 3rd, 4th, 5th and 6th scenarios.

An Animation Manipulation Check was used to determine disfluent animation types. In this regard, the participants were asked to select the manipulation to be used in the 2nd, 4th and 5th scenarios by deciding which of the three pictures was most difficult to see (Eitel et al., 2014; Leder, 2003). As provided in Fig. 2, the same picture was given with squirelleblur, wash and amplush manipulations. Squirelleblur is a kind of blurring, wash is a watercolor effect and amplush is a fuzzy image manipulation. The choice of the participants determined the appearance of the animations in the learning phase.

A Visual Patterns Test (VPT) was used to assess visual working memory (VWM) capacity due to its possible influence on the disfluency manipulations. A computer-adapted form of the test was used (Della Sala, Gray, Baddeley, & Wilson, 1997) which was translated to participants' target languages by professional translators. The process was carried out in line with a previous study making use of the Visual Patterns Test (Eitel et al., 2014). In the test were black and white squares resembling a puzzle structure. These images were displayed on the screen for 3 s, and the participants were asked to mark the black squares on the response sheet. A total of 30 items were presented in increasing complexity, and each figure marked as complete and correct was worth 1 point.

Metacognitive judgments were evaluated by entering numbers on a



Video Görünürlük Testi

Bu testin amacı materyalde yer verilecek videolara karar vermektir. Lütfen bu amaç doğrultusunda size yöneltilen sorulara yanıt veriniz.

Aşağıda yer alan videolardan hangisini görmek daha zor?

Sayfa : 11/30

Fig. 2. Screen shot from animation manipulation check.

continuous scale ranging from 0 (difficult) through to 100 (easy). For EOL, the participants responded to the following question after the animation: “How easy or difficult was it to learn the topic?” (Pieger et al., 2016). In this test, the learners were presented with a 15-s section of the material to be presented in the computer assisted environment that was the first animation in the 1st, 2nd, 5th and 6th scenarios where there were no mixed animations. In other scenarios, 7.5 s of the first animation and 7.5 s of the second animation were used. On the other hand, JOLs were evaluated from the responses to the following question: “What percentage of questions about the text will you answer correctly?” (Pieger et al., 2016). The participants had to answer the question on a scale ranging from “0 = none” through to “100 = all”.

Subjective cognitive load was assessed through the item “How much mental effort did you invest?” (Paas, 1992). The scale, which was translated into the target language by Kılıç and Karadeniz (2004), had a 9-point Likert structure ranging from “1 = Very little” through “9 = Too much”, while the following question was used to address the extraneous cognitive load: “How difficult was it for you to learn with the material?” (Kalyuga, Chandler, & Sweller, 1998). This scale was preferred based on its focus on the material rather than the content (Cierniak, Scheiter, & Gerjets, 2009). The item was rated on a seven-point Likert scale.

A secondary task performance and the total number of animations watched were used to measure objective cognitive load. Secondary task stimuli were used for the objective ratings. It is stated that the use of secondary tasks is fruitful despite its rare usage. In this technique, which is based on the limited capacity hypothesis of the Cognitive Load Theory, a secondary task is given in addition to the primary one (Brünken et al., 2002) in the form of a “Click Here!” shape appearing randomly on one of the narration screens. The time that elapsed between the appearance of the secondary on the screen and participants' clicking behavior was recorded as the response time. The response time of the participants who did not click on the task was accepted as 36 s. The total number of animations watched was also considered as an objective research measure. All of these values were recorded by the computer interface. Every animation had to be watched at least once by the learners, although the participants were free to watch animations as many times as they wanted.

The computer-based test for learning outcomes measured retention, comprehension and achievement performances. The learners were asked to answer all of the questions in the tests, with no time limit applied. The Retention Test consisted of 11 open-ended questions (French et al., 2013; Lehmann et al., 2015) (e.g. Write as much as you can about the basic baseball equipment), with each relevant statement in the answer key being worth 1 point up to a maximum score of 53 points. In the Comprehension Test, there were six open-ended questions similar to those used by Lehmann et al. (2015) (e.g. What happens if three strikers are out from the away team and then three strikers are out from the home team?). Unlike in the Retention Test, in which there were free recall questions, this test had higher-level questions that allowed the learners to make deductions from a specific situation, with each question worth 1 point. The answers given to the open-ended questions were rated by two independent scholars and revealed inter-rater reliability coefficients of 0.90 or above. Finally, an achievement test was developed and piloted twice to measure achievements in the learning content. The test consisted of 15 multiple choice questions (e.g. How many innings does a baseball game last?), with each correct answer being worth 1 point. The test was used to assess the prior knowledge of the learners prior to the learning phase as well. It revealed acceptable (i.e., + 0.70) internal consistency coefficients in each implementation.

3.4. Procedure

Fig. 3 illustrates the data collection process that was carried out over a total of 34 sessions, based on the availability of the participants and the computer laboratories. The participants were tested in groups of three to 27 per session, with a single session lasted for about 30 min. The process was carried out in three identical computer laboratories in terms of hardware and software. Learners were free to sit anywhere in

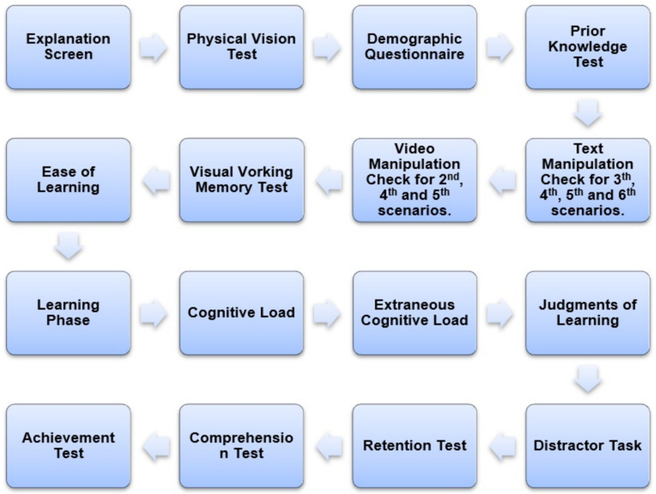


Fig. 3. Data-collection procedure.

the laboratories, and the six conditions were allotted randomly to the 36 computers in each laboratory. The learners were told that they would be participating in a learning experiment that they could stop at any time, and they were also told that written instructions would be provided on each page of the material. The procedure was explained in brief.

The experiment began with the explanation screen, containing ethical issues and the contact details of the researchers, after which the Physical Vision Test was applied. Afterwards, participants with normal vision were requested to fill in the demographic questionnaire and complete the Prior Knowledge Test. After this, the participants in the 3rd, 4th, 5th and 6th scenarios were asked to choose an option from the three text manipulations, followed by the animation manipulation check for the 2nd, 4th and 5th scenarios. In the next step, the VWM test was completed by all groups. Ease of learning judgments were made prior to the animations, and the learning phase then began. The narration was self-paced, although all response times and actions were recorded. The participants were free to return to pages they had watched before, but could not go forward without completing the current page. After studying the material, cognitive load and extraneous cognitive load tests were completed, and learning judgments were made with the JOL Test. After the JOL, a distractor task was given in which basic mathematics problems were solved. The participants who answered the questions in less than 30 s passed to the next page, while those who could not answer on time were automatically directed to the next page after 30 s. At the end, the retention, comprehension and achievements tests were completed. The experiment was terminated with a thank you message on the screen and coffee discount vouchers were distributed.

3.5. Ethical issues

The research proposal was approved by the institutional review board of the university. The study was carried out on a voluntary basis. An explanation screen containing ethical issues were presented to the volunteers before the experiment. The screen contained information regarding the nature of the study. Participants were informed that they were free to leave the study any time and no personal information was collected. In addition, it was stated that the data would be used only for scientific purposes and would not be shared with any third parties.

4. Results

The descriptive statistics pertaining to all variables of interest are summarized in Table 2. The adjusted means are provided at the subsequent paragraphs whenever an ANCOVA or MANCOVA was needed.

4.1. Participants' manipulation type choices with regards to text and animation

The Text Manipulation Check was used to determine the subtitles in the animations of the MF, MD, DD and DF scenarios. As summarized in Table 3, most participants preferred the Haettenschweiler font.

Similarly, the Animation Manipulation Check was used to determine the manipulation types pertaining to animations in the FD, MD, and DD conditions. None of the learners considered the wash manipulation as disfluent, while the other two options had similar distributions, as summarized in Table 4.

4.2. Learning outcomes

To test the influence of the disfluency types on learning outcomes, a one-way between-groups MANCOVA was conducted, with VWM and prior knowledge being covariates, as suggested in previous literature, and these correlated significantly with the dependent variables. That is, the relationships between the VWM and the retention ($r = 0.207$, $p < 0.001$), comprehension ($r = 0.134$, $p = 0.022$) and achievement ($r = 0.222$, $p < 0.001$) were significant. Similarly, prior knowledge was correlated significantly with retention ($r = 0.273$ $p < 0.001$), comprehension ($r = 0.263$, $p < 0.001$) and achievement ($r = 0.307$, $p < 0.001$).

The adjusted means are provided in Table 5. The highest achievement performance was demonstrated by the MF group (Estimate = 6.387, SE = 0.325) followed by the FD (Estimate = 6.251, SE = 0.308), the DF (Estimate = 6.186, SE = 0.322), the FF (Estimate = 6.173, SE = 0.324) and the MD group (Estimate = 5.984, SE = 0.312). The learners in the DD group demonstrated the worst performance (Estimate = 10.980, SE = 0.860), with the DF condition resulting in the lowest scores in terms of retention (Estimate = 12.180, SE = 0.909) whereas the FF scenario had the best results (Estimate = 13.934, SE = 0.916). Comprehension results were similar to the retention performance in terms of the highest (FF group, Estimate = 1.715, SE = 0.200) and lowest scores (DD group, Estimate = 1.395, SE = 0.188). These descriptive statistics suggest that providing disfluent animations with disfluent texts could decrease scores in terms of learning outcomes. However, the MANCOVA revealed no significant effect of disfluency on either retention ($F_{(5,284)} = 1.203$; $p = 0.308$; $\eta_p^2 = 0.021$; *Power* = .426), comprehension ($F_{(5,284)} = 0.429$; $p = 0.829$; $\eta_p^2 = 0.007$; *Power* = .163), or the achievement performance ($F_{(5,284)} = 1.907$; $p = 0.093$; $\eta_p^2 = 0.032$; *Power* = .643). Contrary to our hypotheses, the learning outcomes did not differ across the different scenarios.

4.3. Metacognitive judgments

A one-way between-groups ANCOVA was conducted to assess the disfluency effect on metacognitive judgments (i.e., JOL). A significant correlation existed between the EOL and JOL ($r = 0.282$; $p < 0.001$), and between JOL and VWM ($r = 0.122$; $p = 0.037$). As EOL was measured prior to the experiment and JOL was measured after the experiment, EOL was considered as the covariate. The adjusted JOL means are presented in Table 6.

Interestingly, the participants in the easiest scenario (FF) had the highest scores (*Estimate* = 45.888, *SE* = 3.348), followed by the DD (*Estimate* = 43.068, *SE* = 3.166), the FD (*Estimate* = 40.871, *SE* = 3.178), the MF (*Estimate* = 40.512, *SE* = 3.357) and the DF conditions (*Estimate* = 39.64, *SE* = 3.313). The worst performance was noted among the learners in the MD scenario (*Estimate* = 39.191, *SE* = 3.209). However, the ANCOVA revealed no significant differences between conditions ($F(5,285) = 0.586$; $p = 0.711$; ${\eta_p}^2 = 0.01$;

Table 2
Descriptive statistics pertaining to research variables.

Variable of Interest	Conditions											
	FF (n = 46)		FD (n = 51)		MF (n = 46)		MD (n = 50)		DD (n = 52)		DF (n = 47)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Visual Working Memory (VWM)	21.80	5.001	20.84	4.173	21.37	4.084	20.50	4.888	20.88	4.333	21.98	4.131
Retention	14.11	6.714	12.75	5.851	12.02	6.777	12.56	7.054	10.81	6.481	12.68	6.249
Comprehension	1.74	1.341	1.61	1.343	1.57	1.424	1.64	1.481	1.37	1.428	1.55	1.427
Achievement	6.24	2.162	6.27	2.450	6.30	2.289	5.86	2.474	5.13	2.197	6.38	2.454
Ease of Learning (EOL)	66.37	25.668	61.43	28.360	69.37	30.299	62.06	30.099	56.04	29.282	66.83	25.902
Judgments of Learning (JOL)	46.57	19.575	40.39	27.706	41.89	23.928	38.86	21.831	41.33	23.443	40.43	24.021
Cognitive Load	5.48	1.560	5.84	1.377	5.59	1.654	5.64	1.925	5.19	1.669	5.21	1.641
Extraneous Cognitive Load	3.50	1.169	4.39	1.484	3.63	1.356	4.62	1.354	4.48	1.627	3.77	1.322
Number of Animations Watched	12.80	3.436	12.41	3.517	12.63	2.760	13.86	5.757	11.44	2.191	11.49	2.781
Secondary Task	32.17	9.794	28.14	13.355	30.85	11.169	31.64	10.950	27.90	13.187	29.53	12.436

FF: Fluent text and fluent animation, FD: Fluent text and disfluent animation, MF: Mixed text and fluent animation, MD: Mixed text and disfluent animation, DD: Disfluent text and disfluent animation, DF: Disfluent text and fluent animation.

Table 3
Participants' preferences of text manipulations.

	Conditions									
	MF		MD		DD		DF		Total	
	f	%	f	%	f	%	f	%	f	%
Haettenschweiler	46	100	47	94	50	96.2	45	95.7	188	96.5
Monotype Corsiva	–	–	2	4	1	1.9	–	–	3	1.5
Comic Sans MS	–	–	1	2	1	1.9	2	4.3	4	2.0

MF: Mixed text and fluent animation, MD: Mixed text and disfluent animation, DD: Disfluent text and disfluent animation, DF: Disfluent text and fluent animation.

Table 4
Participants' preferences of animation manipulations.

	Conditions							
	FD		MD		DD		Total	
	f	%	f	%	f	%	f	%
Squirelleblur	24	47.1	21	42	23	44.2	68	44.4
Amplush	27	52.9	29	58	29	55.8	85	55.6
Wash	–	–	–	–	–	–	–	–

FD: Fluent text and disfluent animation, MD: Mixed text and disfluent animation, DD: Disfluent text and disfluent animation.

Power = .214) which was inconsistent with our hypothesis. On the other hand, that fact that the learners felt confident about their JOL in the DD condition, which was perceived as the most difficult scenario, needs further discussion.

Table 6
Adjusted means in Judgments of Learning (JOL).

Conditions	*Mean*	*SE*
FF (Fluent text & fluent animation)	45.888	3.348
FD (Fluent text & disfluent animation)	40.871	3.178
MF (Mixed text & fluent animation)	40.512	3.357
MD (Mixed text & disfluent animation)	39.191	3.209
DD (Disfluent text & disfluent animation)	43.068	3.166
DF (Disfluent text & fluent animation)	39.640	3.313

4.4. Cognitive load

One-way between-groups ANOVA was used to check for differences between the conditions in terms of the cognitive load measures. Cognitive load involved either subjective or objective measures, with the subjective measures being cognitive load and extraneous cognitive load, and the secondary task and the number of animations watched were the objective measures. In this regard, four ANOVAs were conducted, in which the significance level was set at 0.0125 by dividing the Alpha into the number of ANOVAs (i.e., Bonferroni adjustment; 0.05/4 = 0.0125). The ANOVA revealed no significant differences among the conditions in terms of cognitive load ($F_{(5,286)} = 1.192$; $p = 0.313$; $\eta_p^2 = 0.02$; *Power* = .422) and the secondary task performance ($F_{(5,286)} = 1.125$; $p = 0.347$; $\eta_p^2 = 0.019$; *Power* = .399). The conditions seemed to differ in terms of the number of animations watched ($F_{(5,286)} = 3.122$; $p = 0.009$; $\eta_p^2 = 0.052$; *Power* = .874). However, the Levene's Test reached significance ($F_{(5,286)} = 6.874$, $p < 0.001$). Thus, Tamhane post-hoc test was used, which revealed that none of the differences across the groups was statistically significant.

On the other hand, a significant effect of disfluency was noted on extraneous cognitive load ($F_{(5,286)} = 5.857$; $p < 0.001$; ${\eta_p}^2 = 0.093$; *Power* = .994). Scheffe's post hoc test was used for pairwise

Table 5
Adjusted means in learning outcomes.

	Conditions											
	FF (n = 46)		FD (n = 51)		MF (n = 46)		MD (n = 50)		DD (n = 52)		DF (n = 47)	
	Mean	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*
Retention	13.934	.916	12.695	.870	12.223	.916	12.878	.879	10.980	.860	12.180	.909
Comprehension	1.715	.200	1.589	.190	1.612	.200	1.693	.192	1.395	.188	1.462	.199
Achievement	6.173	.324	6.251	.308	6.387	.325	5.984	.312	5.202	.305	6.186	.322

FF: Fluent text and fluent animation, FD: Fluent text and disfluent animation, MF: Mixed text and fluent animation, MD: Mixed text and disfluent animation, DD: Disfluent text and disfluent animation, DF: Disfluent text and fluent animation.

Table 7
Relationships across variables.

	Cognitive load	Extraneous cognitive load	Secondary task	Number of animations watched	Achievement	Retention
Extraneous cognitive load	0.034	–				
Secondary task	−0.052	0.009	–			
Number of animations watched	.241***	.149*	.184**	–		
Achievement	.211***	−0.125*	0.021	.213***	–	
Retention	.268***	−0.085	.112	.333***	.579***	–
Comprehension	.189***	-.144*	0.007	.130*	.442***	.547***

*p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001.

comparisons, which revealed that the FF group ($M = 3.5$; $SD = 1.17$) invested significantly less extraneous cognitive load than those in the MD ($M = 4.62$; $SD = 1.35$; p = 0.002) and the DD group ($M = 4.48$; $SD = 1.63$; p = 0.009). Finally, a significant difference was found between the MD ($M = 4.62$; $SD = 1.35$) and MF conditions ($M = 3.63$; $SD = 1.36$; p = 0.009). The findings suggest that the participants studying with disfluent animations or mixed subtitles invested more extraneous cognitive load than others.

4.5. *The relationship between cognitive load and learning outcomes*

Relationships across the variables of interest are summarized in Table 7.

The findings revealed positive and significant relationships between cognitive load, the number of animations watched and the learning outcomes, as hypothesized. The extraneous cognitive load was negatively related with achievement ($r = -0.125$, $p = 0.032$) and comprehension ($r = -0.144$, $p = 0.014$). In addition, extraneous cognitive load correlated positively with the number of animations watched ($r = 0.149$, $p = 0.01$). Accordingly, a significant correlation was found between the secondary task performance and the number of animations watched ($r = 0.184$, $p = .002$).

5. Discussion

Effect of different fluency manipulations on learning outcomes, metacognitive judgment and cognitive load was assessed in the current study. Undergraduate students were assigned randomly to the disfluency conditions in a computer-based learning environment. Working memory capacity, prior knowledge, cognitive load and metacognitive judgments were considered as well. The originality was the investigation of the role of disfluency through fluent and disfluent texts together with animations. There were not significant differences with regard to learning outcomes and metacognitive judgments in the experimental groups whereas significant differences were observed with regard to the extraneous cognitive load. In addition, significant relationships were observed between the cognitive load, the number of animations watched by the students and the learning outcomes.

The findings related to the participants' manipulation type choices can contribute to literature in the provision of disfluency. Several font types have been used in literature to provide disfluent conditions (Diemand-Yauman et al., 2011; Eitel et al., 2014; French et al., 2013; Lehmann et al., 2015; Rummer et al., 2016; Seufert et al., 2017). The disfluency options provided to the participants were not preferred at the same rate, with most learners choosing the Haettenschweiler font as disfluent, whereas the Monotype Corsiva and Comic Sans MS fonts were not preferred, which may be attributable to contextual features. Besides, learners who are used to materials with similar font types may have adjusted themselves to the current fonts. The role of individual differences has also been discussed in previous studies (Kühl et al., 2014a). In this regard, further topics and font options, along with contextual characteristics, may be taken into account.

The findings were similar also in terms of the selection of disfluent animations. That is, even though the manipulation options were selected from effective ones in literature (Eitel et al., 2014; Leder, 2003), the wash manipulation was not judged as disfluent by any of the participants. As manipulations forcing learners to invest more effort could be helpful in creating disfluency (Kühl & Eitel, 2016; Leder, 2003), the wash option was considered useless in the current context. Disfluency has a sensitive nature that can be influenced by other such variables as learning characteristics, application environments and time (Rummer et al., 2016), and so further studies could be helpful in mapping the role of dynamic images that are rarely studied in literature.

The findings refuted our hypothesis regarding the influence of the disfluency on learning outcomes. That is, the experimental conditions did not trigger System 2, which was inconsistent with the results of studies that revealed significant retention and comprehension differences (Alter et al., 2007; Alter & Oppenheimer, 2008a; Diemand-Yauman et al., 2011; Lehmann et al., 2015; Weissgerber & Reinhard, 2017). On the other hand, the findings were in line with several empirical works (e.g. Carpenter et al., 2013; Eitel et al., 2014; Eitel & Kühl, 2015; Magreehan et al., 2015; Miele et al., 2013; Pieger et al., 2016; Rummer et al., 2016; Strukelj et al., 2016). Despite insignificant differences across the groups, the MF (Mixed text and Fluent animation) group had the highest performance in terms of achievement performance, which can be considered interesting, as this group was exposed to mixed subtitles and fluent animations. This finding could be promising for further studies, indicating that it may be beneficial to concentrate the attention of learners on the important parts of the material (Alter, 2013). In addition, these findings make it evident that it is more useful to provide learners with fluent and disfluent text on the same screen (Rummer et al., 2016). On the other hand, learners in the FF (Fluent text & Fluent animation) group were the most successful participants in retention and comprehension. This implies a drawback of disfluency to activate System 2.

A similar and interesting finding was that the DD (Disfluent text and Disfluent animation) group had the lowest scores in terms of learning outcomes, which was inconsistent with scholarly works that are in favor of the disfluency effect (Alter et al., 2007; Alter & Oppenheimer, 2008a; Diemand-Yauman et al., 2011; Lehmann et al., 2015; Weissgerber & Reinhard, 2017). However, such variations in disfluency implications may also depend on the nature of the material (Weissgerber & Reinhard, 2017), in that dealing with more than one manipulation may not be beneficial for performance (Seufert et al., 2017; Sweller, Ayres, & Kalyuga, 2011), and the provision of disfluent texts and animations together may overshadow the potential of disfluency.

Similar to the findings on the learning outcomes, findings pertaining to metacognitive judgments were not in favor of a positive disfluency effect, which is inconsistent with several studies in literature (e.g. Besken & Mulligan, 2013, 2014; Magreehan et al., 2015; Pieger et al., 2016; Rhodes & Castel, 2008; Rosner et al., 2015; Susser, Mulligan, & Besken, 2013; Yue et al., 2013), but in line with other scholarly works (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Sungkhasettee et al., 2011). Although these differences were not statistically significant, interesting trends were apparent in the data. For instance, the participants in the FF (Fluent text and Fluent animation) group had the highest scores in terms of JOL, and the easier judgement of the fluent material than others could be the reason for the concurrence with literature (Besken & Mulligan, 2013, 2014; Magreehan et al., 2015; Pieger et al., 2016; Rhodes & Castel, 2008; Rosner et al., 2015; Susser et al., 2013; Yue et al., 2013). The participants in the MD (Mixed text and Disfluent animation) group had the lowest scores in terms of the JOL, which refuted our hypothesis that the lowest score would be in the DD (Disfluent text and Disfluent animation) group. Learners in the MD (Mixed text and Disfluent animation) condition saw the disfluent animations and perhaps judged the material as difficult. The provision of mixed texts could be another reason for such variations in the JOL scores. Finally, the provision of animations in the material may have led to such differences. Since it is difficult to perfectly replicate an instructional setting described in literature (Strukelj et al., 2016; Weissgerber & Reinhard, 2017), it may not be plausible to compare the current findings with existing studies. Besides, this may be the first study investigating the metacognitive consequences of disfluency manipulations in animations, and so animation disfluency experiments with larger samples are needed to address the consequences of the disfluency manipulations on metacognitive judgments (Pieger et al., 2016).

We hypothesized that disfluent conditions would lead to higher cognitive load; however, the only difference was observed in terms of the extraneous cognitive load. This lack of difference in cognitive load was in line with literature (Eitel et al., 2014; Kühl et al., 2014a; Lehmann et al., 2015). On the other hand, the significant difference in extraneous load was in contrast with literature (Eitel et al., 2014; Lehmann et al., 2015), and could have stemmed from making the material appear more difficult or unusual. As we hypothesized, the FF group demonstrated the lowest extraneous cognitive load scores, whereas the MD (Mixed text and Disfluent animation) and DD (Disfluent text and Disfluent animation) groups had significantly higher scores than the FF (Fluent text and Fluent animation) group. Challenging the working memory limits of the participants through disfluent text and animations could be the reason for this (Yue et al., 2013).

The lack of significant differences regarding the number of animations watched was in line with relevant research on reading times (Strukelj et al., 2016). However, the finding was inconsistent with other studies in which an increase in reading times was recorded (Eitel & Kühl, 2015; Miele et al., 2013; Miele & Molden, 2010; Pieger et al., 2016; Song & Schwarz, 2008). The measurement of study times is an indirect way of investigating the Disfluency Effect (Kühl & Eitel, 2016). As we were the first to examine the number of animations watched, similar investigations are needed to extend the current data. Similarly, no significant differences were observed with regard to the secondary task performance, which was perceived as a useful way of addressing cognitive load in disfluency studies (Seufert et al., 2017). As one of the first empirical studies to investigate the secondary task performance in terms of animation-related disfluency effect, our findings are in line with an eye tracking study (i.e., Strukelj et al., 2016), but conflicts with another eye tracking work (Sanchez & Jaeger, 2015). In brief, the current findings retain the hypothesis that disfluency is independent of objective cognitive load (Alter & Oppenheimer, 2009b; Oppenheimer, 2008).

Positive and significant relationships were observed between cognitive load, the number of animations watched and all learning variables. We expected that watching disfluent animations would increase the level of cognitive load, although the positive and significant correlation between the subjective cognitive load and learning outcomes was interesting. Making the material more difficult to process caused higher levels of cognitive load which may trigger System 2 (e.g., Lehmann et al., 2015). However, the increase in the subjective cognitive load was associated with better learning outcomes, whereas the increase in the extraneous cognitive load was associated with worse learning outcomes. The coefficients related to the positive relationship between cognitive load and learning outcomes were consistently stronger and significant. Such findings could be promising for an investigation of the instructional implications of disfluency manipulations in further empirical studies. The current findings contradict those of studies revealing no change in the extraneous cognitive load (Eitel et al., 2014) and no relationship between the extraneous load and learning (Lehmann et al., 2016). Finally, the significant and positive relationship between the extraneous cognitive load and the number of animations watched could be promising in terms of the potential offered by these alternative cognitive load measures for further disfluency experiments (Seufert et al., 2017).

In brief, the lack of significant differences across the experimental conditions with regard to current dependent variables may lead scholars to support the assumptions of the CTML rather than the Disfluency Effect, and this inference is also in line with some scholarly works (Kühl et al., 2014a). The present study contributes to literature with further empirical evidence through questioning the positive consequences of the disfluency manipulations. This could be a novelty effect, as proposed previously (Rummer et al., 2016). As we failed to provide empirical support for disfluency, one should refrain from using disfluency manipulations in animations with subtitles, in contrast to previous recommendations (Diemand-Yauman et al., 2011).

6. Limitations and future research directions

The present study has several limitations. We particularly examined the effect of disfluency in a computer-assisted learning environment on learning outcomes, metacognitive judgment and cognitive load, but were unable to observe any significant differences across the conditions aside from in the extraneous cognitive load. This may be due to the complexity of the topic choice for the current research setting (i.e. baseball) or the nature of the target population (i.e. undergraduates), and so replication studies with other conventional topics and samples are needed. Second, the participants in the current study were learning from a disfluent material for the first time, and so accommodation to the material may be a confounding variable. In this regard, a comparison of experienced and inexperienced learners with regard to disfluency could be made. Third, the current study is the first to provide learners with disfluent animations, and so we had no reference studies with which to compare the current findings. The complexity of the topic, along with the animation-based nature of the material, may have led us to different findings than current literature (Diemand-Yauman et al., 2011; Eitel et al., 2014; French et al., 2013; Lehmann et al., 2015; Sungkhasettee et al., 2011). Fourth, providing individuals with both disfluent texts and animations could have challenged the participants and overshadowed the potential contribution of the disfluency. That is, the provision of disfluent subtitles and disfluent animations together could have interfered with the learning outcomes (Ayres & Paas, 2007; Sweller, 2005). In this regard, audio may also be used in further studies to decrease the extraneous load in disfluency conditions. Fifth, only a limited number of disfluency manipulations were included in the current experiment through the conducting of a systematic literature review and resorting to manipulation checks. Other manipulation types may be examined in further research. Sixth, the characteristics of the cognitive load measures might have influenced the current findings. For instance, cognitive load may be addressed through EEG measures, in addition to the number of animations watched and the secondary task performance, to validate the current measures. Seventh, the disfluency effect may appear over time (Yue et al., 2013), which requires monitoring long-term memory performance, as realized in some scholarly works (Weissgerber & Reinhard, 2017). Finally, we resorted to metacognitive judgments that are considered a concurrent measure. Promising data can be obtained through the utilization of both concurrent and non-concurrent methods.

Funding sources

This work was funded by Anadolu University Research Fund (Grant

number 1609E617).

Acknowledgements

This study is the summary of the first author's PhD dissertation, which was supervised by the second author and supported by Anadolu University Research Fund. The design process of the material was presented as an abstract at ICITS 2017 (Turkey) and preliminary findings were presented as an abstract at 2018 The Future of Education International Conference (Italy). We thank the Editor and four anonymous reviewers for their outstanding feedback on the previous versions of the manuscript.

References

Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science, 22*(6), 437–442.

Alter, A. L., & Oppenheimer, D. M. (2008a). Effects of fluency on psychological distance and mental construal (or why New York is a large city, but New York is a civilized jungle). *Psychological Science, 19*(2), 161–167.

Alter, A. L., & Oppenheimer, D. M. (2008b). Easy on the mind, easy on the wallet: The roles of familiarity and processing fluency in valuation judgments. *Psychonomic Bulletin & Review, 15*(5), 985–990.

Alter, A. L., & Oppenheimer, D. M. (2009a). Suppressing secrecy through metacognitive ease cognitive fluency encourages self-disclosure. *Psychological Science, 20*(11), 1414–1420.

Alter, A. L., & Oppenheimer, D. M. (2009b). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal Of Experimental Psychology: General, 136*(4), 569–576.

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: Acognitive load approach. *Applied Cognitive Psychology, 21*, 695–700.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*(5), 610–632.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition, 41*(6), 897–903.

Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: Analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/a0034407.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.). *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: MA: MIT Press.

Bjork, E. L., & Bjork, R. (2011). Making things hard on yourself, but in a good way. In M. A. Gernsbacjer, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.). *Psychology in the real world* (pp. 59–68). New York, NY: Worth.

Brown, A. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. *Advances in Child Development and Behavior, 10*, 103–151.

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. L. Plass, R. Moreno, & R. Brünken (Eds.). *Cognitive load theory* (pp. 181–182). Cambridge: Cambridge University Press.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load within multimedia learning by the dual task methodology. *Experimental Psychology, 49*(2), 109–119.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review, 20*(6), 1350–1356.

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2015). Animations showing Lego manipulative tasks: Three potential moderators of effectiveness. *Computers and Education, 85*, 1–13.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25*(2), 315–324.

Della Sala, S., Gray, C., Baddeley, A., & Wilson, L. (1997). *The visual patterns test: A new test of short-term visual recall.* Suffolk: Thames Valley Test Company.

Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: How you test is what you get. *Metacognition and Learning, 3*(3), 189–206.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the ( ): Effects of disfluency on educational outcomes. *Cognition, 118*(1), 111–115.

Driscoll, M. (2005). *Psychology of learning for instruction.* Boston, MA: Pearson.

Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. In J. Dunlosky, & R. A. Bjork (Eds.). *Handbook of metamemory and memory* (pp. 11–28). New York, NY, US: Psychology Press.

Dunlosky, J., & Mueller, M. L. (2016). *Recommendations for exploring the disfluency hypothesis for establishing whether perceptually degrading materials impacts performance.* Metacognition and Learninghttps://doi.org/10.1007/s11409-016-9155-9.

Eggen, P. D., & Kauchak, D. P. (2007). *Strategies for teachers: Teaching content and Thinking skills* (7th ed.). Needham Heights: Allyn and Bacon.

Eitel, A., & Kühl, T. (2015). *Effects of disfluency and test expectancy on learning with text.* Metacognition and Learninghttps://doi.org/10.1007/s11409-015-9145-3.

Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: Does harder-to-read mean better-to-understand? *Applied Cognitive Psychology, 28*(4), 488–501.

Eitel, A., Scheiter, K., & Schüler, A. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology, 27*(4), 451–461.

Faber, M., Mills, C., Kopp, K., & D'mello, S. (2017). The effect of disfluency on mind wandering during text comprehension. *Psychonomic Bulletin & Review, 24*(3), 914–919.

French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, A., et al. (2013). Changing fonts in education: How the benefits vary with ability and dyslexia. *The Journal of Educational Research, 106*(4), 301–304.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshtaki (Eds.). *Human mental workload* (pp. 139–183). Amsterdam, the Netherlands: North-Holland.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*(6), 722–738.

James, W. (1950). *The principles of psychology. (2 vols).* New York: Henry Holt and Co. Reprint.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*(1), 1–17.

Karamitroglou (1998). A proposed set of subtitling standards in Europe. *Translation Journal, 2*(2), 1–15.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36–69.

Kühl, T., & Eitel, A. (2016). *Effects of disfluency on cognitive and metacognitive processes and outcomes.* Metacognition and Learninghttps://doi.org/10.1007/s11409-015-9145-3.

Kühl, T., Eitel, A., Damnik, G., & Körndle, H. (2014a). The impact of disfluency, pacing, and students' need for cognition on learning with multimedia. *Computers in Human Behavior, 35*, 189–198.

Kühl, T., Eitel, A., Scheiter, K., & Gerjets, P. (2014b). A call for an unbiased search for moderators in disfluency research: Reply to Oppenheimer and Alter (2014). *Applied Cognitive Psychology, 28*(5), 805–806.

Kılıç, E., & Karadeniz, Ş. (2004). Hiper ortamlarda öğrencilerin bilişsel yüklenme ve kaybolma düzeylerinin belirlenmesi. *Kuram ve Uygulamada Egitim Yönetimi Dergisi, 10*(4), 562–579.

Leder, H. (2003). Familiar and fluent! Style-related processing hypotheses in aesthetic appreciation. *Empirical Studies of the Arts, 21*(2), 165–175.

Lee, M. H. (2013). Effects of disfluent kanji fonts on reading retention with e-book. *Proceeding 13th international conference on advanced learning technologies (ICALT)* (pp. 481–482). Beijing, China, 15-18 July, 2013.

Lehmann, J., Goussios, C., & Seufert, T. (2015). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning, 11*, 89–105.

Liao, Y. C. (2007). Effects of computer-assisted instruction on students' achievement in Taiwan: A meta-analysis. *Computers and Education, 48*(2), 216–233.

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2015). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning, 11*(1), 35–56.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390–397.

Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review, 14*(1), 87–99.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

Miele, D. B., & Molden, D. C. (2010). Naive theories of intelligence and the role of processing fluency in perceived comprehension. *Journal of Experimental Psychology: General, 139*(3), 535–557.

Miele, D. B., Son, L. K., & Metcalfe, J. (2013). Children's naive theories of intelligence influence their metacognitive judgments. *Child Development, 84*(6), 1879–1886.

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review, 20*(2), 378–384.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.). *The psychology of learning and motivation* (pp. 125–173). NY: Academic Press.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences, 12*(6), 237–241.

Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology, 28*(4), 502–504.

Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition, 106*(3), 1178–1194.

Ornstein, A. C., & Levine, D. U. (1993). *Foundations of educations* (5th ed.). Boston: Houghton Mifflin Co.

Owusu, K. A., Monney, K. A., Appiah, J. Y., & Wilmot, E. M. (2010). Effects of computer-assisted instruction on performance of senior high school biology students in Ghana. *Computers and Education, 55*(2), 904–910.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 45*(3), 255.

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency–Does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction, 44*, 31–40.

Pieger, E., Mengelkamp, C., & Bannert, M. (2017). Fostering analytic metacognitive processes and reducing overconfidence by disfluency: The role of contrast effects. *Applied Cognitive Psychology, 31*(3), 291–301.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615–625.

Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica, 160*, 11–22.

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacognition and Learning, 11*(1), 57–70.

Ryoo, K., & Linn, M. C. (2012). Can dynamic visualizations improve middle school students' understanding of energy in photosynthesis? *Journal of Research in Science Teaching, 49*(2), 218–243.

Sanchez, C. A., & Jaeger, A. J. (2015). If it's hard to read, it changes how long you do it: Reading time as an explanation for perceptual fluency effects on judgment. *Psychonomic Bulletin & Review, 22*(1), 206–211.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*(2), 141–156.

Serra, M. J., & Metcalfe, J. (2009). Effective implementation of metacognition. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.). *Handbook of metacognition in education* (pp. 278–298). New York: Psychology Press.

Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science, 45*(2), 221–238.

Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority-does the medium, screen versus paper, affect performance even with brief tasks? *Metacognition and Learning, 11*(1), 15–33.

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do processing fluency affects effort prediction and motivation. *Psychological Science, 19*(10), 986–988.

Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning, 11*, 71–88.

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review, 18*(5), 973–978.

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition, 41*(7), 1000–1011.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In I. R. E. Mayer (Ed.). *The cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In J. Sweller, P. Ayres, & S. Kalyuga (Eds.). *Cognitive load theory* (pp. 71–85). New York, NY: Springer.

Thompson, D. V., & Ince, E. C. (2013). When disfluency signals competence: The effect of processing difficulty on perceptions of service agents. *Journal of Marketing Research, 50*(2), 228–240.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., et al. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237–251.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247–262.

Van Gerven, P. W., Paas, F., Van Merriënboer, J. J., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*(2), 167–174.

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*(1), 16–26.

Weissgerber, S. C., & Reinhard, M. A. (2017). Is disfluency desirable for learning? *Learning and Instruction, 49*, 199–217.

Xie, H., Zhou, Z., & Liu, Q. (2018). Null effects of perceptual disfluency on learning outcomes in a text-based educational context: A meta-analysis. *Educational Psychology Review, 30*(3), 745–771.

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition, 41*(2), 229–241.

Ulaş İlic is a research assistant at the Department of Computer Education and Instructional Technology at Pamukkale University, Turkey. He has a Ph.D. in Computer Education and Instructional Technology. His research interests are multimedia learning and the implementation of desirable difficulties in instructional settings.

Yavuz Akbulut is a professor at the Department of Educational Sciences at Anadolu University, Turkey. He has an M.A. in computer assisted language learning and a Ph.D. in instructional design and technology. He conducts research on cyberpsychology and learning, multitasking and online gaming behaviors.