Metacognition and Disfluency –

The Effects of Disfluency on Monitoring and Performance

Inaugural-Dissertation

zur Erlangung der Doktorwürde der

Fakultät für Humanwissenschaften

der

Julius-Maximilians-Universität Würzburg

Vorgelegt von

Elisabeth Pieger

aus Kirchehrenbach

Würzburg

2017

Betreuung:	Professor Dr. Maria Bannert
Erstgutachterin:	Professor Dr. Gerhild Nieding
Zweitgutachterin:	PD Dr. Stephanie Pieschl

Tag des Kolloquiums: 20.09.2017

Acknowledgements

I would like to thank all people, who supported me during my PhD.
Particularly, I want to name the following:

Prof. Dr. Maria Bannert
My supervisor

Dr. Christoph Mengelkamp
My co-author & mentor

Prof. Dr. Gerhild Nieding & PD Dr. Stephanie Pieschl
My jury members

Christoph Sonnenberg, Regina Parzl, Markus Hörmann, Svenja Finn, Thorsten Göbel, Hannes Münchow, & Johannes Loudwin
My office-colleagues

Jessica Hartleb, Tamara Schmidtner, Katharina Bruckner, Anika Faber, Sophia Kleinhenz, Laura Häpp, Linda Bauer, Vanessa Flammiger, Anna Horrer, & Emily Lubaway
My research assistants, students, & BA-candidates

... and all colleagues and friends who supported me

...as well as my family, without you it would not have been possible.

Content

German Summary (Zusammenfassung)

Ziel dieser Dissertation ist es, den Einfluss von Disfluency auf metakognitive Einschätzungen sowie deren Genauigkeit und deren Einfluss auf die Leistung zu untersuchen. Dabei wird die Metakognitions- mit der Fluency-Forschung verbunden. *Metakognition* kann als das Denken über das eigene Denken beschrieben werden (Flavell, 1979), während *Fluency* als Verarbeitungsflüssigkeit definiert werden kann (z. B. Alter & Oppenheimer, 2009; Lanska, Olds & Westerman, 2014; Schwarz, 2010). Theorien der Metakognition und des Selbstregulierten Lernens postulieren, dass metakognitive Einschätzungen den Lernprozess sowie die Leistung beeinflussen (z. B. Nelson & Narens, 1990; Winne & Hadwin, 1998). Beispielsweise könnte sich die Einschätzung bzgl. der Schwierigkeit des Lernmaterials oder der eigenen Leistung darauf auswirken, wie viel Zeit und Anstrengung beim weiteren Lernen investiert wird. Deshalb ist die Förderung metakognitiver Einschätzungen während des gesamten Lernprozesses relevant. Genaue Einschätzungen der eigenen Leistung sind notwendig, um adäquate Lernprozesse zu initiieren und die Leistung zu verbessern (z. B. Thiede, Anderson & Therriault, 2003). In der bisherigen Forschung zeigte sich allerdings, dass Personen ihre eigene Leistung häufig falsch einschätzen, v. a. beim Lernen mit Texten (z. B. Dunlosky & Lipko, 2007; Dunlosky, Hartwig, Rawson & Lipko, 2011). Als Folge ungenauer Einschätzungen können dysfunktionale Lernprozesse sowie schlechtere Leistung resultieren. Beispielsweise kann eine Überschätzung der eigenen Leistung zur vorzeitigen Beendigung des Lernprozesses führen, was sich wiederum negativ auf die Leistung auswirkt (Dunlosky & Rawson, 2012).

Ein möglicher Grund für ungenaue Einschätzungen ist die Verwendung von *Cues* (Hinweisreizen), die keine validen Prädiktoren für die Leistung sind. Ein solcher Cue könnte Fluency sein. Fluente Texte sind leicht zu verarbeiten und sollten deshalb auch als leicht zu lernen und als leicht zu erinnern eingeschätzt werden (siehe ELER-heuristic: „Easily Learned, Easily Remembered"; Miele, Finn & Molden, 2011). Umgekehrt sollten Texte, deren Verarbeitungsflüssigkeit beispielsweise durch eine disfluente Schriftart (z. B. Mistral) oder durch Löschung von Buchstaben (z. B. B_chst_b_n) herabgesetzt ist, als schwerer zu lernen und zu erinnern eingeschätzt werden. Während diese Annahme v. a. dann bestätigt wurde, wenn Personen sowohl mit fluentem als auch disfluentem Material lernen, zeigen einige Studien, dass Fluency keinen Einfluss auf Einschätzungen hat, wenn Fluency zwischen Personen manipuliert wird (z. B. Susser, Mulligan & Besken, 2013, Experiment 1). Susser et al. (2013) folgerten daraus, dass Fluency im letzteren Fall ein weniger offensichtlicher Cue ist. Allerdings ist die Anzahl an Studien, die Fluency zwischen Personen manipuliert haben, begrenzt. In dieser Dissertation wurde sowohl in Experiment 1 für die Löschung von Buchstaben, als auch in Experiment 4 für die Schriftart Mistral bestätigt, dass auch eine Manipulation zwischen Personen Einschätzungen beeinflussen kann, wobei jeweils die Interaktion mit der Art der Einschätzung und dem Zeitpunkt im Lernprozess berücksichtigt werden müssen.

Eine weitere Bedingung, unter der Fluency die Einschätzungen beeinflusst, wurde in Experiment 2 und 3 untersucht. Ziel dieser Experimente war es u. a. herauszufinden, ob Disfluency zu einer analytischeren Überwachung des eigenen Lernprozesses führt und ob diese auch für nachfolgendes fluentes Material anhält. Als Folge analytischer Überwachungsprozesse für disfluentes und fluentes Material sollte Fluency nicht länger als Cue für Einschätzungen verwendet werden. Die Ergebnisse bestätigen dies weitgehend sowohl für die Löschung von Buchstaben (Experiment 2) als auch für die Verwendung der Schriftart Mistral (Experiment 3). Wenn Studierende zuerst mit einem disfluenten und anschließend mit einem fluenten Text lernen (*Kontrast disfluent-fluent*), wird sowohl die Schwierigkeit als auch die Leistung des fluenten und des disfluenten Textes als gleich hoch eingeschätzt. Auch hier gibt es allerdings Interaktionen mit der Art der Einschätzung und dem Zeitpunkt im Lernprozess. Wird dagegen zuerst mit einem fluenten und anschließend mit einem disfluenten Text gelernt (*Kontrast fluent-disfluent*), findet man den *Fluency Effekt*, d. h. einen Unterschied zwischen fluentem und disfluentem Text, bei den Einschätzungen.

Neben der Untersuchung, unter welchen der genannten Bedingungen Fluency Effekte auf verschiedene Arten von Einschätzungen im Lernprozess gefunden werden, ist es ein weiteres Ziel dieser Dissertation zu untersuchen, ob Disfluency durch die Initiierung analytischer Überwachungsprozesse auch zu genaueren Einschätzungen führt. Bei der Genauigkeit von Einschätzungen kann zwischen der *absoluten* und der *relativen* Genauigkeit unterschieden werden (z. B. Dunlosky & Metcalfe, 2009). Während die absolute Genauigkeit die Überschätzung oder Unterschätzung der eigenen Leistung beschreibt, kennzeichnet die relative Genauigkeit, wie gut zwischen eigenen Leistungen, z. B. beim Lernen aus unterschiedlichen Texten, unterschieden werden kann. Die Befunde aus Experiment 3 und 4 zeigen, dass Mistral die oft gefundene Überschätzung reduziert, wenn die Fluency zwischen Personen manipuliert wird oder zuerst mit fluentem und anschließend mit disfluentem Text gelernt wird. Abhängig von der Art der Einschätzung und des Zeitpunkts im Lernprozess kann Mistral sogar zu Unterschätzung führen und auch die relative Genauigkeit verbessern (Experiment 4).

Eine Verbesserung der Genauigkeit von Einschätzungen ist v. a. dann sinnvoll, wenn genauere Einschätzungen auch zu besseren Lernprozessen und zu besserer Leistung führen (Dunlosky & Rawson, 2012; Efklides, 2012). Dies wurde in Experiment 4 untersucht. Hier zeigte sich allerdings, dass genauere Einschätzungen nicht zu einer besseren Leistung führen. Deshalb sollten in der weiteren Forschung Interventionen entwickelt werden, die nicht nur die Genauigkeit von Einschätzungen, sondern auch deren Transfer in bessere Lernprozesse und in bessere Leistung fördern.

Insgesamt ist es also das Ziel dieser Dissertation zu untersuchen, wann Disfluency verschiedene Einschätzungen im Lernprozess beeinflusst, ob Disfluency zu analytischeren und genaueren Einschätzungen führt, und ob genauere Einschätzungen letztendlich in besserer Leistung resultieren. Die Verbindung der Metakognitions- mit der Fluency-Forschung ermöglicht dabei, nicht nur Theorien zur Metakognition sondern auch zur Fluency zu spezifizieren. Wie die Befunde zeigen, müssen nicht nur verschiedene Arten der Fluency und des Designs, sondern auch verschiedene Arten von Einschätzungen und deren Genauigkeit, sowie der Zeitpunkt der Einschätzung im Lernprozess berücksichtigt werden. Die Kenntnis von Zusammenhängen und Bedingungen sowie von zugrundeliegenden Mechanismen ist wichtig, um bisherige inkonsistente Befunde systematisieren zu können und schließlich lernförderliche Interventionen ableiten zu können (Efklides, 2012).

1 Metacognition and Disfluency

In this thesis, metacognition research is connected with fluency research. Thereby, the focus lies on how disfluency can be used to improve metacognitive monitoring (i.e., students` judgments during the learning process). Improving metacognitive monitoring is important in educational contexts (De Bruin & van Gog, 2012; Dunlosky & Metcalfe, 2009; Efklides, 2012) in order to foster performance. This is also supposed by different theories about metacognition and self-regulated learning (Boekaerts, 1997; Nelson & Narens, 1990; Winne & Hadwin, 1998; Zimmerman, 1990). Based on these theories, accurate monitoring is necessary to initiate adequate control that results in better performance (Thiede, Anderson, & Therriault, 2003). However, previous research shows that students are often not able to accurately monitor their learning with meaningful text material (e.g., Dunlosky & Lipko, 2007; see also Glenberg, Wilkinson, & Epstein, 1982, for the *illusion of knowing*). For example, students often judge their performance too highly (e.g., Dunlosky, Hartwig, Rawson, & Lipko, 2011). One reason for inaccurate monitoring is that students use cues for their judgments that are not valid predictors of their performance. Because fluency might be such a cue, the first aim of this thesis is to investigate under which conditions fluency is used as a cue for judgments during the learning process and how disfluency can be used to improve monitoring accuracy.

Moreover, there is very little research that has tested if better monitoring indeed leads to better performance (see Thiede et al., 2003, for an exception). However, evidence supporting this connection is necessary to justify interventions that improve monitoring accuracy (Dunlosky & Rawson, 2012; Efklides, 2012). Therefore, another aim of this thesis is to test if improved monitoring accuracy leads to improved performance. This investigation can further contribute to disclosing conditions for the *fluency effect* (i.e., the difference between fluent and disfluent material) on performance. In fluency research, there are some open questions regarding the conditions, under which fluency affects performance. Evidence for the fluency effect on performance is inconsistent (see Kühl & Eitel, 2016, for an overview), and research is far from being able to derive compelling conclusions. Because accurate judgments are supposed to affect performance, metacognitive processes might be a mediator for fluency effects on performance. Thus, fluency effects on monitoring might be a prerequisite for fluency effects on performance. However, in this thesis, the focus is not on fluency effects on performance. Rather, the focus is on the conditions when fluency affects monitoring, how disfluency can be used to improve monitoring, and if better monitoring leads to better control and performance (which could explain fluency effects on performance). Thereby, meaningful text material instead of word-pairs is investigated to be able to transfer findings to educational contexts (Schwartz & Efklides, 2012).

1.1 Basic Theoretical Concepts

Before connecting metacognition with fluency research and deriving the theoretical model of this thesis, the concepts of metacognition and fluency are described. Because this thesis investigates fluency and its effects on metacognition when learning with meaningful text material (also known as *metacomprehension*, Maki & Berry, 1984; see also Maki & McGuire, 2002), these concepts are referred to the concept of text-processing (Kintsch, 1994; see also Redford, Thiede, Wiley, & Griffin, 2012). This is also addressed in the papers of this thesis.

1.1.1 Metacognition and text-processing. Metacognition is defined as thinking about one´s own thoughts and cognitions (Flavell, 1979). Two important components of metacognition are monitoring and control (e.g., Nelson & Narens, 1990; Winne & Hadwin, 1998). "[M]onitoring refers to the subjective assessment of one´s own cognitive processes and knowledge, whereas control refers to the processes that regulate cognitive processes and behavior" (Koriat, Ma'ayan, & Nussinson, 2006, p. 38; see also Ball, Klein, & Brewer, 2014; Schwartz & Perfect, 2002). To monitor their learning, students can make different types of judgments during the learning process (Nelson & Narens, 1990). Different types of judgments can represent different aspects of monitoring (see also Leonesio & Nelson, 1990). Furthermore, they can be related to different levels of text-processing that occur at different stages of the learning process and blend into each other (see Figure 1).

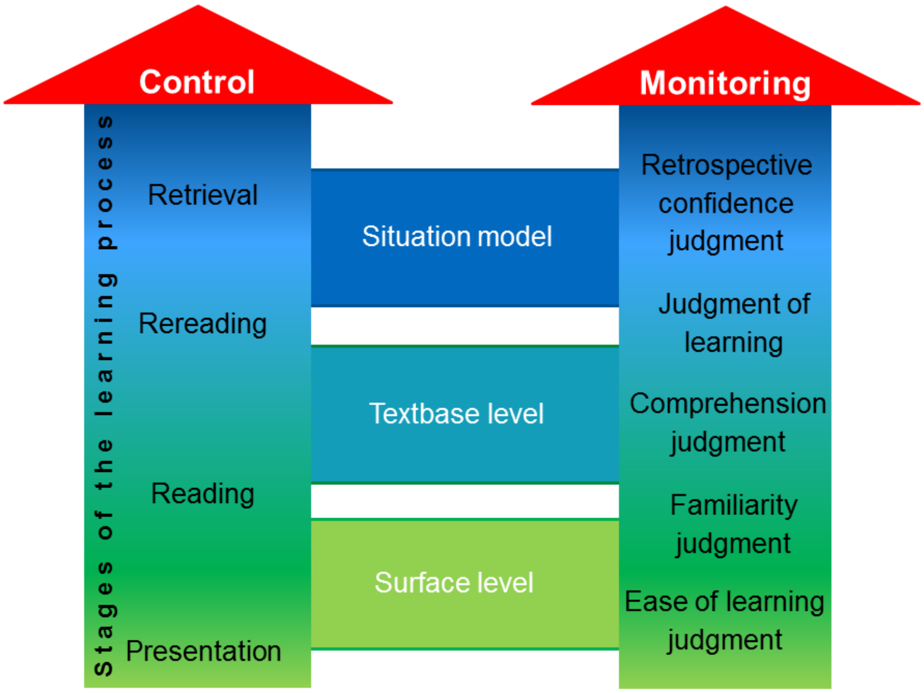

Figure 1. Relation of metacognitive processes to levels of text-processing during the learning process.

At early learning stages, students can make *ease of learning judgments* (EOL judgments) by judging how difficult it will be to learn the text. When making EOL judgments, students usually have not studied the material yet, but the material has just been presented to them. Thus, EOL judgments should refer to the *surface level* of text-processing (see Figure 1). At this level of text-processing, the students might have deciphered words or phrases but have not yet reached a deep understanding of the text (e.g., Kintsch, 1994). EOL judgments are of high importance for making plans for further learning. However, they were often not considered in previous research (Jönsson & Lindström, 2010). This is also true for *familiarity judgments*, although domain familiarity can affect not only further learning but also monitoring of the further learning process (Shanks & Serra, 2014). Familiarity judgments capture students` familiarity with the material (e.g., Westerman, Lanska, & Olds, 2015).

During the knowledge acquisition process, students should read and also reread the text in order to gain a deeper understanding of the text. When learning word-pairs, it is sufficient to recall the word-partner to reach the criteria of learning. Inversely, to learn a text, simply reproducing the wording of this text is not sufficient to understand the text. Encoding the wording (surface level) is necessary when learning with texts. But, students further have to build sense within and between sentences (*textbase level*) and to connect the ideas from the text into existing knowledge structures in order to memorize them (*situation model*, Kintsch, 1994; see also Redford et al., 2012). Hence, the textbase level as well as the situation model is necessary to reach a deeper understanding of the text. To assess monitoring of the textbase level, students can be asked for *comprehension judgments* immediately after learning. In this case, comprehension judgments capture students` actual comprehension level. *Judgments of learning* (JOLs) further refer to the situation model because these judgments capture students` predictions about their future performance, for which memorizing of the understood content is required (see Figure 1).

Retrieving the memorized content and ideas of the text might also be the basis for *retrospective confidence judgments* (RC judgments). Inversely to JOLs and also to EOL judgments, which are *prospective judgments*, *retrospective judgments* are usually made during a knowledge test by judging the confidence in a retrieved answer (also called *postdictions*, see Pieschl, 2009). Because the situation model is required to answer the questions in the test, RC judgments should refer to this situation model (see Figure 1). However, Dinsmore and Parkinson (2013) found that text characteristics are also used as cues for these judgments.

Summing up, to monitor different levels of text-processing, students should make different types of judgments at different learning stages of the learning process. Because this thesis investigates the effects of fluency on different types of judgments, the concept of fluency is described next.

1.1.2 Fluency. Fluency can be defined as the ease or speed of processing a stimulus (e.g., Alter & Oppenheimer, 2009; Lanska, Olds, & Westerman, 2014; Oppenheimer, 2008; Schwarz, 2010) and thus is also called *processing fluency*. Whereas *fluent* material is easy to process, *disfluent* material is difficult to process. Thus, the term fluency is used as a generic term that describes the continuum from fluent to disfluent processing. Moreover, fluency can also signify the pole of this continuum that refers to material that is easy to process (instead of difficult to process). The difference between fluent and disfluent material is called the fluency effect.

In the literature (e.g., Alter & Oppenheimer, 2009; Schwarz, 2010), different types of fluency are distinguished. With respect to processing fluency, Schwarz (2010) distinguishes between *perceptual fluency* and *conceptual fluency*, which are in the focus of this thesis. When reading a text, perceptual fluency refers to the ease of identifying words. A possible manipulation of perceptual fluency is font type (e.g., Mistral), which affects the deciphering and, hence, the ease of identifying words. Thus, perceptual fluency affects the surface level of text-processing. Conceptual fluency refers to the ease of identifying the meaning of words and integrating them into existing knowledge structures. A possible manipulation of conceptual fluency is letter deletion (e.g., l_tt_r d_l_ti_n). When learning with a deleted letter text, students have to generate the deleted letters by making sense of the context in order to identify the meaning of the words (Bjork & Storm, 2011; see also Bjork & Yue, 2016). Thus, conceptual fluency affects higher levels of text-processing, which are also referred as *conceptual processing* or *propositional processing*.

Although perceptual and conceptual fluency might affect different levels of text-processing, even perceptual disfluency has been found to improve performance of higher levels of text-processing (e.g., Eitel, Kühl, Scheiter, & Gerjets, 2014, Experiment 1, for transfer; see also Diemand-Yauman, Oppenheimer, & Vaughan, 2011, Experiment 2, for exams). However, research regarding the fluency effect on performance is inconsistent. Particularly for perceptual fluency, research is far from being able to predict fluency effects on performance, although some conditions have been investigated (Kühl & Eitel, 2016, for an overview). In Appendix A (Table A1), an overview with important studies that investigated the effects of fluency on performance is presented, thereby focusing on perceptual fluency but also including studies about letter deletion in texts, which is relevant for this thesis.

For conceptual fluency, some conditions under which deleted letters improve performance are already described by McDaniel and Butler (2011; see also Bjork, DeWinstanley, & Storm, 2007). Based on the framework of *desirable difficulties* (Bjork, 1994), McDaniel and Butler (2011) systemized conditions under which difficulties like letter deletion are desirable because they improve performance and conditions under which difficulties are not beneficial or even obstructive for performance: Whether difficulties are desirable, depends on the interplay of encoding operations, the learning material, the test, and learner characteristics. Most relevant for this thesis, McDaniel and Butler (2011) stated that deleted letters are only useful for material that stimulates different encoding operations than letter deletion. Expository texts as well as deleted letters initiate propositional processing, and hence, deleted letters should not foster performance of expository texts. This is supported by findings from McDaniel, Einstein, Dunay, and Cobb (1986, Experiment 2) that deleted letters improved performance only for fairy tales but not for descriptive/expository texts. However, findings by Rawson and Dunlosky (2002, Experiment 4; see also Maki, Foley, Kajer, Thompson, & Willert, 1990, Pilot and Experiment 1) show that another aspect of encoding seems to matter: When students write-in deleted letters in expository texts performance can be improved, whereas this often is not the case when mentally filling in deleted letters. Moreover, learning with a deleted letter text seems more likely to increase performance in a within-person design (e.g., McDaniel, 1984) than in a between-person design (e.g., Maki et al., 1990, for an overview).

Summing up, in prior research, fluency effects on performance were in the focus. Because of the inconsistent findings, conditions for fluency effects on performance have been investigated. However, there is less research about fluency effects and their conditions on monitoring. This is the case for both perceptual and conceptual fluency, although metacognition might be a mediator for the effects of both types of fluency on performance. Fluency effects on monitoring are addressed in more detail next.

1.2 Fluency as a Cue for Monitoring

As described in paper I (see also paper II, III, and IV), the assumption that fluency is used as a cue for monitoring can be derived from different theories. Koriat (1997) states in his *cue utilization approach* that fluency is an experience-based cue for metacognitive judgments. Similarly, Schwartz and Efklides (2012; see also Efklides, 2009) describe judgments as metacognitive experiences, which is similar to the mechanism described by *control-based monitoring* (Koriat, 2012; Koriat et al., 2006). Perceptually as well as conceptually disfluent material is difficult to process; students judge it as difficult to remember. Inversely, students judge material that is easy to process as easy to remember, which is stated in the *ELER*-heuristic ("Easily Learned, Easily Remembered", Koriat, 2008; Miele, Finn, & Molden, 2011) and the *ease of processing heuristic* (Kornell, Rhodes, Castel, & Tauber, 2011). Thus, students infer the difficulty of learning and remembering from disfluency in processing.

There is evidence that disfluent material requires longer reading- or study-times than fluent material (e.g., McDaniel, 1984, Pilot of Experiment 1; McDaniel et al., 1986, Experiment 1 and 2, for conceptual fluency; Eitel & Kühl, 2016; Miele & Molden, 2010, Experiment 3; Sanchez & Jaeger, 2015, Experiment 1 and 2, for perceptual fluency), supporting the assumption that fluency affects control of cognitive processes: Processing of disfluent material is slower and less automatic than processing of fluent material. Moreover, there is evidence that fluency affects monitoring: Studies found that fluency affects the magnitude of metacognitive judgments (see Appendix A, Table A1, for an overview). This has been found for different types of fluency, for different types of judgments, and when learning with different learning material. For example, students judge disfluent compared with fluent material as less familiar (Westerman et al., 2015; see also Lanska et al., 2014, for an overview), and they judge disfluent expository texts as more difficult to comprehend (e.g., Maki et al., 1990, Experiment 1; Rawson & Dunlosky, 2002, Experiment 4, for conceptual fluency; Miele & Molden, 2010, Experiment 3, for perceptual fluency). Furthermore, students predict lower performance for disfluent than for fluent word-lists (e.g., Kornell et al., 2011; Yue, Castel, & Bjork, 2013, Experiment 1a, 2a, 2b, and 3, for perceptual fluency), word-pairs (e.g., Magreehan, Serra, Schwartz, & Narciss, 2016, Experiment 4 and 5; Mueller, Tauber, & Dunlosky, 2013, Experiment 2, for perceptual fluency), and texts (e.g., Rawson & Dunlosky, 2002, Experiment 4, for conceptual fluency).

However, besides this evidence, there are also some studies that found no fluency effect on judgments when learning with word-lists (e.g., Sungkhasettee, Friedman, & Castel, 2011, Experiment 1 and 2; Yue et al., 2013, Experiment 1b, for perceptual fluency), word-pairs (e.g., Magreehan et al., 2016, Experiment 1, 2, and 3, for perceptual fluency), or texts (e.g., Maki et al., 1990; McDaniel et al., 1986, for conceptual fluency; Faber, Mills, Kopp, & D'Mello, 2016, for perceptual fluency). In some of these studies (e.g., Maki et al., 1990, Experiment 1), students seem to have used retrieval as an additional cue besides fluency for their judgments: Performance (i.e., retrieval) in these studies was better for the disfluent than for the fluent material. Retrieval as a cue for judgments might have affected judgments in an opposite direction than fluency, and thus, no fluency effects on judgments have been found. This assumption is confirmed by Rawson and Dunlosky (2002, Experiment 4) who asked students to mentally fill in deleted letters. This fluency-manipulation did not affect performance (see section 1.1.2), but it affected JOLs. Therefore, to be able to investigate the effects of fluency on judgments, it is important to ensure that fluency does not improve performance at the same time, although improving performance is desirable in educational contexts. In other studies that found no fluency effects on judgments, fluency was manipulated between-person (e.g., Faber et al., 2016, for texts; Magreehan et al., 2016, Experiment 1 and 2, for word-pairs; Maki et al., 1990, Experiment 2, for texts; McDaniel et al., 1986, for texts; Susser, Mulligan, & Besken, 2013, Experiment 1 and 3, for pure word-lists; Yue et al., 2013, Experiment 1b, for word-lists). Therefore, students experienced either fluency or disfluency in processing, but they did not experience a contrast between fluent and disfluent learning material as they did in within-person designs. Thus, the experience of fluency might not have been an obvious cue for judgments in these studies. This assumption is supported by the findings by Susser et al. (2013, Experiment 1; see also Magreehan et al., 2016, for word-pairs) who found fluency effects on judgments for mixed word-lists (within-person design) but not for pure word-lists (between-person design). Nevertheless, there might be fluency-manipulations that might even be obvious in between-person designs, as also suggested by Susser et al. (2013). Furthermore, this might depend on the type of judgment and the learning stage. For example, when judging comprehension, fluency has been found to be used as a cue (e.g., Miele & Molden, 2010, Experiment 3), whereas JOLs are often not affected by fluency (e.g., Maki et al., 1990, Experiment 2). However, systematic research that investigates conditions for fluency effects on different types of judgments is rare. This is especially the case when looking for studies that used texts as learning material. Previous research focused on the effects of perceptual disfluency on JOLs when learning with word-lists or word-pairs, and as noted by Susser et al. (2013), fluency usually was manipulated within-person. Thus, an open question is if fluency is also used as a cue for different types of judgments during the learning process, when fluency is manipulated between-person and when learning with texts.

Summing up, previous research shows that fluency can be a cue for judgments. However, under some conditions, fluency is not used as a cue for judgments. Next, it will be outlined that fluency could be more than just a cue for judgments. Disfluency might be a way to activate analytic monitoring, which could also be a condition for fluency effects on judgments.

1.3 Disfluency and Analytic Monitoring

In this section, the effects of disfluency on analytic processes are described (see also paper II, III, and IV). These analytic processes are specified with respect to metacognitive monitoring and control. This specification is necessary to understand why the activation of analytic monitoring by disfluency can be a condition for finding fluency effects on judgments in empirical studies: Once activated, analytic monitoring might sustain for succeeding fluent material, and as a consequence, no fluency effects on monitoring should be found.

Disfluency-theories (e.g., Alter, Oppenheimer, Epley, & Eyre, 2007) suppose that fluency activates quick, effortless, and surface processes (*System 1 processes*), whereas disfluency activates slow, effortful, and analytic processes (*System 2 processes*). Based on this assumption, disfluency-theory predicts higher performance for disfluent than for fluent material. However, prior research does not fully support this assumption (e.g., Meyer et al., 2015, for an overview) because findings on fluency effects on performance are inconsistent (see section 1.1.2 and Appendix A). Similarly, there are inconsistent findings regarding the assumption that disfluency activates effortful processes. *Perceived difficulty* and *mental effort* are often used (Schmeck, Opfermann, van Gog, Paas, & Leutner, 2015) as subjective measures for cognitive resources students invest during learning (Rey & Nieding, 2010). In some studies (see Appendix A, for an overview), students reported that disfluent material was more difficult to read or to learn (perceived difficulty) than fluent material (e.g., Miele & Molden, 2010, Experiment 3; Sanchez & Jaeger, 2015, Experiment 1 and 2; Song & Schwarz, 2008, Experiment 1, 2, and 3). Moreover, students reported higher mental effort for disfluent than for fluent material (Eitel et al., 2014, Experiment 1). However, some studies found no significant fluency effects on mental effort (e.g., Eitel & Kühl, 2016; Eitel et al., 2014, Experiment 2–4) or perceived difficulty (e.g., Eitel & Kühl, 2016; Eitel et al., 2014, Experiment 1–4).

Thus, there are some discrepancies between the predictions of disfluency-theory and the empirical findings. This illustrates the need to specify the processes described in disfluency-theory. Alter et al. (2007; see also Diemand-Yauman et al., 2011) suggest that the processes that are activated by disfluency are metacognitive processes. However, these metacognitive processes seem to be mainly about analytic monitoring and not necessarily about metacognitive control. This assumption can be derived from theories about metacognition (e.g., Nelson & Narens, 1990; Winne & Hadwin, 1998): Performance depends on the quality of control of cognitive processes. If disfluency would activate analytic control, performance should be improved when compared to fluent material. But, as described in section 1.1.2 (see also Appendix A), this is often not the case. However, disfluency might activate analytic monitoring that can result in improved performance when students implement it into analytic control (see section 1.5 and paper IV for details regarding *monitoring-based control*). The assumption that disfluency activates analytic monitoring (which potentially activates analytic control and improves performance) is supported by findings by Alter et al. (2007, Experiment 2). They found that arguments in a persuasion task were more analytically processed when the masthead of the text presenting the arguments was disfluent than when it was fluent. However, the arguments were presented in a fluent way. Thus, fluency has not directly led to analytic control of cognitive processes. Inversely, the disfluency of the masthead activated analytic monitoring that sustained, even when reading the fluent arguments, and enabled analytic control of cognitive processes of the succeeding fluent arguments.

Therefore, the *type of contrast* between fluent and disfluent material seems to be a condition for the fluency effect on judgments. If analytic monitoring that is activated by disfluency remains for succeeding fluent material, monitoring should be analytic for disfluent and succeeding fluent text (see Figure 2a). As a result, no fluency effect on monitoring should be found when students first learn with a disfluent and afterwards with a fluent text (*contrast disfluent-fluent*).

Inversely, when students first learn with a fluent and afterwards with a disfluent text (*contrast fluent-disfluent*), only monitoring of the disfluent but not of the fluent text should be analytic, whereas monitoring of the fluent text should be surface (see Figure 2b). The assumption that fluency-contrasts moderate fluency effects is also supported by findings from Dreisbach and Fischer (2011). They found significantly longer reaction-times for fluent after disfluent tasks compared to fluent after fluent tasks.

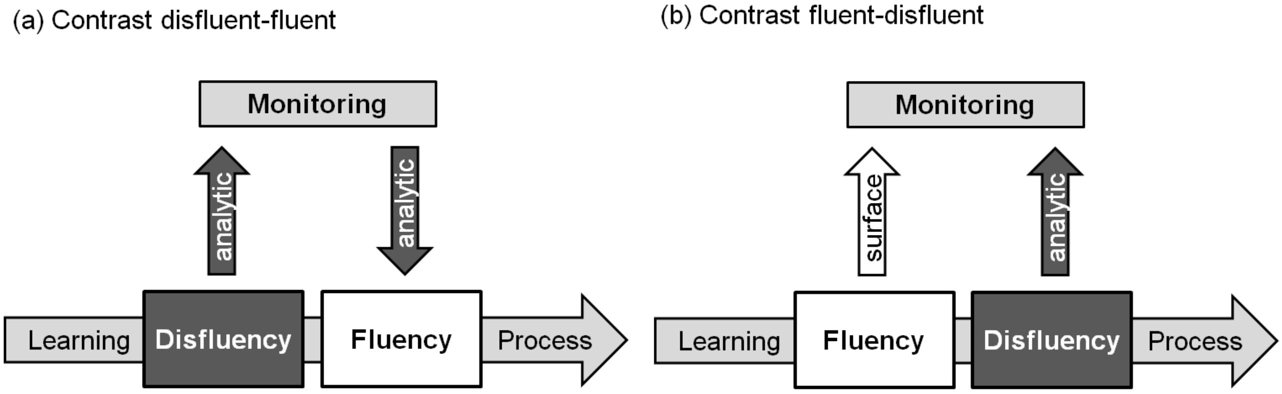

Figure 2. Analytic monitoring for (a) contrast disfluent-fluent and (b) contrast fluent-disfluent.

To sum up, disfluency-theory supposes that fluency leads to quick, effortless, and surface processes, whereas disfluency leads to slow, effortful, and analytic processes (e.g., Alter et al., 2007). However, these analytic processes seem to be about metacognitive monitoring: Disfluency is a way to foster analytic monitoring, which remains for succeeding material. Analytic monitoring should enable students to use cues for their judgments that are valid for performance (i.e., conceptual cues like the difficulty of described concepts). Thus, disfluency should be a way to improve monitoring accuracy. Next, different types of monitoring accuracy and how they can be affected by disfluency (see also paper III and IV) are described.

1.4 Disfluency and Monitoring Accuracy

Monitoring accuracy is defined as the relation between students´ judgments and their performance (Dunlosky & Metcalfe, 2009; Pieschl, 2009). When students use cues that are valid predictors for their performance (*cue diagnosticity*), monitoring is *accurate* (e.g., Dunlosky & Thiede, 2013; see also Pyc, Rawson, & Aschenbrenner, 2014). In the literature, there are different terms and indexes for different types (or sometimes for the same type) of monitoring accuracy (Nelson, 1984; see Pieschl, 2009; Schraw, Kuch, & Gutierrez, 2013, for an overview). In this thesis, the following two types of accuracy are distinguished (see also Dunlosky & Lipko, 2007;

Dunlosky & Metcalfe, 2009; Mengelkamp & Bannert, 2010): *absolute accuracy* and *relative accuracy*.

Absolute accuracy (also known as *calibration*, see Dunlosky & Metcalfe, 2009) refers to the congruence between the level of students` judgment magnitudes and their performance. When students exactly predict the performance they show in a knowledge test (e.g., an exam), they are perfectly calibrated. When their performance in a knowledge test is higher than their judged performance, they are *underconfident*. Inversely, when students judge their performance as higher compared with their actual performance in a knowledge test, they are *overconfident*. In previous research, overconfidence has often been found in educational contexts (e.g., Baker, Dunlosky, & Hertzog, 2010, for texts; De Bruin, Kok, Lobbestael, & Grip, 2016, for classroom settings; Dunlosky et al., 2011, for concepts; Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008, for realistic/classroom settings; Eitel, 2016, for texts and pictures; Miesner & Maki, 2007, for texts) as well as in other areas (e.g., see Ehrlinger & Eichenbaum, 2016, for an overview). Overconfidence is a problem because students might not invest enough time and effort during learning when they are overconfident. Thus, it is important to find ways to reduce overconfidence. Disfluency should be a way to reduce overconfidence. As described in section 1.2, disfluent material leads to slower, less automatic processing. As a consequence, students might judge disfluent material as more difficult, and judge their performance as lower for disfluent than for fluent material. Due to lower judgments, overconfidence should be reduced if performance is the same for disfluent and fluent material. Inversely, fluent material is quickly processed, and this experience leads to the judgment that the material is easily learned and remembered. Thus, fluency might induce overconfidence (see Dunlosky et al., 2011; Ehrlinger & Eichenbaum, 2016, for other explanations of overconfidence). Moreover, as described in section 1.3, disfluency should not only be a way to lower judgments, but disfluency also is assumed to activate analytic monitoring. Due to analytic monitoring, students should use cues for their judgments that are valid for performance (e.g., conceptual cues). This should lead not only to improved absolute but also to improved relative accuracy.

Relative accuracy (also known as *resolution*; see Dunlosky & Metcalfe, 2009) is a measure for being able to discriminate between materials for which higher scores are obtained and materials for which lower scores are obtained. Thus, not the absolute performance magnitude but the ability to discriminate between high and low performance is relevant for relative accuracy. Absolute and relative accuracy are independent from each other; absolute accuracy can be high, even if relative accuracy is low and inversely, which can be illustrated by the example from Dunlosky and Lipko (2007): Students can judge that they will attain 80% of the maximal possible points in an exam, but in fact reach just 60% of maximal possible points. Although these students are overconfident, they can be able to discriminate between higher and lower scored answers, and therefore, they can reach perfect relative accuracy. However, also relative accuracy has often found to be low (see Dunlosky & Lipko, 2007; Dunlosky & Metcalfe, 2009, for an overview). Previous research focused on how to improve relative accuracy (see also Dunlosky & Lipko, 2007, for an overview). As described in section 1.3, disfluency should be a way to activate analytic monitoring. When learning with texts, analytic monitoring should enable students to realize difficulties in text-processing, which they would not realize without analytic monitoring. As reasoned in paper IV, students` aim is the comprehension and memorization of the text. Thus, analytic monitoring should foster monitoring of conceptual processing: When making judgments, this enables students to use conceptual cues, which are valid for performance. Therefore, students should be able to accurately discriminate between high and low comprehension. Hence, analytic monitoring should foster relative accuracy. This is supported by findings of higher relative accuracy of performance predictions for disfluent (deleted letter) texts than for fluent (intact) texts (e.g., Maki et al., 1990, Experiment 1 and 2). They concluded that this effect was due to increased processing for disfluent texts (see also Lin & Zabrucky, 1998, for this conclusion).

Summing up, disfluency should be a way to improve both absolute and relative monitoring accuracy at the same time, which is rarely investigated. Students should predict lower performance for disfluent than for fluent texts, and thus, overconfidence should be reduced. Furthermore, disfluency should be a way to activate analytic monitoring. Because of analytic monitoring, not only absolute but also relative monitoring accuracy should be improved. This is important in order to activate adequate control and thus, to improve performance. Next, the interplay between monitoring, control, and performance as well as the role of fluency for this interplay is described.

1.5 Fluency and the Relation Between Monitoring, Control, and Performance

Theories of metacognition and self-regulated learning (e.g., Nelson & Narens, 1990; Winne & Hadwin, 1998) assume that monitoring affects control (monitoring-based control) and thus, performance. Consequently, high monitoring accuracy should enable adequate control of cognitive processes, resulting in high performance. However, there is little research that investigates if accurate monitoring leads to adequate control and improved performance (e.g., Thiede et al., 2003, for an exception), although evidence is necessary to justify interventions that foster monitoring (Dunlosky & Rawson, 2012; Efklides, 2012). Furthermore, most of the few studies that tried to investigate this link had some shortcomings, e.g., because students did not have the chance to implement their judgments into control (see Dunlosky & Rawson, 2012, for a more detailed overview), or because the link between monitoring accuracy and performance was not investigated in an experimental design to test causality (see Thiede et al., 2003, for an overview). Moreover, some research, rather than investigating the link between monitoring accuracy and performance, investigated the link between monitoring and control, thereby using mostly words or word-pairs as learning material (see Dunlosky & Ariel, 2011; Metcalfe, 2002, 2009; Metcalfe & Kornell, 2005; Nelson & Leonesio, 1988; Nelson, Dunlosky, Graf, & Narens, 1994; Son & Kornell, 2008; Son & Metcalfe, 2000, for overviews about *study time allocation*). The *discrepancy-reduction hypothesis* states that students spend more time on material they judge as difficult, whereas the *region-of-proximal-learning hypothesis* states that students spend more time on material they judge as low or medium in difficulty than on material they judge as highly difficult (see also Dunlosky & Metcalfe, 2009, for an overview). Whereas the former seems to pertain to situations without time-pressure, the latter seems to pertain to situations in which students do not have enough time for learning (e.g., because they started to learn the night before the exam, see also Dunlosky & Metcalfe, 2009, for an overview).

Evidence that judgments can affect control when learning with words or word-pairs, however, is not sufficient to make conclusions about the connection between monitoring and control when learning with texts because learning with texts is more complex (see section 1.1). Moreover, this evidence is not sufficient to draw conclusions about the relation between accurate monitoring and performance (see also Dunlosky & Rawson, 2012). Basing control on monitoring will only lead to better performance, if monitoring is accurate. Otherwise, students do not invest time and effort into learning material that indeed (and not only judged) requires further learning and thus, performance should not be improved. Hence, further research is required to test if accurate monitoring leads to improved performance.

To be able to investigate the effect of monitoring accuracy on control and performance, at least two learning phases are required (see also paper IV). In the first phase, monitoring accuracy should be improved, and in the second phase, students should have the opportunity to reread the learning material in order to implement accurate monitoring into adequate control and better performance. In the first learning phase, only monitoring and not performance should be improved because under this condition, better performance in the second phase is caused by accurate monitoring. Thereby, it is important to foster both types of monitoring accuracy because absolute and relative accuracy might have different implications for control (see also De Bruin & van Gog, 2012; Dunlosky & Rawson, 2012): High relative accuracy might be a presupposition to select texts for rereading that indeed require further learning (*text selection*); high absolute accuracy might be a presupposition to invest enough time during rereading the material (*termination of study*). Although there are some studies that aimed to improve monitoring accuracy, the focus was often on relative accuracy, although during the last few years, there has been an increase in research investigating absolute accuracy (see Alexander, 2013; De Bruin & van Gog, 2012). However, there is little research that investigates both relative and absolute monitoring accuracy (see also Maki & McGuire, 2002). Disfluency should improve absolute and relative monitoring accuracy at the same time (see section 1.4). Moreover, perceptual fluency (and conceptual fluency under certain conditions, see section 1.1.2) should not affect performance after reading a text once because perceptual fluency affects only surface levels of text-processing. However, in the second phase, performance for the disfluent texts will be improved if students implement accurate monitoring into adequate control (monitoring-based control), e.g., by selecting texts that indeed require further learning and by spending enough time when rereading these texts. Thus, fluency effects on monitoring might be a presupposition for fluency effects on control and performance.

Summing up, to be able to investigate if better monitoring leads to better control and better performance, at least two learning phases are required. In the first phase, absolute and relative monitoring accuracy must be enhanced, and in the second learning phase, students must have the chance to implement better monitoring into better control. Disfluency enables the investigation of the causal chain from monitoring to control and performance by affecting both absolute and relative monitoring accuracy, but not performance in the first learning phase. However, by implementing improved monitoring into improved control in the second learning phase, even perceptual fluency, which does not foster higher levels of processing (see section 1.1.2), should improve performance.

1.6 Theoretical Model, Research Questions, and Hypotheses

Based on the theories and empirical findings described so far, the model for this thesis is derived. Thereby, theories about metacognition and about disfluency are combined: This model (see Figure 3) unites the effects of fluency on monitoring as well as the interplay between fluency, monitoring, control, and performance. Thereby, different types of fluency (i.e., perceptual and conceptual fluency) and different aspects of monitoring are included: fluency (see section 1.1.1) as a cue for monitoring (see section 1.2) and as a way to foster analytic monitoring (see section 1.3) and monitoring accuracy (see section 1.4). Based on this model, the main research questions and hypotheses as well as their interplay are described. Afterwards, the experiments and the specific research questions of the experiments are summarized.

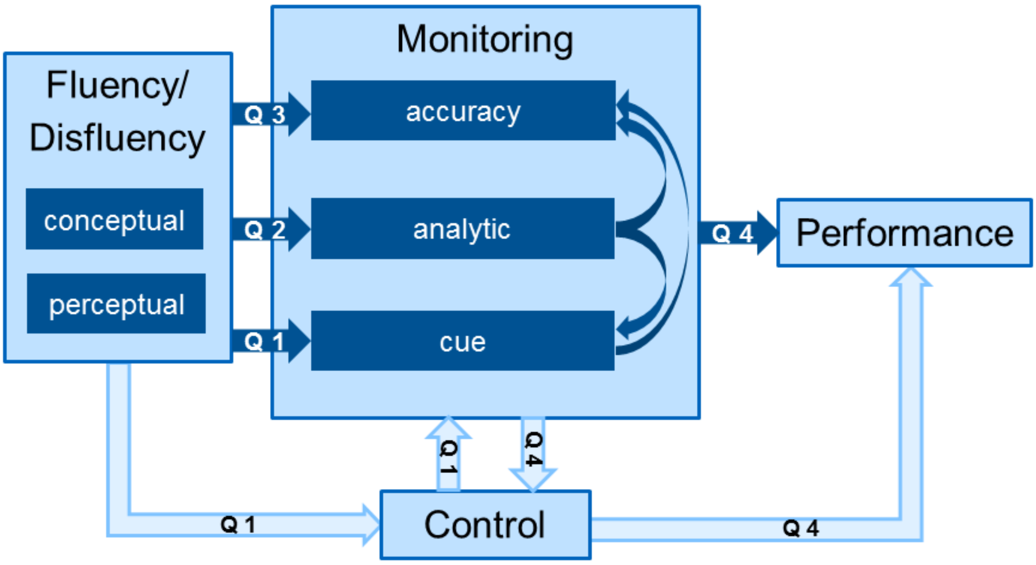

Figure 3. Theoretical model and research questions of the thesis.

The first research question (Q1) is: *Is fluency used as a cue for monitoring?* This research question focuses on the link between fluency and monitoring (see the arrows Q1 in Figure 3). Although different types of processing fluency (e.g., perceptual and conceptual fluency) might be related to different levels of processing (see section 1.1.2), all of these types of fluency should lead to slower processing. Hence, students control their learning by slowing down processing (see the arrow Q1 from fluency to control in Figure 3). The experience of slow, less automatic processing should inform students´ judgments (control-based monitoring; see the arrow Q1 from control to monitoring/cue in Figure 3). Moreover, students can also directly infer from disfluency the difficulties in learning or remembering (ELER-heuristic, see section 1.2, and arrow Q1 from fluency to monitoring in Figure 3). Therefore, fluency should be a cue for different types of judgments at different learning stages, even in a between-person design (see section 1.2).

However, these fluency effects on judgments might be moderated by the type of contrast in fluency because disfluency might activate analytic monitoring that remains for succeeding fluent material (see arrow from analytic to cue in Figure 3). Hence, the second research question (Q2) addresses the question: *Does disfluency activate analytic monitoring?* Different approaches can be used to investigate the second research question (Q2). One approach is to investigate *if the type of contrast in fluency affects the fluency effect on monitoring*, which is in the focus of the second research question (Q2). As described in section 1.3, disfluency should initiate analytic monitoring that should remain for succeeding fluent material, which was concluded from findings by Alter et al., (2007, Experiment 2). Hence, for contrast disfluent-fluent, no fluency effect on monitoring is expected. Inversely, for contrast fluent-disfluent, only monitoring of the disfluent text should be analytic, and thus, fluency effects on judgments can be expected. This is also the case when students either read only fluent or disfluent texts (between-person design). Therefore, the second research question (Q2) is related to the first research question (Q1) because the activation of analytic monitoring can affect if fluency is used as a cue for judgments or not (see arrow from analytic to cue in Figure 3).

Furthermore, the second research question is also related to the third research question (Q3) that investigates the question: *Does disfluency improve monitoring accuracy?* Thus, investigating monitoring accuracy is another approach to investigate the second research question (see arrow from analytic to accuracy in Figure 3). Due to analytic monitoring, students should use cues that are valid for their performance and thus, monitoring should be accurate (see also section 1.4). Particularly, relative monitoring accuracy can be an indicator of analytic monitoring because analytic monitoring enables the use of conceptual cues; hence students should be able to discriminate between the difficulty and their performance of different texts. Inversely, absolute monitoring accuracy can also be improved when disfluency is used as a cue for judgments and lowers judgment magnitudes (without activating analytic monitoring): Because students are often overconfident, predicting lower performance for disfluent than for fluent text can reduce overconfidence for disfluent text (see arrow from cue to accuracy of monitoring in Figure 3).

Improved absolute and relative monitoring accuracy might be necessary in order to initiate adequate control and to foster performance. The connection between accurate monitoring and performance is investigated in the fourth research question (Q4, see arrow from monitoring to performance): *Does improved monitoring improve performance?* Better performance can be expected when students implement accurate monitoring into adequate control (see section 1.5). For example, accurate monitoring should enable students to select texts for rereading that require further reading and to invest enough time and effort when rereading them. Therefore, fluency effects on performance can be mediated by monitoring-based control (see arrows Q4 in Figure 3).

Overall, this model integrates theories about metacognition and disfluency into one model. Thereby, the fluency effects on metacognitive processes as well as their interplay are specified. Furthermore, disfluency-theory is specified, which is required because of inconsistent findings in previous research.

2 Overview of the Experiments

In this section, the four experiments of this thesis are summarized. Thereby, each experiment is related to one of the four research questions (see section 1.6). More detailed information about each experiment and further research questions can be found in the papers (see Appendix G). Each paper addresses one experiment, and the order of the papers corresponds to the order of the experiments. Furthermore, in the appendixes, detailed information about the fluency-manipulations (see Appendix B), the material (see Appendix C), the knowledge tests (see Appendix D), the instruments of monitoring and control (see Appendix E) and the procedure (see Appendix F) of the experiments can be found.

2.1 Experiment 1: Fluency as a Cue for Monitoring

The paper on Experiment 1 "Fluency and metacognition – Is fluency used as a cue for judgments during and after the learning process?" can be found in Appendix G1. In this paper, the focus was on the research question Q1: Is fluency used as a cue for monitoring? Thereby, different types of judgments during and after the learning process were considered, and fluency was manipulated between-person.

As described in section 1.2, students use fluency as a cue for JOLs when learning with word-lists or word-pairs in within-person designs because they experience both fluent and disfluent processing. Inversely, there is evidence that fluency is often not used as a cue for judgments in between-person designs. However, there are only a few studies that investigated the effects of fluency on judgments in between-person designs, and there is even less research when learning with texts. Susser et al. (2013) mentioned that there could be fluency-manipulations that might be obvious as cues for judgments even in between-person designs. Letter deletion as a manipulation of conceptual fluency affects processing on higher levels of text-processing. Thus, this fluency-manipulation should be obvious as a cue for judgments in a between-person design.

In Experiment 1 of this thesis, the effects of conceptual fluency (i.e., intact text vs. deleted letter text) on different types of judgments during the learning process were investigated. Students learned with two expository texts about motivational psychology (each about 1000 words). Because conceptual fluency affects processing on higher levels of text-processing, fluency effects on judgments are expected if students learn either with only fluent or disfluent texts. To test this assumption, students made a first EOL judgment after a text-presentation for 2 seconds, a second EOL judgment as well as a first JOL after reading the text once. A second JOL was made after rereading. This procedure was done with the first and then with the second text (order of texts was randomized). Afterwards, students made RC judgments during the knowledge test on both texts.

Results from $N = 63$ students show that fluency was used as a cue for judgments after reading (EOL judgment and JOL) and after rereading (JOL). However, conceptual fluency was not used as a cue for the EOL judgment after a text-presentation for 2 seconds on the screen or for RC judgments during the test. One reason for these findings is that the experience of disfluency was not obvious when making the EOL judgment or the RC judgments. A text-presentation for 2 seconds seems to be insufficient to initiate conceptual processing and thus, conceptual fluency did not affect EOL judgments. Moreover, when making RC judgments during the test, students seem not to experience disfluency because performance was equal between fluent and disfluent texts, and therefore, the experience of fluency during the test seems to be equal. These results indicate that fluency effects on judgments depend on the type of judgment and the learning stage. Moreover, they show that there are fluency-manipulations that can be obvious in between-person designs at least after reading and rereading texts.

2.2 Experiment 2: Fluency and Analytic Monitoring

The paper on Experiment 2 "Judgments during the learning process – Does disfluency activate analytic monitoring not only for disfluent but also for succeeding fluent text?" can be found in Appendix G2. This experiment was designed to answer the research question if disfluency is a way to activate analytic monitoring that remains for succeeding fluent text (Q2). Thereby, disfluency-theory is specified regarding the processes that are activated by disfluency: Disfluency-theory supposes that disfluency leads to analytic processes and thus to better performance than fluency. However, often no fluency effects on performance are found (see section 1.1.2). Therefore, disfluency is assumed to activate analytic monitoring, but not necessarily analytic control of cognitive processing (see section 1.3).

If disfluency activates analytic monitoring that remains for succeeding fluent text, students` monitoring should be similar for the disfluent and the fluent text, and thus, fluency effects on judgments should be reduced if students first learn with a disfluent text and afterwards with a fluent text (contrast disfluent-fluent). Inversely, if students first learn with a fluent and afterwards with a disfluent text (contrast fluent-disfluent), monitoring of the fluent text should not be analytic. Thus, fluency effects on judgments should be found. Hence, this experiment uses a new approach to investigate the effects of disfluency on metacognition: Analytic control of cognitive processes is inferred from the fluency effect on performance, and analytic monitoring is inferred from moderated fluency effects on judgments. This approach enables an investigation of whether disfluency affects not only monitoring of disfluent but also of succeeding fluent material. If analytic monitoring remains for succeeding fluent text, the type of contrast in fluency will moderate fluency effects on judgments.

The fluency-manipulation, the texts, the instruments, and the procedure were the same as in Experiment 1 (see also Appendix B to F). However, fluency (intact text vs. deleted letter text) was manipulated within-person, and type of contrast (contrast fluent-disfluent vs. contrast disfluent-fluent) was manipulated between-person.

Results from $N = 65$ students in Experiment 2 widely confirm the assumptions: Fluency effects on all types of judgments during the learning process were found when students first learned with a fluent and afterwards with a disfluent text (contrast fluent-disfluent). For contrast disfluent-fluent, no fluency effect was found for JOLs after reading and rereading or for RC judgments. For the EOL judgment after a text-presentation for 2 seconds, a fluency effect was found for both types of contrast, and for the EOL judgment after reading, an inverted fluency effect was found for contrast disfluent-fluent: Students judged the fluent text as more difficult than the disfluent text. Performance was not significantly affected by fluency. Thus, disfluency might be a way to activate analytic monitoring that remains for succeeding fluent text. Inversely, disfluency did not activate analytic control of cognitive processing, which supports the derived specification of disfluency-theory. Moreover, for the EOL judgment after a first reading of the text, the fluency effect can even be inverted. This might be due to hyper-compensation, which requires further research. Hence, in Experiment 2, as in Experiment 1, the type of judgment and the learning stage also seem to play a role for fluency effects on judgments.

2.3 Experiment 3: Fluency and Monitoring Accuracy

The paper on Experiment 3 “Fostering analytic metacognitive processes and reducing overconfidence by disfluency: The role of contrast effects” is published in the Journal Applied Cognitive Psychology (see Appendix G3). The focus of Experiment 3 was on the effects of disfluency on monitoring accuracy (Q3). More specifically, it was tested if the type of contrast moderates the fluency effect on judgments and their accuracy. Thus, Experiment 3 is built on Experiment 2 but goes beyond the scope of Experiment 2 by using monitoring accuracy as a further approach to investigate analytic monitoring. Moreover, in Experiment 3, perceptual instead of conceptual fluency was investigated to test if manipulating fluency on a lower level of text-processing can also lead to analytic and thus, to more accurate monitoring. Furthermore, in order to gain deeper insights into students´ control of cognitive processes, in Experiment 3, students were asked about their invested mental effort and perceived difficulty.

As described in section 1.4, students are often overconfident when predicting their performance. Analytic monitoring should enable students to accurately predict their performance and thus, to reduce overconfidence. Because analytic monitoring remains for succeeding material, overconfidence should be reduced not only for disfluent but also for succeeding fluent texts.

Thus, no fluency effect on monitoring accuracy should be found for contrast disfluent-fluent. Inversely, when students first learn with a fluent and afterwards with a disfluent text (contrast fluent-disfluent), a fluency effect on monitoring accuracy can be expected.

Students learned with two texts about social psychology (each about 1000 words). Times New Roman (16 point) was used for fluent text and Mistral (18 point) was used for disfluent text. Like in Experiment 2, fluency was manipulated within-person and type of contrast was manipulated between-person. Students made an EOL judgment and a familiarity judgment after reading a text once. After rereading the text, a comprehension judgment and a JOL was made. In order to investigate control of cognitive processes in more detail, students further reported their perceived difficulty and invested mental effort. The same procedure was carried out with the second text, and then a knowledge test with RC judgments on the texts followed. In this experiment, $N = 75$ students participated because smaller effect sizes were expected for perceptual fluency than for conceptual fluency and because monitoring accuracy was used as an additional variable.

Results in Experiment 3 revealed that the type of contrast moderates not only the fluency effect on judgment magnitudes (see also Experiment 2) but also on absolute monitoring accuracy. For contrast fluent-disfluent, we found perfect absolute monitoring accuracy for the disfluent text, whereas students were overconfident for the fluent text. This was the case even though performance was worse for disfluent than for fluent texts. For contrast disfluent-fluent, there was no significant difference between the fluent and the disfluent text for absolute monitoring accuracy. Moreover, for contrast disfluent-fluent, students were overconfident for both texts, although overconfidence was descriptively reduced for the disfluent and the succeeding fluent text compared to the fluent text for contrast fluent-disfluent. Students` ratings about perceived difficulty and mental effort were also moderated by the type of contrast: Whereas a fluency-effect was found for contrast fluent-disfluent, no fluency effect was found for contrast disfluent-fluent. Inversely, for performance, only a mariginal moderation effect of type of contrast was found. Univariate analyses found lower performance for the disfluent than for the fluent text for contrast fluent-disfluent, whereas performance was comparable between the fluent and the disfluent text for contrast disfluent-fluent. Hence, for contrast disfluent-fluent, not only students` performance but also their judgments were almost comparable between the fluent and the disfluent text. Therefore, students might be able to judge the relation within their performance between the fluent and the disfluent text, but they were not able to accurately judge the absolute performance magnitude. However, to investigate the effects of disfluency on relative monitoring accuracy, more than two texts should be used. In Experiment 4, relative monitoring accuracy is used as another approach to investigate analytic monitoring.

2.4 Experiment 4: Fluency and the Relation Between Monitoring, Control, and Performance

The paper on Experiment 4 “Metacognitive judgments and disfluency – Does disfluency lead to more accurate judgments, better control, and better performance?” is published in the Journal Learning and Instruction (see Appendix G4). The aim of this experiment was to test if improved monitoring leads to better performance (Q4). Thereby, perceptual disfluency was used to improve monitoring accuracy.

Assuming that disfluency is a way not only to lower judgment magnitudes (see Experiment 1, 2, and 3) but also to initiate analytic monitoring (see Experiment 2 and 3), both absolute as well as relative monitoring accuracy should be improved. Thus, in Experiment 4, relative accuracy was used as another approach to investigate if disfluency activates analytic monitoring. It was expected that analytic monitoring enables students to focus on conceptual cues when reading a text, and thus, students would realize difficulties in comprehension. Because conceptual cues are predictive for performance, the use of conceptual cues should result not only in improved absolute monitoring accuracy but also in improved relative monitoring accuracy. Improved monitoring accuracy should lead to improved control of cognitive processes and thus, to improved performance. According to theories about metacognition and self-regulated learning (e.g., Nelson & Narens, 1990; Winne & Hadwin, 1998), students use their judgments in order to control their learning. Thus, accurate judgments should enable students to select texts for rereading that in fact require further reading and to invest enough time for these texts (see section 1.5).

In order to investigate this research question, students learned with four instead of two texts to be able to compute not only absolute but also relative monitoring accuracy and to ensure that students can select between different texts for rereading. Fluency was manipulated between-person to ensure that students did not only select fluent texts for rereading and to ensure that the activation of analytic monitoring improves relative monitoring accuracy (see section 1.4). As has been shown in Experiment 1, fluency-manipulations can affect judgments in between-person designs. In Experiment 4, the findings of Experiment 1 were extended to perceptual fluency. Students made EOL judgments after a text-presentation for 2 seconds and JOLs after reading. RC judgments were made during the test. Performance was captured after reading to compute monitoring accuracy for the first reading phase, not only after rereading (second learning phase) like in Experiment 1, 2, and 3. In the second learning phase, students were allowed to reread the texts. They selected which texts they wanted to reread (before and after the knowledge test) and how much time they wanted to invest for each text; they could reread a text at a maximum of 7 minutes. Afterwards, the test was repeated to capture final performance.

Results from $N = 83$ students show that perceptual disfluency is a way to improve monitoring accuracy, dependent on the type of judgment and the stage of the learning process: Whereas disfluency led to reduced overconfidence and, hence, to perfect calibration for EOL and RC judgments, disfluency resulted in underconfidence for JOLs. Furthermore, relative monitoring accuracy was improved only for RC judgments. Nevertheless, performance in the final test was not improved because students did not use their judgments as a basis to decide which texts to reread or for how long. This finding that students did not use their judgments to control learning is most important for the research question if monitoring affects control and performance. It is important to develop instructions that help students to implement accurate monitoring into adequate control and therefore, into improved performance.

3 Discussion and Conclusion

3.1 Implications for the Theoretical Model

Whereas the results of the four experiments are discussed in more detail in each paper, the aim of this section is to discuss results in relation to each other, to previous research, and to the research questions from the theoretical model (see section 1.6). Based on the results, the theoretical model is extended to provide guidelines for systematic research in the future. Afterwards, limitations of the experiments and implications for future research and educational settings are discussed.

3.1.1 Fluency and monitoring. Regarding the first research question (Q1), if fluency is used as a cue for monitoring, results of the four experiments (see Table 1, for an overview), partly confirm that fluency affects judgments. Based on these results, it can be concluded that fluency effects on monitoring depend on the interplay between the design/the type of contrast, the type of fluency, and the stage of the learning process at which a specific type of judgment is made.

Among these factors, the design/type of contrast seems to be the most important condition for fluency effects on judgments: For contrast fluent-disfluent (which implies a within-person design), fluency effects on monitoring were found regardless of the fluency-manipulation (perceptual vs. conceptual), the type of judgment, or the learning stage (see Table 1). Moreover, lower judgments were found regardless of whether disfluency reduced performance (Experiment 3) or not (Experiment 2). This is consistent to previous research in which fluency effects on judgments were often found in within-person designs, but not in between-person designs (e.g., Susser et al., 2013, Experiment 1; see also Appendix A), regardless of whether fluency reduced performance (e.g., Mueller et al., 2013, Experiment 2; Yue et al., 2013, Experiment 1a and 2a) or did not affect performance (e.g., Magreehan et al., 2016; Yue et al., 2013, Experiment 2b and 3; see Appendix A, for further studies). Only in cases when disfluency improved performance were fluency effects sometimes not found because retrieval was used as another cue that affected judgments in an opposite direction and therefore, compensated for the fluency effect (e.g., Maki et al., 1990, Experiment 1; see Appendix A, for further studies). The findings for contrast fluent-disfluent extend previous research to different judgments at different learning stages that refer to different levels of text-processing, even to RC judgments during the test.

Table 1. *Overview of the Results from the Experiments Regarding Judgment Magnitudes*

	Judgments after text-presentation	**Judgments after reading a text once**		**Judgments after rereading a text**		**Judgments during the test**
Experiment 1 (conceptual fluency)	EOL	EOL	JOL	JOL		RC judgments
Experiment 2 (conceptual fluency)						
Contrast fluent-disfluent	EOL	EOL	JOL	JOL		RC judgments
Contrast disfluent-fluent	EOL	EOL	JOL	JOL		RC judgments
Experiment 3 (perceptual fluency)						
Contrast fluent-disfluent	/	EOL	Familiarity	Comprehension	JOL	RC judgments
Contrast disfluent-fluent	/	EOL	Familiarity	Comprehension	JOL	RC judgments
Experiment 4 (perceptual fluency)	EOL	JOL		Rereading after the test		RC judgments

Note. Fluency effect, no fluency effect, inverted fluency effect on judgments; EOL = ease of learning judgment; JOL = judgment of learning; RC judgments = retrospective confidence judgments; Familiarity = familiarity judgment; Comprehension = comprehension judgment.

The findings from Experiment 1 and 4 also extend prior research: Results show that fluency effects on monitoring can be found in between-person designs in which performance was not affected by fluency. Although, the type of fluency, the type of judgment, and the learning stage seem to play a more important role (see Table 1). Results from both experiments show that letter deletion (Experiment 1) as well as Mistral (Experiment 4) can affect judgments after reading and/or rereading (see Table 1) in a between-person design. This finding supports the assumption by Susser et al. (2013) that there might be fluency-manipulations that can affect judgments even in between-person designs (see section 1.2). Nevertheless, fluency effects on monitoring in between-person designs seem to be less stable as fluency effects in within-person designs using a contrast fluent-disfluent: RC judgments were not affected by fluency in the between-person experiments (Experiment 1 and 4), and EOL judgments after a text presentation for 2 seconds were only affected by perceptual fluency (Experiment 4). In these cases, fluency does not seem to be obvious and is not used as a cue for judgments (see paper I and paper IV, for further and more detailed explanations).

Summing up, the interplay between the design, the type of fluency, and the type of judgment at a specific learning stage seems to play a role for fluency effects on judgments. Moreover, regarding the within-person design further, the type of contrast in fluency seems to condition the fluency effect on judgments. When students experience contrast disfluent-fluent, almost no fluency effects were found for judgments after reading and rereading. These findings are rather new and can be explained by analytic monitoring, which is discussed next.

3.1.2 Disfluency and analytic monitoring. The second research question asked if disfluency activates analytic monitoring. As described in section 1.6, different approaches were used to investigate this question. This section discusses the first approach that investigates if the type of contrast in fluency affects the fluency effect on monitoring, which was in the focus of the second research question (Q2). No fluency effects for contrast disfluent-fluent were expected (see also section 1.3) because it was assumed that analytic monitoring remains for succeeding fluent material, which was concluded from findings by Alter et al. (2007, Experiment 2). Results for contrast disfluent-fluent in Experiment 2 support this assumption for judgments towards the end of the learning process (JOLs and RC judgments) but not for judgments at early learning stages (EOL judgments). For the EOL judgment after reading, an inverted fluency effect was even found (see paper II, for possible explanations). Moreover, results for contrast disfluent-fluent in Experiment 3 support this assumption for all types of judgments, except for RC judgments. Inversely, for contrast fluent-disfluent, fluency effects for all judgments during the learning process were found (see Table 1). The finding that the type of contrast in fluency moderated fluency effects on judgments goes beyond prior research that only investigated within-person designs and sometimes between-person designs. Moreover, this pattern has been found not only in Experiment 2 but has been replicated in Experiment 3, in which an even more consistent pattern was found for perceptual fluency. Furthermore, in Experiment 3, a similar pattern was found for perceived difficulty and mental effort: Students reported higher perceived difficulty and mental effort for the disfluent than for the fluent text for contrast fluent-disfluent, whereas for contrast disfluent-fluent, no fluency effect was found. Hence, although these measures are usually used as subjective measures for cognitive processing (Rey & Nieding, 2010; Schmeck et al., 2015), they might refer to monitoring (see paper III, for more details).

Summing up, the type of contrast seems to moderate fluency effects on monitoring for conceptual (i.e., letter deletion) as well as for perceptual (i.e., *Mistral*) fluency. Hence, disfluency seems to activate analytic monitoring that remains for succeeding fluent text. However, the interplay with the learning stage and the type of judgment thereby seems to play a role.

3.1.3 Disfluency and monitoring accuracy. In order to investigate analytic monitoring, not only fluency effects on judgments but also on monitoring accuracy can be considered (see section 1.6). Hence, the third research question (Q3) asked if disfluency improves monitoring accuracy. Monitoring accuracy is calculated from the judgments in relation to performance. Thus, even if fluency affects performance, conclusions about analytic monitoring can be made. Because in Experiment 3 disfluency led to lower performance than fluency particularly for contrast fluent-disfluent, absolute monitoring accuracy was investigated in addition to judgment magnitudes. Previous research shows that students are often overconfident when judging their own performance (e.g., Baker et al., 2010). It was assumed that disfluency is a way to reduce overconfidence not only because of lower judgments but also because of analytic monitoring (see section 1.4).

Results in Experiment 3 support this assumption for contrast fluent-disfluent, but not for contrast disfluent-fluent. Whereas perfect absolute monitoring accuracy of the disfluent text for contrast fluent-disfluent indicates analytic monitoring, overconfidence for the disfluent and the fluent text for contrast disfluent-fluent does not support this assumption (see also paper III). As described in section 1.4, absolute monitoring accuracy might not only be affected by analytic monitoring but can also be affected by students´ judgment magnitudes. For contrast fluent-disfluent, students made lower judgments for the disfluent than for the fluent text and this might also have affected monitoring accuracy. However, lower judgments are not sufficient to find perfect absolute monitoring accuracy for the disfluent text for contrast fluent-disfluent: Performance also was lower for the disfluent than for the fluent text, and thus, accurate monitoring has to be due to analytic monitoring to some extent. Moreover, for contrast disfluent-fluent, there are also some hints for analytic monitoring: Students´ overconfidence was at least descriptively reduced for the disfluent and also for the succeeding fluent text compared to the fluent text when experiencing contrast fluent-disfluent. Additionally, students seemed to be able to accurately judge the relation between the two texts (relative accuracy): They realized that the performance of the disfluent text was almost comparable to the performance of the fluent text for contrast disfluent-fluent. However, in Experiment 3, students learned only with two texts and relative monitoring accuracy was not in the focus of the study. In order to draw conclusions of whether fluency affects relative monitoring accuracy, more than two texts should be investigated. This was done in Experiment 4.

In Experiment 4, not only absolute but also relative monitoring accuracy was considered as an indicator for analytic monitoring. In this experiment, not only perfect absolute monitoring accuracy of EOL judgments and RC judgments was found, but furthermore, disfluency improved relative accuracy of RC judgments. This supports the assumption that disfluency activates analytic monitoring. Because of analytic monitoring, students seem to have used conceptual cues for their RC judgments, and thus, relative accuracy was improved. However, for EOL judgments and for JOLs (see also Susser et al., 2013, Experiment 1, for JOLs), there was no significant difference between fluent and disfluent texts regarding relative monitoring accuracy. These results do not support the assumption that disfluency improves monitoring accuracy via analytic monitoring. One reason for this finding might be that these judgments were made immediately after each text, and thus, students did not use conceptual cues for these judgments. Students just seem to lower their judgments for all texts to a similar extent because fluency/disfluency was the most obvious cue (see paper IV, for a more detailed discussion). For JOLs, these lowering of the judgments even resulted in underconfidence for the disfluent texts, whereas students were perfectly calibrated for the fluent texts. Therefore, analytic monitoring seems to be not only necessary to improve relative accuracy but also helpful to improve absolute accuracy. Simply lowering students´ judgments might not always be sufficient to improve absolute monitoring accuracy because this could result in underconfidence (Experiment 4).

Summing up, results on monitoring accuracy in Experiment 3 and in Experiment 4 support our conclusion that the design/type of contrast, the type of judgment, and the stage of the learning process should be considered. Additionally, the type of accuracy (absolute vs. relative accuracy) should be taken into account. Regarding the question if disfluency activates analytic monitoring, the results are not completely coherent, but it can be concluded that disfluency might possibly activate analytic monitoring under specific conditions (see above).

3.1.4 Fluency and the relation between monitoring, control, and performance. Absolute and relative monitoring accuracy are also important to answer the fourth research question (Q4) if improved monitoring leads to improved performance (see section 1.5 and 1.6). This research question was in the focus of Experiment 4. It was assumed that performance should be improved if students implement accurate monitoring into adequate control of cognitive processes. Results show that monitoring was not implemented widely into control, and thus, improved monitoring accuracy did not lead to improved performance via improved control of cognitive processes. Hence, the interplay between monitoring and control seems to be more complex than it is assumed in theories about metacognition and self-regulated learning (see also paper IV). In previous studies, students often learned with words or word-pairs, made JOLs, and selected the words or word-pairs for restudy and/or decided how long to restudy them (e.g., Mazzoni, Cornoldi, & Marchitelli, 1990; Metcalfe & Finn, 2008; Thiede & Dunlosky, 1999). However, when learning with texts, students have to monitor different levels of processing. Therefore, they make different types of judgments (see section 1.1.1). All of these judgments can potentially affect control. When making different types of judgments, it is more complex to decide which judgments to use as a basis for control. Besides not knowing which judgments to use as a basis for control, students might further have false beliefs about control (see also Schwartz & Efklides, 2012). As shown by Kornell and Bjork (2009), students do not realize the benefits of rereading, and thus, they underestimate the effects of rereading as a learning strategy. Similarly, in Experiment 4, most of the students did not invest enough time or effort for rereading in order to improve performance (see paper IV, for a more detailed discussion and further reasons). This was also the case in Experiment 1, 2, and 3: In none of these experiments did disfluency lead to improved performance. For example, in Experiment 3, students did not invest enough time or effort during rereading, even though they accurately judged their performance immediately after rereading for the disfluent text for contrast fluent-disfluent. Thus, they terminated rereading, although further learning would have been necessary because performance was lower for the disfluent than for the fluent text.

Based on the finding that disfluency did not improve performance in any of the four experiments, it can be concluded that neither improved monitoring (see Experiment 4) nor disfluency improved control of cognitive processes and performance. This finding supports the specification of disfluency-theory (see section 1.3) that the effects of disfluency should be separated with respect to monitoring and control. Disfluency can activate analytic monitoring but does not necessarily activate analytic control of cognitive processes. In the experiments of this thesis, fluency-manipulations were used that were not expected to improve performance in order to investigate the research questions. However, as previous research shows, there might be fluency-manipulations that can improve performance (see Appendix A). As described in section 1.1.2, this might also depend on the interplay between different factors like the level of processing, the encoding operation, the design, and the material. For example, regarding the design, disfluency has been found to improve performance in within-person designs (e.g., McDaniel, 1984) but not in between-person designs (e.g., Maki et al., 1990, Experiment 2). Moreover, in prior research, inverted fluency effects on performance were rarely found (see Appendix A). Hence, Bjork and Yue (2016) assumed that the hypothesis that disfluency decreases performance can be rejected. The results in this thesis support the assumption that the design affects fluency effects on performance and also on monitoring. Consistent to prior research, no fluency effects on performance were found in between-person designs (Experiment 1 and 4). However, for contrast fluent-disfluent in Experiment 3, performance was lower for the disfluent than for the fluent text, and hence, an inverted fluency effect was found. But, at the same time, disfluency improved absolute monitoring accuracy for contrast fluent-disfluent. Thus, disfluency might have improved monitoring at the cost of analytic control of cognitive processes. Students might not have enough working memory capacity to adequately control cognitive processes because analytic monitoring already requires working memory capacity, and working memory capacity is limited. Hence, learner-characteristics, like working memory capacity, should be considered in future research (see also Lehmann, Goussios, & Seufert, 2016, for working memory capacity and disfluency). Moreover, when investigating fluency effects on performance, metacognitive processes should be considered because disfluency can have different effects on monitoring and control of cognitive processes. In prior research, fluency effects on performance were in the focus: Performance was investigated to test if disfluency activates analytic processes. However, analytic processes should be distinguished with respect to monitoring and control of cognitive processes. Fluency effects on performance are more closely related to analytic control of cognitive processes than to analytic monitoring. Analytic monitoring can mainly be inferred by students´ cue use or by judgment accuracy.

Summing up, results show that disfluency can improve monitoring, but improved monitoring does not necessarily improve control of cognitive processes and performance. Furthermore, disfluency does not necessarily improve control of cognitive processes and performance, but fluency effects on performance might depend on different factors (e.g., encoding operations). Hence, not only disfluency-theory but also theories about metacognition and self-regulated learning should be specified: If students implement monitoring into control seems to depend on conditions like the number and the type of judgments as well as knowledge and beliefs about adequate control processes (e.g., beliefs about the efficacy of rereading strategies). Further research is required to investigate how better monitoring can be implemented into better control.

3.1.5 Specification of the theoretical model. Based on the results described so far (see section 3.1.1 to 3.1.4), the theoretical model that was deduced in section 1.6 is extended (see Figure 4). The blue boxes in Figure 4 describe the aspects that were investigated in this thesis. The white boxes describe aspects that were not investigated in this thesis but that also seem to be relevant given the results of the four experiments and should be considered in future research (see also section 3.2, for implications for future research). Hence, the theoretical model can be used to systematize research about conditions for fluency effects on monitoring, control, and performance. Moreover, this model specifies disfluency-theory and theories about metacognition by connecting these theories. Next, the most relevant factors are summarized, but of course, further factors can be added into this model and investigated in future research (e.g., further factors like learner characteristics; see Winne & Hadwin, 1998).

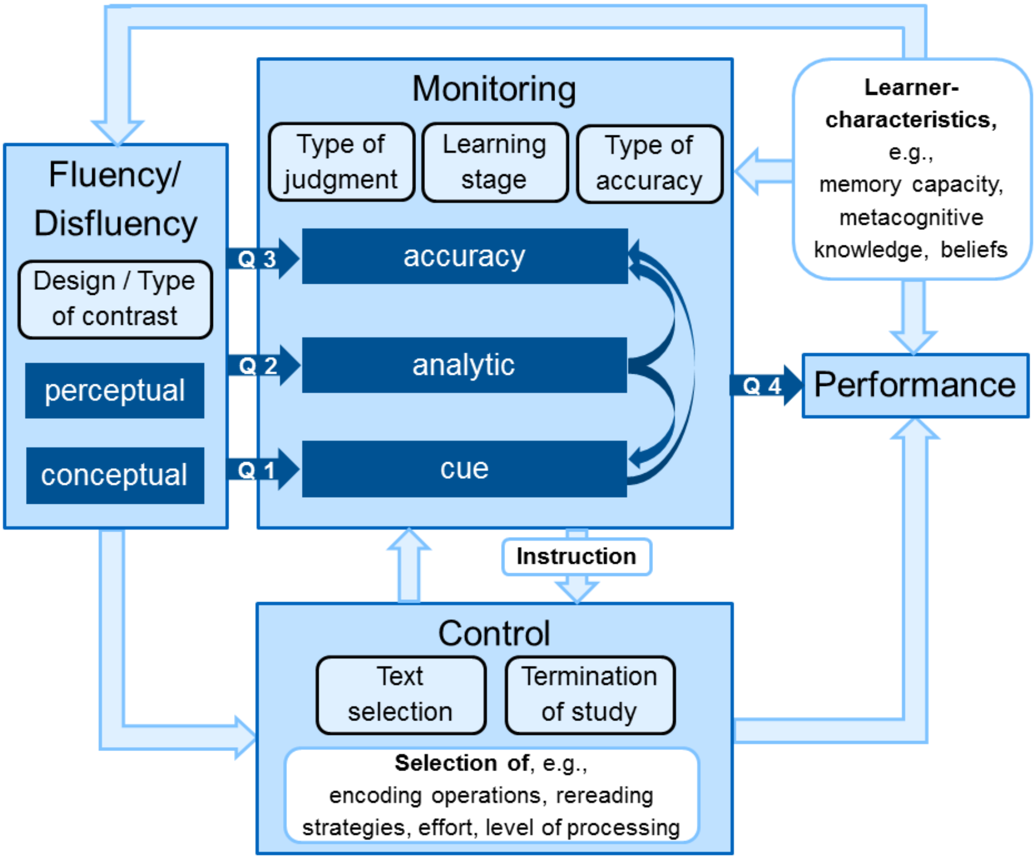

Figure 4. Specification of the theoretical model regarding fluency effects on monitoring, control, and performance.

When investigating fluency effects on monitoring and the effects of monitoring on performance, the interplay of the following conditions seems to be relevant, as concluded from the results of the four experiments. With respect to fluency, not only the type of fluency but also the design and the type of contrast in a within-person design should be considered (see the blue boxes in the field of fluency/disfluency in Figure 4). Our results show that letter deletion as a manipulation of conceptual fluency can have slightly different effects on metacognitive processes than *Mistral* as a manipulation of perceptual fluency (see, e.g., section 3.1.1 and Table 1). To investigate if fluency is a cue for monitoring (Q1), if disfluency activates analytic monitoring (Q2) and improves monitoring accuracy (Q3), as well as the relationship between these questions, research should be specified with respect to the type of judgment that can be made at different learning stages and to the type of accuracy that is of interest (see the light blue boxes in the field of monitoring in Figure 4). This is also relevant to investigate if improved monitoring leads to improved performance (Q4) because different types of judgments at different learning stages can be used as a basis for control (see section 3.1.4). Moreover, different types of monitoring accuracy could affect different types of control (e.g., text selection or termination of study, see the light blue boxes in the field of control in Figure 4). However, as shown in the experiments of this thesis, particularly in Experiment 4, monitoring is not implemented into control. Thus, instructions seem to be necessary to implement accurate monitoring into better control (see the arrow from monitoring to control in Figure 4) and performance. In the four experiments, disfluency also did not directly improve control and performance (see section 3.1.4). Hence, fluency can have different (and even inverted) effects on monitoring and control and it is important to separate these metacognitive processes in future research. Systematic research is required under which conditions disfluency fosters monitoring or control, or both metacognitive processes, and performance. Thereby, monitoring (cue use and accuracy) can be used as an indicator for analytic monitoring. Text selection and termination of study as well as more fine-grained measures, like eye-tracking or think-aloud data, can be used as an indicator for analytic control of cognitive processes, which should result in better performance. Understanding this interplay and its underlying mechanisms is necessary to derive implications for educational settings (see also De Bruin & van Gog, 2012; Efklides, 2012). Next, further implications for future research are described in relation to limitations of the thesis. Moreover, implications for educational settings are derived.

3.2 Limitations and Implications for Future Research and for Educational Settings

Overall, this thesis aimed to investigate fluency effects on monitoring and to investigate how the effects of disfluency can be used to improve monitoring and performance. The results of the four experiments give insights into fluency effects on monitoring, control, and performance, but they also have some limitations. Moreover, based on the results of the four experiments, questions for future research can be derived, as well as implications for educational settings.

With respect to fluency, the experiments were limited to one fluency-manipulation for conceptual fluency (intact texts vs. deleted letter texts) and for perceptual fluency (Times New Roman vs. Mistral). Alter and Oppenheimer (2009) stated that different types of fluency have uniform effects on judgments, even in different domains. However, as described in Figure 4 (see also section 1.1.2 and 3.1), the type of fluency as well as the specific manipulation within each type of fluency can have different impacts on monitoring, control, and performance. Hence, systematic research that considers different fluency-manipulations is necessary. Moreover, the different fluency-manipulations should be specified to their extent of disfluency (Bjork & Yue, 2016). For example, Mistral might be more disfluent than font size, and moreover, although perceptual fluency is assumed to affect lower levels of processing than conceptual fluency, there might be manipulations of conceptual fluency that are less disfluent than perceptual disfluency and hence, do not affect performance (see section 1.1.2). Using different types of disfluency ensures that students do not adapt to one type of disfluency. In educational contexts, many different types of fluency are already used: For example, in textbooks, some passages or headings are printed in italics, bold, or in different font sizes without knowing the effects on metacognitive processes or performance. However, knowing which type of disfluency under which conditions fosters monitoring, control, and performance is necessary in order to design beneficial learning material (see also Bjork & Yue, 2016, for disfluency as a desirable difficulty).

3.2.1 Monitoring. In order to investigate the fluency effects on metacognitive monitoring, different types of judgments were used in the experiments. These judgments were made at different stages of the learning process and referred to different levels of text-processing (see section 1.1.1). In research, there are many more judgments that students can make, and these judgments differ in the concrete wording. For example, in some studies, students were asked to rate the confidence to answer questions correctly (e.g., Maki et al., 1990; Rawson & Dunlosky, 2002, Experiment 4; Susser et al., 2013), whereas in other studies, students were not asked for their confidence, but simply about the number of correct questions (e.g., Besken & Mulligan, 2013, Experiment 1). Some of these judgments might capture similar, others different, aspects of metacognitive monitoring. Dependent on the specific question and the timing of the judgment in the learning process, furthermore, one type of judgment might refer to different levels of text-processing (see also paper I and III). For example, whereas judgments that are made immediately after reading mainly refer to the textbase level, delayed judgments might refer also to the situation model. Accordingly, students might use different cues for the same type of judgment depending on the timing in the learning stage. This assumption is consistent with the findings in previous research that delayed judgments are more accurate than immediate judgments (e.g., Thiede et al., 2003): Students seem to use different cues for these judgments. Also, results in the experiments of this thesis show that fluency is not always used as a cue for judgments, but this depends on the interplay of different factors like the stage of the learning process, the type of judgment, or the design/type of contrast (see Figure 4 and section 3.1).

Besides the wording and the timing, the number and sequence of judgments might further play a role for cue use. Whereas in previous studies, often only one or two types of judgments were made (e.g., Maki et al., 1990; Rawson & Dunlosky, 2002, Experiment 4; see also Appendix A), in the experiments of this thesis, different types of judgments during the learning process were made. Further research should systematically test if the number and the sequence of judgments play a role for fluency effects on judgments. Moreover, research is required to investigate if different types of judgments affect each other, which might also be relevant to explain the inverted fluency effect on the EOL judgment after reading in Experiment 3 (see section 3.1.2 and Table 1). Because students judged the disfluent text as more difficult than the fluent text after a text-presentation for 2 seconds, they might have hyper-compensated when making EOL judgments after reading (see paper III, for further discussion).

For educational contexts, it is important to be aware that different types of judgments can be made, and that the wording, the learning stage, and the number of judgments can matter. In order to get a conclusive understanding of students` monitoring, asking for different judgments that capture different aspects of monitoring might be useful. Moreover, these judgments could be asked at different learning stages because monitoring is important during the entire learning process: In advance of learning to make realistic learning plans, during learning to realize difficulties in understanding, and during retrieval to learn from failure.

3.2.2 Analytic monitoring. In order to investigate if disfluency activates analytic monitoring, not only monitoring accuracy but also the type of contrast was investigated as a new approach. This approach enabled the testing if the effects of disfluency remain for succeeding fluent text. Indeed, the type of contrast mattered for students` cue use. As results show, for contrast fluent-disfluent, fluency effects on all types of judgments were found, independent from the type of fluency, whereas this was not the case for contrast disfluent-fluent. For contrast fluent-disfluent, disfluency (Mistral) in Experiment 3 at the same time reduced performance. Hence, further research is required to investigate the underlying mechanisms of fluency effects on monitoring. Does fluency affect judgments via control, via the belief that a fluent text is easier than a disfluent text (see also Dunlosky & Mueller, 2016), and/or because of reduced performance, which should be related to control? The results of this thesis show that fluency can affect judgments independent from fluency effects on performance (see section 3.1.1). Furthermore, judgments can be affected when students read disfluent text longer than fluent text, but also when students did not reread disfluent text longer than fluent text (e.g., Experiment 1). In Experiment 1, EOL judgments after a short text-presentation were not affected by disfluency (conceptual fluency), whereas they were affected in Experiment 4 (perceptual fluency). Hence, the underlying mechanisms might also differ dependent on the interplay of the conditions mentioned in section 3.1 (e.g., design, type of fluency). For example, as mentioned by Susser et al. (2013), disfluency might not often be obvious in between-person designs, but in within-person designs. Moreover, learner characteristics (e.g., beliefs) seem also to play a role because Miele et al. (2011; see also Koriat, Nussinson, & Ackerman, 2014) found that fluency effects depend on how students interpret invested effort: If effort is interpreted as reaching the bound of ability, students will predict lower performance for disfluency than for fluency, whereas this will not be the case, if students interpret effort as a learning strategy to improve performance. Thus, future research should investigate further conditions and underlying mechanisms, including when and how fluency affects monitoring.

If fluency has not been used as a cue for judgments, further research is also required to test which cues have been used instead (see Koriat, 1997, for cues). Potential cues might be conceptual cues or retrieval, as found by Dinsmore and Parkinson (2013) for RC judgments, which could also be true for the experiments in this thesis. Dinsmore and Parkinson (2013) asked students to report cues they used for their RC judgments, and this would be interesting for other types of judgments, too. Analytic monitoring should enable students to use cues that are diagnostic for performance. However, in order to do so, further research is also required to investigate if students know which cues are diagnostic for their performance.

In educational contexts, it is important to enable students to use cues that are diagnostic for performance. This can be done either by instructing students, e.g., which cues to use or by giving feedback about cue use, or by making these cues obvious in the learning material. In the experiments of this thesis, fluency was manipulated as a cue that is inherent in the learning material. An advantage of implementing disfluency into the learning material is that further instructions are not necessary because disfluency is applied when learning with the material. However, fluency does not always seem to be a diagnostic cue for performance because fluency effects on performance are inconsistent. In between-person designs as well as when a disfluent text is presented before a fluent text, performance was not affected by fluency. However, disfluency seems to activate monitoring that remains for succeeding fluent text, which could be useful in educational settings. As shown by Alter et al. (2007, Experiment 2), presenting only a heading in a disfluent way might be useful for the processing of the subsequent material. But, knowing when disfluency is useful is necessary in order to use disfluency in a beneficial way in educational contexts.

3.2.3 Monitoring accuracy. Monitoring accuracy is computed from students` judgment and performance magnitudes. Thus, in order to investigate if students make accurate judgments, not only students` judgments but also their performance should be taken into account. This is important in order to draw conclusions from accurate monitoring to analytic monitoring, but also to interpret the measures of monitoring accuracy, which is described next.

Absolute monitoring accuracy can result not only from analytic monitoring but also from lower judgments as well as from improved performance. Previous research shows that students with low performance often show overconfidence (*Dunning-Kruger effect*, see Dunning, Johnson, Ehrlinger, & Kruger, 2003; Hacker, Bol, Horgan, & Rakow, 2000; Kruger & Dunning, 1999; see also De Bruin et al., 2016, for an overview), even when incentives for making accurate judgments are given (Ehrlinger et al., 2008). Hence, even in cases where high and low performing students make equal judgments, high-performing students might be more accurate or even underconfident whereas low-performing students might be overconfident. Results in this thesis show that students can make accurate judgments even when performance is low (see disfluent text for contrast fluent-disfluent in Experiment 3). Inversely to the studies mentioned above, in this thesis, participants were not divided into groups according to their given performance, but performance differences (see Experiment 3) were due to the experimental manipulation of fluency. Hence, monitoring accuracy does not seem to be caused by performance but by analytic monitoring (see section 3.1.3). In order to conclude if accurate monitoring is due to analytic monitoring, future research should consider the reasons for absolute monitoring accuracy (e.g., lower judgments, improved performance, and/or analytic monitoring). Moreover, different types of accuracy (i.e., absolute and relative accuracy) should be considered because absolute monitoring accuracy can be high, even if relative monitoring accuracy is low (e.g., EOL judgments in Experiment 4). Relative monitoring accuracy is not improved if students simply lower their judgments, but analytic monitoring is required to enable students to discriminate performance between different texts (see section 1.6). In educational settings, it is useful to foster both absolute and relative monitoring accuracy as well as performance at the same time. Inversely, improving performance would not have been useful to investigate the research questions of this thesis (see section 1.2 and 1.4). With respect to performance, another limitiation should be considered that might affect monitoring accuracy.

Performance can be affected not only by fluency or by the difficulty of the learning material but also by the difficulty of the test. For absolute monitoring accuracy, students can be overconfident when the test is difficult, whereas they can be underconfident when the test is easy, without changing judgments (see *also hard-easy effect*, Lichtenstein, Fischhoff, & Phillips, 1982). Further, relative monitoring accuracy can be affected by the difficulty between different questions. This limitation is important when metacognition research is transferred to metacomprehension research. In metacognition research, students often learn with word-pairs, and thus, they exactly know the task in the performance test (i.e., retrieval of the second word when the first word is presented) and can consider the difficulty of the performance test when making judgments. When learning with texts, students do not know all the tasks of the performance test. Furthermore, they have to infer their answers from the textbase level and their mental model, simply retrieving a word is not sufficient to answer questions on comprehension and transfer. Thus, further research should investigate how to solve these issues (see paper IV, for a more detailed discussion).

In educational contexts, providing examples for questions and the difficulty of the test could be helpful. However, in exams, the difficulty of questions often differs within the test. Moreover, students usually do not know the questions of an exam during learning. This makes it difficult to predict performance in a test. In order to obtain the best possible performance, students should completely understand the learning material. In this case, they should be able to understand both easy and difficult material and to answer even difficult questions. Furthermore, even if students know the test questions and thus their difficulty, students have been found to be overconfident (see Dunlosky & Metcalfe, 2009, for an overview). Hence, not only information about the test but also instructions to improve monitoring accuracy is required in educational contexts. Thereby, different instructions might be necessary for different learners, dependent on the age group (see, e.g., van Loon, De Bruin, van Gog, & van Merriënboer, 2013, for monitoring accuracy in children), or the performance level. However, fostering monitoring is particularly helpful if monitoring is implemented into control (Dunlosky & Rawson, 2012; Efklides, 2012).

3.2.4 Relation between monitoring, control, and performance. Regarding the connection between monitoring, control, and performance, results show that this connection is not as compelling as suggested by theory. Future research is required to understand reasons why monitoring is not used for control of cognitive processes. Thereby, more fine-grained measures than termination of study or text selection should be used, e.g., eye-tracking data. This is necessary because control can be affected by other factors besides monitoring (see Figure 4), like learning strategies, or beliefs about effective control processes (see section 3.1.4). Prior research further shows that study time allocation can further be affected by students` motivation and learning goals (Pintrich, 2000), their interest in the topic (Son & Metcalfe, 2000), or the reward (Ariel, Dunlosky, & Bailey, 2009).

Besides these factors, the number of texts can also affect measures (i.e., gamma correlations) for the transition between monitoring and control (see also paper IV). In order to compute gamma, students` judgments and the texts selected for rereading have to vary within a participant: Gamma cannot be computed if a participant selects all texts for rereading or judges all texts as being equally difficult. Again, in metacognition research, this issue is less problematic because students learn with many word-pairs (e.g., Thiede & Dunlosky, 1999: up to 30), and thus, a non-variation is unlikely. In metacomprehension research, a smaller number of texts is used because learning with texts requires more time and effort. Future research should investigate how many texts should at least be used in order to interpret gamma and compare these results to results from metacognition research, where connections between monitoring and control have sometimes been found (e.g., Nelson & Leonesio, 1988). Hence, future research should consider theoretical as well as methodological issues.

Based on the results of this thesis, an implication for educational settings is that students need instructions to implement accurate monitoring into adequate control (see also Schwartz & Efklides, 2012). This could be realized, e.g., by information about which judgments to use as a basis for control. In Experiment 4, absolute and relative accuracy of RC judgments was high, and thus, these judgments might be useful to control learning. Instructions could also focus on providing students with information about effective learning strategies (e.g., using prompts, see Bannert, 2006; Bannert & Mengelkamp, 2013; Bannert & Reimann, 2012; Bannert, Sonnenberg, Mengelkamp, & Pieger, 2015). In educational settings, students have to learn a lot of material (e.g., Dunlosky & Rawson, 2012; Maki & McGuire, 2002); to acquire knowledge, the content can be presented in multiple ways, e.g., in textbooks, pictures, figures, or digital media. Providing knowledge about how to focus and what to reread, based on accurate monitoring about different learning material and different learning situations, is complex but necessary for educational settings.

Summing up, future research should develop instructions that improve monitoring in combination with instructions to implement accurate monitoring into better control, which is necessary to foster performance.

3.3 Conclusion

This thesis contributes to the field of metacognition research as well as fluency research. By connecting both areas, not only disfluency-theory but also theory about metacognition and metacomprehension was specified. Disfluency enabled the investigation of the interplay between monitoring, control, and performance that seems to be more complex than suggested by theory. However, interventions that improve monitoring accuracy are only useful when monitoring is implemented into control and performance (Dunlosky & Rawson, 2012; Efklides, 2012). Hence, the next step should be to investigate not only interventions that improve monitoring but also interventions that help to implement accurate monitoring into better performance. Thereby, it is necessary to understand the underlying mechanisms and factors that affect the interplay between metacognitive processes and performance (De Bruin & van Gog, 2012). However, to understand these mechanisms not only learning materials like words and word-pairs should be investigated, but research should also investigate meaningful text material. Meaningful material is more complex and thus, further mechanisms might be relevant. This is necessary to apply findings into educational settings as the next step.

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1–3. doi:10.1016/j.learninstruc.2012.10.003

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. doi:10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576. doi:10.1037/0096-3445.136.4.569

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, *138*(3), 432–447. doi:10.1037/a0015928

Baker, J. M. C., Dunlosky, J., & Hertzog, C. (2010). How accurately can older adults evaluate the quality of their text recall? The effect of providing standards on judgment accuracy. *Applied Cognitive Psychology, 24*, 134–147. doi:10.1002/acp.1553

Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied*, *20*(4), 336–348. doi:10.1037/xap0000023

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research*, *35*(4), 359–375. doi:10.2190/94V6-R58H-3367-G388

Bannert, M., & Mengelkamp, C. (2013). Scaffolding hypermedia learning through metacognitive prompts. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 171–186). New York, NY: Springer. doi:10.1007/978-1-4419-5546-3_12

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, *40*(1), 193–211. doi:10.1007/s11251-011-9167-4

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short-and long-term effects of students` self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, *52*, 293–306. doi:10.1016/j.chb.2015.05.038

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. doi:10.3758/s13421-013-0307-8

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

Bjork, E. L., DeWinstanley, P. A., & Storm, B. C. (2007). Learning how to learn: Can experiencing the outcome of different encoding strategies enhance subsequent encoding?. *Psychonomic Bulletin & Review*, *14*(2), 207–211. doi:10.3758/BF03194053

Bjork, E. L., & Storm, B. C. (2011). Retrieval experience as a modifier of future encoding: Another test effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1113–1124. doi:10.1037/a0023549

Bjork, R. A., & Yue, C. L. (2016). Commentary: Is disfluency desirable?. *Metacognition and Learning, 11,* 133–137. doi:10.1007/s11409-016-9156-8

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*(2), 161–186. doi:10.1016/S0959-4752(96)00015-1

De Bruin, A. B., Kok, E. M., Lobbestael, J., & Grip, A. (2016). The impact of an online tool for monitoring and regulating learning at university: Overconfidence, learning strategy, and personality. *Metacognition and Learning, 1,* 21–43. doi:10.1007/s11409-016-9159-5

De Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*(4), 245–252. doi:10.1016/j.learninstruc.2012.01.003

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 114–118. doi:10.1016/j.cognition.2010.09.012

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students` explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14. doi:10.1016/j.learninstruc.2012.06.001

Dreisbach, G., & Fischer, R. (2011). If it's hard to read… try harder! Processing fluency as signal for effort adjustments. *Psychological Research*, *75*(5), 376–383. doi:10.1007/s00426-010-0319-y

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol. 54* (pp. 103–140). San Diego, CA: Elsevier Academic Press. doi:10.1016/B978-0-12-385527-5.00004-8

Dunlosky, J., Hartwig, M. K., Rawson, K. A., & Lipko, A. R. (2011). Improving college students' evaluation of text learning using idea-unit standards. *The Quarterly Journal of Experimental Psychology*, *64*(3), 467–484. doi:10.1080/17470218.2010.502239

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228–232. doi:10.1111/j.1467-8721.2007.00509.x

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Los Angeles, CA: SAGE.

Dunlosky, J., & Mueller, M. L. (2016). Recommendations for exploring the disfluency hypothesis for establishing whether perceptually degrading materials impacts performance. *Metacognition and Learning, 11,* 123–131. doi:10.1007/s11409-016-9155-9

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280. doi:10.1016/j.learninstruc.2011.08.003

Dunlosky, J., & Thiede, K. W. (2013). Metamemory. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 283–298). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780195376746.013.0019

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3), 83–87. doi:10.1111/1467-8721.01235

Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema*, *21*(1), 76–82.

Efklides, A. (2012). Commentary: How readily can findings from basic cognitive psychology research be applied in the classroom?. *Learning and Instruction, 22*(4), 290–295. doi:10.1016/j.learninstruc.2012.01.001

Ehrlinger, J., & Eichenbaum, A. (2016). The dark (and light) sides of overconfidence. In V. Zeigler-Hill & D. K. Marcus (Eds.), *The dark side of personality: Science and practice in social, personality, and clinical psychology* (pp. 251–266). Washington, DC: American Psychological Association. doi:10.1037/14854-013

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, *105*(1), 98–121. doi:10.1016/j.obhdp.2007.05.002

Eitel, A. (2016). How repeated studying and testing affects multimedia learning: Evidence for adaptation to task demands. *Learning and Instruction*, *41,* 70–84. doi:10.1016/j.learninstruc.2015.10.003

Eitel, A., & Kühl, T. (2016). Effects of disfluency and text expectancy on learning with text. *Metacognition and Learning, 11*(1), 107–121. doi: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. doi:10.1002/acp.3004

Faber, M., Mills, C., Kopp, K., & D'Mello, S. (2016). The effect of disfluency on mind wandering during text comprehension. *Psychonomic Bulletin & Review*, *24*(3), 914–919. doi:10.3758/s13423-016-1153-z

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. doi:10.1037/0003-066X.34.10.906

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*(6), 597–602. doi:10.3758/BF03202442

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. doi:10.1037/0022-0663.92.1.160

Jönsson, F. U., & Lindström, B. R. (2010). Using a multidimensional scaling approach to investigate the underlying basis of ease of learning judgments. *Scandinavian Journal of Psychology*, *51*(2), 103–108. doi:10.1111/j.1467-9450.2009.00762.x

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294–303. doi:10.1037/0003-066X.49.4.294

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. doi:10.1037/0096-3445.126.4.349

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*(2), 416–428. doi:10.3758/MC.36.2.416

Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction*, *22*(4), 296–298. doi:10.1016/j.learninstruc.2012.01.002

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. doi:10.1037/0096-3445.135.1.36

Koriat, A., Nussinson, R., & Ackerman, R. (2014). Judgments of learning depend on how learners interpret study effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1624–1637. doi:10.1037/xlm0000009

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General, 138*, 449–468. doi:10.1037/a0017350

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794. doi:10.1177/0956797611407929

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. doi:10.1037/0022-3514.77.6.1121

Kühl, T., & Eitel, A. (2016). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning, 11,* 1–13. doi:10.1007/s11409-016-9154-x

Lanska, M., Olds, J. M., & Westerman, D. L. (2014). Fluency effects in recognition memory: Are perceptual fluency and conceptual fluency interchangeable?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 1–11. doi:10.1037/a0034309

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning, 11*(1), 89–105. doi:10.1007/s11409-015-9149-z

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 464–470. doi:10.1037/0278-7393.16.3.464

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of subjective probabilities: The state of the art up to 1980. In D. Kahneman, P. Slovjc, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York, NY: Cambridge University Press. doi:10.1007/978-94-010-1834-0_8

Lin, L., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, *23*(4), 345–391. doi:10.1006/ceps.1998.0972

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning, 11*(1), 35–56. doi:10.1007/s11409-015-9147-1

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 663–679. doi:10.1037/0278-7393.10.4.663

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 609–616. doi:10.1037/0278-7393.16.4.609

Maki, R. H., & McGuire, M. J. (2002). Metacognition for text: Findings and implications for education. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 68–92). Cambridge, UK: University Press. doi:10.1017/CBO9780511489976.004

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation?. *Memory & Cognition*, *18*(2), 196–204. doi:10.3758/BF03197095

McDaniel, M. A. (1984). The role of elaborative and schema processes in story memory. *Memory & Cognition*, *12*(1), 46–51. doi:10.3758/BF03196996

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 175–199). New York, NY: Psychology Press. doi:10.4324/9780203842539

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, *25*(6), 645–656. doi:10.1016/0749-596X(86)90041-0

Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: Stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, *38*(4), 441–451. doi:10.3758/MC.38.4.441

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

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159–163. doi:10.1111/j.1467-8721.2009.01628.x

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. doi:10.3758/PBR.15.1.174

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*(4), 463–477. doi:10.1016/j.jml.2004.12.001

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., Pennycook, G., Ackerman, R., Thompson, V. A., & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16–e30. doi:10.1037/xge0000049

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered?: It depends on your beliefs about intelligence. *Psychological Science*, *22*(3), 320–324. doi:10.1177/0956797610397954

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. doi:10.1037/a0019745

Miesner, M. T., & Maki, R. H. (2007). The role to test anxiety in absolute and relative meta-comprehension accuracy. *European Journal of Cognitive Psychology*, *19*(4–5), 650–670. doi:10.1080/09541440701326196

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. doi:10.3758/s13423-012-0343-6

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133. doi:10.1037/0033-2909.95.1.109

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, *5*(4), 207–213. doi:10.1111/j.1467-9280.1994.tb00502.x

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the 'labor-in-vain effect.'. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686. doi:10.1037/0278-7393.14.4.676

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York, NY: Academic Press. doi:10.1016/S0079-7421(08)60053-5

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. doi:10.1016/j.tics.2008.02.014

Pieschl, S. (2009). Metacognitive calibration – An extended conceptualization and potential applications. *Metacognition and Learning, 4*(1), 3–31. doi:10.1007/s11409-008-9030-4

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). New York, NY: Academic Press. doi:10.1016/B978-012109890-2/50043-3

Pyc, M. A., Rawson, K. A., & Aschenbrenner, A. J. (2014). Metacognitive monitoring during criterion learning: When and why are judgments accurate?. *Memory & Cognition*, *42*(6), 886–897. doi:10.3758/s13421-014-0403-4

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 69–80. doi:10.1037/0278-7393.28.1.69

Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, *22*(4), 262–270. doi:10.1016/j.learninstruc.2011.10.007

Rey, G. D., & Nieding, G. (2010). Arbeitsgedächtnis und Cognitive Load [Working memory and cognitive load]. In H.-P. Trolldenier, W. Lenhard, & P. Marx (Eds.), *Brennpunkte der Gedächtnisforschung. Entwicklungs- und pädagogisch-psychologische Perspektiven [Focal points of memory research. Developmental and educational-psychological perspectives]* (pp. 67–85). Göttingen, Germany: Hogrefe.

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,* 206–211. doi:10.3758/s13423-014-0658-6

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*, *43*(1), 93–114. doi:10.1007/s11251-014-9328-3

Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction*, *24,* 48–57. doi:10.1016/j.learninstruc.2012.08.007

Schwartz, B. L., & Efklides, A. (2012). Metamemory and memory efficiency: Implications for student learning. *Journal of Applied Research in Memory and Cognition*, *1*(3), 145–151. doi:10.1016/j.jarmac.2012.06.002

Schwartz, B. L., & Perfect, T. J. (2002). Introduction: Toward an applied metacognition. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 1–11). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511489976

Schwarz, N. (2010). Meaning in context: Metacognitive experiences. In B. Mesquita, L. F. Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 105–125). New York, NY: Guilford Press.

Shanks, L. L., & Serra, M. J. (2014). Domain familiarity as a cue for judgments of learning. *Psychonomic Bulletin & Review*, *21*(2), 445–453. doi:10.3758/s13423-013-0513-1

Son, L. K., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1890 to the present (and beyond). In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 333–351). New York, NY: Psychology Press.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204–221. doi:10.1037/0278-7393.26.1.204

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. doi:10.1111/j.1467-9280.2008.02189.x

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

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. doi:10.3758/s13421-013-0323-8

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. doi:10.1037/0022-0663.95.1.66

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1024–1037. doi:10.1037/0278-7393.25.4.1024 van Loon, M. H., de Bruin, A. H., van Gog, T., & van Merriënboer, J. G. (2013). The effect of delayed-JOLs and sentence generation on children's monitoring accuracy and regulation of idiom study. *Metacognition and Learning*, *8*(2), 173–191. doi:10.1007/s11409-013-9100-0

Westerman, D. L., Lanska, M., & Olds, J., M. (2015). The effect of processing fluency on impressions of familiarity and liking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 426–438. doi:10.1037/a0038356

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: LEA.

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. doi:10.3758/s13421-012-0255-8

Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review*, *2*(2), 173–201. doi:10.1007/BF01322178

Appendix

Content

Appendix A – Overview Fluency Effects

Table A1. *Overview of Studies that Investigated Fluency Effects on Performance and Partly on Judgments (not Exhaustive)*

	Fluency-Manipulation	Material	Design[a]	Performance[b]	Judgments[c]
Alter et al. (2007):					
Experiment 1	*font* fluent: black Myriad Web, 12 disfluent: gray italicized Myriad Web, 10	Cognitive Reflection Test (CRT)	between	Answers	/
Experiment 2	*Manipulation of the masthead only* fluent: easy-to-read typeset disfluent: difficult-to-read combination of letter-resembling symbols	Persuasion: text about a short review of a new MP3 player	between	Use of systematic (vs. heuristic) cues	/
Experiment 4	*Same as in Experiment 1*	Questionnaire with six syllogistic reasoning problems	between	Answers	Pilot-ratings: JOL, Perceived difficulty
Besken & Mulligan (2013):					
Experiment 1	*Encoding condition* fluent: presentation for 2,500 ms disfluent: presentation for 83 ms, replaced by a row of Xs for 2,417 ms	Word-list	within	Recalled words	JOL
Experiment 2	*Same as in Experiment 1*	Word-list	within	Recalled words	JOL
Bjork & Storm (2011):					
Experiment 1	*Letter deletion* fluent: intact disfluent: deleted letters	Sentences (text presented sentence by sentence)	within	Recall of the first text-passage, Recall of the second text-passage	/

	Fluency-Manipulation	Material	Design[a]	Performance[b]	Judgments[c]
Bjork & Storm (2011):					
Experiment 2	*Same as in Experiment 1*	Sentences (text presented sentence by sentence)	within	Recall of the first text-passage, Recall of the second text-passage	/
Experiment 3	*Same as in Experiment 1 and 2*	Sentences (text presented sentence by sentence)	within	Free recall of the first text-passage, Fill-in-the-blank-test of the second text-passage	/
Experiment 4	*Same as in Experiment 1, 2, and 3*	Sentences (text presented sentence by sentence)	within	Free recall of the first text-passage, Fill-in-the-blank-test of the first text-passage, Free recall of the second text-passage, Fill-in-the-blank-test of the second text-passage	/
Burnett & Bodner (2014):					
Experiment 1	*Letter deletion* fluent: intact disfluent: deleted letters	Sentences (text presented sentence by sentence)	Mixed: within and between	Recall in test 1 in the within-person design, Recall in test 2 in the within-person design, Recall in the between-person design	/

Burnett & Bodner (2014):					
Experiment 2	*Same as in Experiment 1*	Sentences (text presented sentence by sentence)	within	Recall in test 1 Recall in test 2	/
Experiment 3	*Same as in Experiment 1 and 2*	Sentences (text presented sentence by sentence)	within	Recall	JOL, RC judgment
DeWinstanley & Bjork (2004):					
Experiment 1a	*Letter deletion* fluent: intact disfluent: deleted letters	Sentences (text presented sentence by sentence)	within	Fill-in-the-blank-test of the first text-passage, Fill-in-the-blank-test of the second text-passage	/
Experiment 1b	*Same as in Experiment 1a*	Sentences (text presented sentence by sentence)	within	Fill-in-the-blank-test of the first text-passage, Fill-in-the-blank-test of the second text-passage	/
Experiment 2	*Same as in Experiment 1a and 1b*	Sentences (text presented sentence by sentence)	within	Fill-in-the-blank-test (first and second passage)	/
Experiment 3	*Same as in Experiment 1a, 1b, and 2*	Sentences (text presented sentence by sentence)	between	Fill-in-the-blank-test (first and second passage)	/

	Fluency-Manipulation	Material	Design[a]	Performance[b]	Judgments[c]
Diemand-Yauman et al. (2011):					
Experiment 1	*font* fluent: Arial, 16, pure black font disfluent: Comic Sans MS or Bodoni MT 60% grayscale font, 12	Three species of aliens, each with 7 features	between	Answered questions	/
Experiment 2	*font* fluent: unedited disfluent: Haettenschweiler, mono-type Corsiva, Comic Sans Italicized, disfluent copy	Worksheets and PowerPoint slides in class-room settings	between	Exams	Perceived difficulty
Eitel & Kühl (2016):					
	font fluent: Arial, 16 disfluent: Brush Script MT, 16	Text with two static pictures (pictures were not manipulated by fluency)	between	Retention and transfer	Mental effort, Perceived difficulty
Eitel et al. (2014):					
Experiment 1	*First Page* fluent: Arial, 14, black disfluent: text: Haettenschweiler, 14, grayscale 50%; picture: low-quality copy, wavily blurred and deformed *Second Page* fluent: Arial, 14, black disfluent: low-quality copy wavily deformed and blurred	Multimedia: Text and pic-tures	between	Transfer questions on the text, Retention	Mental effort on texts, Perceived difficulty

Eitel et al. (2014):					
Experiment 2	*Similar to Experiment 1* fluent: Arial, 14 or 11, black disfluent: text: Haettenschweiler, 14, grayscale 50% or wavily deformed and blurred; picture: low-quality copy, wavily blurred and deformed or no picture	Text only or text with pictures	between	Retention and transfer, Pictorial recall	Mental effort, Perceived difficulty
Experiment 3	*Similar to Experiment 1 and 2*	Text and pictures	between	Retention and transfer, Pictorial recall	Mental effort, Perceived difficulty
Experiment 4	*font* fluent: Arial, 10, black disfluent: Haettenschweiler, 10, italic, grayscale 35%	Text and pictures	between	Retention and transfer, Pictorial recall	Mental effort, Perceived difficulty
Faber et al. (2016):					
	font fluent: Arial typeface disfluent: Comic Sans typeface	Text (presented sentence by sentence, not possible to go back)	between	Text-level comprehension and interference-level comprehension	Comprehension judgment, Mental effort
Gao et al. (2011):					
	Visual noise fluent: no noise disfluent: visual noise (low noise); dynamic visual noise (high noise)	Sentences	within	Recall of less central propositions, Recall of moderately memorable propositions, Recall of core propositions	/
Gao et al. (2012):					
Experiment 1	*Dynamic noise* disfluent: low, medium, high	Sentences	within	Recall, Recall of core propositions for older adults	/

	Fluency-Manipulation	Material	Design[a]	Performance[b]	Judgments[c]
Gao et al. (2012):					
Experiment 2	*Dynamic noise* disfluent: low, medium, high (higher levels than in Experiment 1)	Sentences	within	Recall, Recall of core propositions (for older and younger adults)	/
Kornell et al. (2011):					
Experiment 1	*Type size* fluent: large disfluent: small	Word-list	within	Recall	JOL
Experiment 2	*Same as in Experiment 1*	Word-list	within	Recall	JOL
Experiment 3	*Description of Experiment 1*				JOL
Lehmann et al. (2016):					
	font fluent: Arial, 12, black disfluent: Haettenschweiler, 12, grayscale 35%	Text (printed)	between	Retention, comprehension, and transfer test (open answer)	/
Magreehan et al. (2016):					
Experiment 1	*Font* (all in CAPITAL letters) fluent: black bold Arial, 56 disfluent: light gray (RGB code 217, 217, 217) italicized Times New Roman, 32	Word-pairs	between	Recall	JOL
Experiment 2	*Same as in Experiment 1* but with a third group (“even-more disfluent group”): disfluent with background: slightly lighter shade of gray (RGB code 225, 225, 225)	Word-pairs	between	Recall	JOL
Experiment 3	*Same as in Experiment 1*	Word-pairs	within	Recall	JOL
Experiment 4	*Same as in Experiment 2*	Word-pairs	within	Recall	JOL
Experiment 5	*Same as in Experiment 2 and 4*	Word-pairs	within	Recall	JOL

Maki et al. (1990):					
Pilot	*Letter deletion* fluent: intact disfluent: deleted letters	Expository text	within	Fill-in-the-blank test (Cued recall)	/
Experiment 1	*Letter deletion* fluent: intact disfluent: deleted letters	Expository text	within	Cued recall	Comprehen-sion judg-ment, JOL, RC judgment
Experiment 2	*Letter deletion* fluent: intact disfluent: deleted letters	Expository text	between	Cued recall	JOL, RC judgment
McDaniel (1984):					
Experiment 1	*Letter deletion* fluent: intact disfluent: deleted letters	Narrative text	within	Recall of the story	/
Experiment 2	*Letter deletion* fluent: intact disfluent: deleted letters	Narrative text	within	Recall of the story	/
Experiment 3	*Letter deletion* fluent: intact disfluent: deleted letters	Narrative text	within	Recall of the story	/
McDaniel et al. (1986):					
Experiment 1	*Letter deletion or sentence reordering* fluent: intact disfluent: letter deletion or sentence reordering	Texts (fairy tale or descriptive text)	between	Recall	Comprehen-sion judg-ment

	Fluency-Manipulation	Material	Design[a]	Performance[b]	Judgments[c]
McDaniel et al. (1986):					
Experiment 2	*Letter deletion or sentence reordering* fluent: intact disfluent: letter deletion or sentence reordering	Text (fairy tale or descriptive text)	between	Recall of the fairy tale (highest for letter deletion), Recall of the descriptive text only for sentence reordering, Recall of the descriptive text: no effect of letter deletion	Comprehension judgment
McDaniel et al. (1989):					
	Letter deletion fluent: intact disfluent: letter deletion: hard (vowels and some consonants) or easy (vowels)	Narrative text	within	Cued recall	/
Meyer et al. (2015):					
	font 17 Experiments with different font types	Cognitive Reflection Test (CRT)	between	Answers	/
Miele et al. (2011):					
Experiment 2	*font* fluent: Arial 48 disfluent: Arial, 18	Word-lists	within	Recall	Perceived difficulty, JOL for entity theorists, JOL for incremental theorists

Miele & Molden (2010):					
Experiment 3	*font* fluent: Times New Roman, black, 12 disfluent: italicized Juice ITC, 12	Expository text	between	Comprehension	Perceived difficulty, Comprehen-sion judg-ment
Mueller et al. (2013):					
Experiment 2	*Encoding* fluent: normal format disfluent: alternating format	Word-pairs	within	Recall	JOL
Mulligan (1996):					
Experiment 1	*Encoding condition* fluent: presentation for 2.5 s disfluent: presentation for 110 ms, replaced by a row of Xs for 2,390 ms	Word-list	within	Category-cued recall tests, Category-exemplar reduction test	/
Experiment 2	*Encoding condition* fluent: presentation for 4 s disfluent: presentation for 110 ms, replaced by a row of Xs for 3,890 ms	Word-list	within	Category-cued recall tests, Category-exemplar reduction test	/
Experiment 3	*Same as in Experiment 1*	Word-list	within	Category-cued recall tests, Category-exemplar reduction test	/
Experiment 4	*Encoding condition* fluent: presentation for 2.5 s disfluent: presentation for 100 ms, replaced by a row of Xs for 2,400 ms	Word-list	within	Yes-no recognition test, Rhyme-recognition test	/

	Fluency-Manipulation	Material	Design[a]	Performance[b]	Judgments[c]
Rawson & Dunlosky (2002):					
Experiment 4	*Letter deletion* fluent: intact disfluent: letter deletion: easy-completion vs. difficult-completion	Expository text	within	Fill-in-the-blank test (cued recall) of difficult-completion, Fill-in-the-blank test (cued recall) of easy-completion	JOL, Comprehension judgment
Rummer et al. (2016):					
Experiment 1	*font* fluent: black Arial, 16 disfluent: 60% grayscale Comic Sans MS, 12	Five species of aliens, each with 7 features	within	Recall	/
Experiment 2	*Same as in Experiment 1* but paper version	Five species of aliens, each with 7 features	within	Recall	/
Song & Schwarz (2008):					
Experiment 1	*font* fluent: black Arial, 12 disfluent: grey Brush Script MTm 12	Questions	between	Distorted question (moses illusion), Undistorted question	/
Experiment 2	*Same as in Experiment 1*	Questions	between	Distorted question, Undistorted question	/
Strukelj et al. (2016):					
	Encoding fluent: Arial, 48 disfluent: blurred (low-pass filtering each color band, average: 8x8 pixels)	Text	between	Free recall	/
Sungkhasettee et al. (2011):					
Experiment 1	*Word orientation* fluent: upright disfluent: inverted	Word-list	within	Recall	JOL

Sungkhasettee et al. (2011):					
Experiment 2	*Word orientation* fluent: upright disfluent: inverted	Word-lists	within	Recall	JOL
Susser et al. (2013):					
Experiment 1	*font* fluent: Arial 48 disfluent: Arial 18 withe font on black background	Word-list	Mixed: within and between	Recall	JOL in the within-person design, JOL in the between-person design
Experiment 3	*Letter-transposition generation* fluent: intact disfluent: first two letters reversed	Word-list	Mixed: within and be-tween	Recall in the within-person design, Recall in the between-person design	JOL
Weltman & Eakin (2014):					
	Mistakes or font fluent: no manipulation disfluent: workshop material with pedagogical mistakes or unusual font	Workshop handouts (Pow-erPoint slides)	between	Multiple-choice ques-tions (especially medi-um and higher levels of comprehension)	Feeling of mastery
Yue et al. (2013):					
Experiment 1a	*Clarity of words* fluent: clear font disfluent: blurred font	Word-lists	within	Recall	JOL
Experiment 1b	*Similar to Experiment 1a*	Word-lists	between	Recall	JOL
Experiment 2a	*Similar to Experiment 1a*	Word-lists	within	Recall	JOL
Experiment 2b	*Similar to Experiment 1a*	Word-lists	within	Aural recognition test	JOL
Experiment 3	*Similar to Experiment 1a*	Word-lists	within	Recall	JOL

Note. Fluency effect, no fluency effect, inverted fluency effect; auditory fluency and relatedness is not considered; focus is on students, younger age-groups are not considered.

[a]refers to the manipulation of fluency, there could be further variables that might have been manipulated.

[b]describes the most important findings regarding the fluency effect on performance.

[c]the concrete wording of the judgments was slightly different in different studies, even in cases when coded as the same judgment; JOL = judgment of learning; RC judgment = retrospective confidence judgment; perceived difficulty and mental effort were also included although they are often used as subjective measures of cognitive processes.

References

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. doi:10.1037/0096-3445.136.4.569

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. doi:10.3758/s13421-013-0307-8

Bjork, E. L., & Storm, B. C. (2011). Retrieval experience as a modifier of future encoding: Another test effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1113–1124. doi:10.1037/a0023549

Burnett, A. N., & Bodner, G. E. (2014). Learnin' 'bout my generation? Evaluating the effects of generation on encoding, recall, and metamemory across study-test experiences. *Journal of Memory and Language*, *75*, 1–13. doi:10.1016/j.jml.2014.04.005

DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, *32*(6), 945–955. doi:10.3758/BF03196872

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 114–118. doi:10.1016/j.cognition.2010.09.012

Eitel, A., & Kühl, T. (2016). Effects of disfluency and text expectancy on learning with text. *Metacognition and Learning, 11*(1), 107–121. doi: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. doi:10.1002/acp.3004

Faber, M., Mills, C., Kopp, K., & D'Mello, S. (2016). The effect of disfluency on mind wandering during text comprehension. *Psychonomic Bulletin & Review, 24*(3), 914–919. doi:10.3758/s13423-016-1153-z

Gao, X., Levinthal, B. R., & Stine-Morrow, E. L. (2012). The effects of ageing and visual noise on conceptual integration during sentence reading. *The Quarterly Journal of Experimental Psychology*, *65*(9), 1833–1847. doi:10.1080/17470218.2012.674146

Gao, X., Stine-Morrow, E. L., Noh, S. R., & Eskew, R. J. (2011). Visual noise disrupts conceptual integration in reading. *Psychonomic Bulletin & Review*, *18*(1), 83–88. doi:10.3758/s13423-010-0014-4

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794. doi:10.1177/0956797611407929

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning, 11*(1), 89–105. doi:10.1007/s11409-015-9149-z

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning, 11*(1), 35–56. doi:10.1007/s11409-015-9147-1

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 609–616. doi:10.1037/0278-7393.16.4.609

McDaniel, M. A. (1984). The role of elaborative and schema processes in story memory. *Memory & Cognition*, *12*(1), 46–51. doi:10.3758/BF03196996

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, *25*(6), 645–656. doi:10.1016/0749-596X(86)90041-0

McDaniel, M. A., Ryan, E. B., & Cunningham, C. J. (1989). Encoding difficulty and memory enhancement for young and older readers. *Psychology and Aging*, *4*(3), 333–338. doi:10.1037/0882-7974.4.3.333

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., Pennycook, G., Ackerman, R., Thompson, V. A., & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16–e30. doi:10.1037/xge0000049

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered?: It depends on your beliefs about intelligence. *Psychological Science*, *22*(3), 320–324. doi:10.1177/0956797610397954

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. doi:10.1037/a0019745

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. doi:10.3758/s13423-012-0343-6

Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1067–1087. doi:10.1037/0278-7393.22.5.1067

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 69–80. doi:10.1037/0278-7393.28.1.69

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacognition and Learning, 11*(1), 57–70. doi:10.1007/s11409-015-9151-5

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. doi:10.1111/j.1467-9280.2008.02189.x

Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning, 11*(1), 71–88. doi:10.1007/s11409-015-9146-2

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,* 973–978. doi:10.3758/s13423-011-0114-9

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. doi:10.3758/s13421-013-0323-8

Weltman, D., & Eakin, M. (2014). Incorporating unusual fonts and planned mistakes in study materials to increase business student focus and retention. *INFORMS Transactions on Education*, *15*(1), 156–165. doi:10.1287/ited.2014.0130

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. doi:10.3758/s13421-012-0255-8

Appendix B – Fluency-Manipulations

In all experiments of this thesis, fluency-manipulations that were assumed to affect metacognitive processes but that do not improve performance were used. This was necessary to investigate the research questions of this thesis (see section 1.6).

B.1 Conceptual Fluency: Experiment 1 and 2

In Experiment 1 and 2, fluency was manipulated by letter deletion, which is a manipulation of conceptual fluency. Intact texts were used as fluent texts and for the disfluent texts, all vowels in nouns, adjectives and verbs, except for initial letters, were deleted. This algorithm was adapted from the algorithm by McDaniel (1984; also used by Maki, Foley, Kajer, Thompson, & Willert, 1990, and Rawson & Dunlosky, 2002, Experiment 4) for German texts. After a pilot study with $N = 10$ students, all letters of four technical terms were completed because less than 50% of these students were unable to recognize these words. Figure B1 presents an example of a deleted letter text (in German).

Dies ist ein T_xt mit
f_hl_nd_n B_chst_b_n.
Alle V_k_l_ in N_m_n,
Adj_kt_v_n und V_rb_n f_hl_n.

Figure B1. Example of a deleted letter text.

References

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 609–616. doi:10.1037/0278-7393.16.4.609

McDaniel, M. A. (1984). The role of elaborative and schema processes in story memory. *Memory & Cognition*, *12*(1), 46–51. doi:10.3758/BF03196996

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 69–80. doi:10.1037/0278-7393.28.1.69

B.2 Perceptual Fluency: Experiment 3 and 4

In Experiment 3 and 4, perceptual fluency was manipulated in order to extent findings from Experiment 1 and 2. For disfluent texts, Mistral (18 point) was used and for fluent texts, Times New Roman (16 point) was used. The sizes of these font types were adapted so that they were comparable (see also Dreisbach & Fischer, 2011). Figure B2 presents an example of a disfluent text in the font type Mistral.

This is a disfluent text in the
font type mistral.
Font type is a manipulation
of perceptual fluency.

Figure B2. Example of a disfluent text in the font type Mistral.

References

Dreisbach, G., & Fischer, R. (2011). If it’s hard to read… try harder! Processing fluency as signal for effort adjustments. *Psychological Research*, *75*(5), 376–383. doi:10.1007/s00426-010-0319-y

Appendix C – Texts

In all experiments, expository texts from textbooks about psychology were used (Aronson, Akert, & Wilson, 2008; Kanning, 2002; Kauffeld, 2011; Rudolph 2009; Ulich & Marc, 2010). These texts were adapted in a way that students could understand the texts without having prior knowledge. Across all texts, the overall word count was comparable. In Experiment 1, 2, and 3, two texts were used, in Experiment 4, four texts were used. To ensure that the duration of the experiments was comparable, the texts in Experiment 4 were shorter than in Experiment 1, 2, and 3. For these shorter texts (Experiment 4), texts with somewhat higher Flesch-Kincaid grade-level scores were used than for longer texts (e.g., in Experiment 1, 2, and 3). As Flesch-Kincaid grade-level scores (computed by a tool from Michalke, 2012) show, the texts were comparable in difficulty, particularly within each experiment. Information about the topic, text length, and Flesch-Kincaid grade-level scores of the texts can be found in Table C1.

Table C1. *Topic, Text Length, and Flesch-Kincaid Grade-Level-Score for the German Texts*

	Topic	Text length	Flesch-Kincaid grade-level score
Experiment 1 and 2			
Text A	Rubicon model	936 words	18.94
Text B	Causal dimensions in attribution theory	929 words	20.92
Experiment 3			
Text A	Attitudes	1,019 words	19.69
Text B	Social skills	900 words	21.71
Experiment 4			
Text A	Personnel selection	449 words	23.89
Text B	Industrial engineering	406 words	25.90
Text C	Personnel marketing	430 words	24.31
Text D	Socio-technical systems approach and the human-technic-organization concept	390 words	24.57

References

Aronson, E., Akert, R. M., & Wilson, T. D. (2008). *Sozialpsychologie [Social psychology]*. München, Germany: Pearson Deutschland GmbH.

Kanning, U. P. (2002). Soziale Kompetenz-Definition, Strukturen und Prozesse [Social skills definition, structure, and processes]. *Journal of Psychology, 210*(4), 154–163. doi:10.1026//0044-3409.210.4.154

Kauffeld, S. (2011). *Arbeits-, Organisations- und Personalpsychologie für Bachelor. Lesen, Hören, Lernen im Web [Work-, oranizational, and personnel psychology for bachelor. Reading, listenting, learning on the web].* Berlin Heidelberg, Germany: Springer-Verlag.

Michalke, M. (2012). koRpus: A R package for text analysis (version 0.04-36). Available from http://reaktanz.de/?c=hacking&s=koRpus (Retrieved October 30, 2013).

Rudolph, U. (2009). *Motivationspsychologie kompakt [Motivational psychology compact]* (2nd ed.). Weinheim, Germany: Beltz PVU.

Ulich, E., & Marc, W. (2010). *Gesundheitsmanagement in Unternehmen [Health management in companies]* (4th ed.). Wiesbaden, Germany: Gabler Verlag. doi:10.1007/978-3-658-04485-5

Appendix D – Tests

In all experiments, performance was captured by a knowledge test. Additionally, in Experiment 3, one question about retention was asked. Table D1 gives an overview about the number of questions and statements, measures of the difficulty (mean), and Cronbach´s alpha of the knowledge tests. The tests were widely comparable in difficulty across the experiments. The reason for the low value of performance in the first test of Experiment 4 is that students were allowed to read the text only once, whereas they were allowed to learn the texts for all other test.

Table D1. *Reliability (Cronbach´s Alpha), Number of Items, and Descriptive Statistics (Means and Standard Deviations in Percentage) of the Knowledge Tests*

	Number of items	*M*	*SD*	Cronbach´s alpha
Experiment 1	23 (questions)	46.53	12.29	.85
	x 6 (statements)			
Experiment 2	+ 24 (questions)	51.07	10.38	.81
	x 6 (statements)			
Experiment 3	2 open questions			
	+ 30 (questions)			
	x 6 statements	53.24	10.40	.86
	x 2 (texts)			
Experiment 4				
First test	6 (questions)			
	x 4 (statements)	39.47	17.19	.70
	x 4 (texts)			
Final test	6 (questions)			
	x 4 (statements)	53.23	15.81	.70
	x 4 (texts)			

Each test consisted of questions with statements. Each statement had to be classified as true or false. Students received one point for each statement that was correctly classified as true or false. Performance scores on a text were computed, calculating the mean of all correct answers on this text. If different texts were presented in one type of fluency (Experiment 1 and 4), the mean of all text scores was built as a performance score. In Experiment 4, the 17 statements that correlated negatively with the remaining statements across all texts were excluded based on the item analysis (see paper IV, for more details). In all experiments, performance scores were corrected for guessing because the chance to guess a correct answer was 50%. The algorithm $200 \cdot x - 100$ transforms the value of guessing to zero ($200 \cdot 0.50 - 100 = 0$), resulting in a scale from 0 to 100, which is comparable to the judgment scales. This is a prerequisite to compute absolute monitoring accuracy.

Appendix E – Instruments: Monitoring and Control

E.1 Measuring Monitoring

To measure monitoring, students were asked for different types of judgments during the learning process. All of these judgments were answered on a continuos visual analogue scale from 0 to 100 by typing in the chosen value (integer number).

E.1.1 Ease of learning judgment (EOL judgment)

In all experiments, students were asked for ease of learning judgments (EOL judgments): "How easy or difficult is it to learn the text?" (0 = *difficult*, 50 = *middle*, 100 = *easy*).
Figure E1 presents the German question and the scale.

Wie leicht bzw. schwer schätzen Sie den Text zu lernen ein?

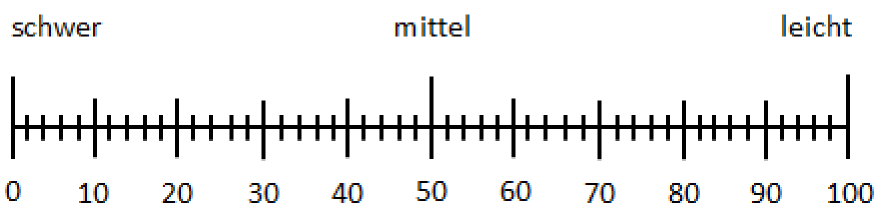

Figure E1. Measuring ease of learning judgments.

E.1.2 Familiarity judgment

In Experiment 3, students were further asked for familiarity judgments:
"How familiar are you with the topic of the text?" (0 = *not familiar*, 50 = *middle*, 100 = *familiar*). Figure E2 presents the German question and the scale.

Wie gut sind Sie mit dem Thema des Textes vertraut?

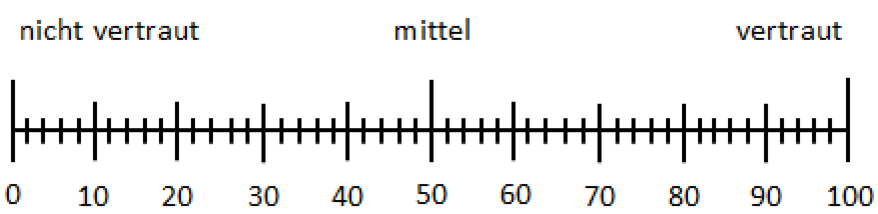

Figure E2. Measuring familiarity judgments.

E.1.3 Comprehension judgment

Moreover, comprehension judgments were captured in Experiment 3: "How easy was it to understand the text?" (0 = *difficult*, 50 = *middle*, 100 = *easy*). Figure E3 presents the German question and the scale.

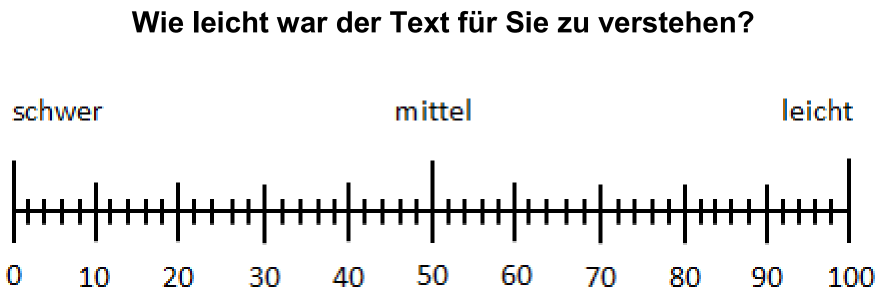

Figure E3. Measuring comprehension judgments.

E.1.4 Judgment of learning (JOL)

In all experiments, students were asked for judgments of learning (JOL): "What percentage of questions about the text will you answer correctly?" (0 = *none*, 50 = *half*, 100 = *all*). Figure E4 presents the German question and the scale.

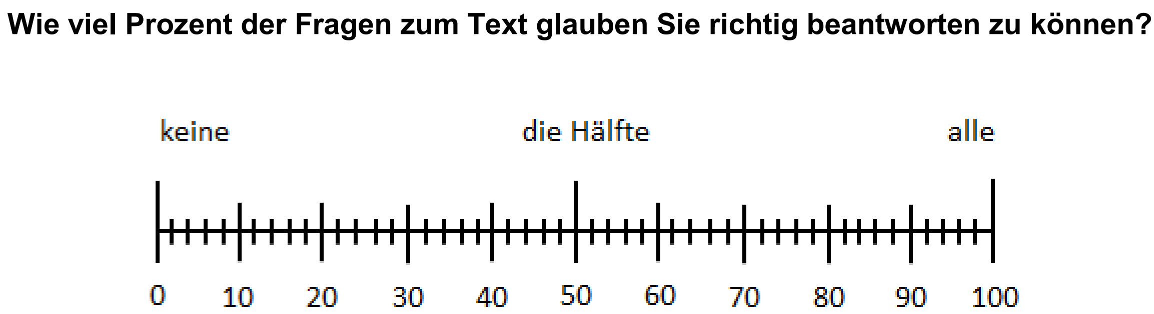

Figure E4. Measuring judgments of learning.

E.1.5 Retrospective confidence judgment (RC judgment)

In all experiments, students were asked for retrospective confidence judgments (RC judgments): “How confident are you that your answer is correct?” (0 = *unconfident*, 50 = *middle*, 100 = *confident*). Figure E5 presents the German question and the scale.

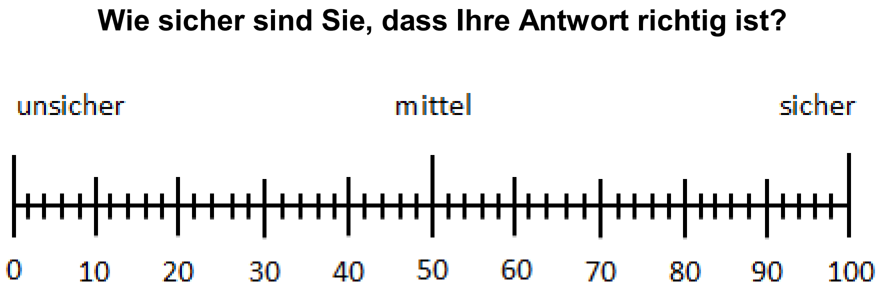

Figure E5. Measuring retrospective confidence judgments.

E.2 Measuring Control

As measures for control, reading-time and termination of study were used in all experiments. Reading-time served as a manipulation-check because disfluency should lead to less automatic processes and thus, should directly affect reading-time when reading a text once. Inversely, when rereading the texts, students were allowed to make notes. Making notes as well as further factors (e.g., monitoring) might affect termination of study, too. Moreover, in Experiment 3, two subjective measures of control of cognitive processes (see Paas, 1992; Rey & Nieding, 2010) were used to investigate not only how much time students invest in rereading a text, but also to investigate the perceived difficulty and the invested mental effort during rereading. Moreover, in Experiment 4, students were further allowed to select texts for rereading before (first text selection) and after (second text selection) a knowledge test by typing in the numbers of the texts that they wanted to select for rereading. Next, the questions and scales of the two subjective measures of perceived difficulty and mental effort are presented.

References

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), 429–434. doi:10.1037/0022-0663.84.4.429

Rey, G. D., & Nieding, G. (2010). Arbeitsgedächtnis und Cognitive Load [Working memory and cognitive load]. In H.-P. Trolldenier, W. Lenhard, & P. Marx (Eds.), *Brennpunkte der Gedächtnisforschung. Entwicklungs- und pädagogisch-psychologische Perspektiven [Focal points of memory research. Developmental and educational-psychological perspectives]* (pp. 67–85). Göttingen, Germany: Hogrefe.

E.2.1 Perceived difficulty

In Experiment 3, students were asked for their perceived difficulty during rereading: “How easy or difficult was it to learn the text?” (0 = *difficult*, 50 = *middle*, 100 = *easy*). Figure E6 presents the German question and the scale.

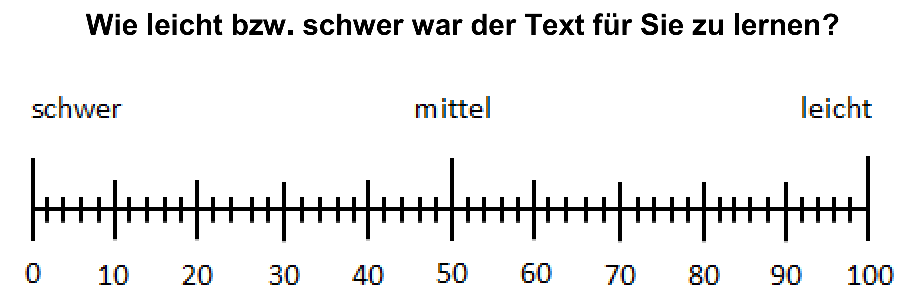

Figure E6. Measuring perceived difficulty.

E.2.2 Mental effort

In Experiment 3, students were further asked for their mental effort during rereading: “How much mental effort did you invest during learning?” (0 = *none*, 50 = *middle*, 100 = *high*). Figure E7 presents the German question and the scale.

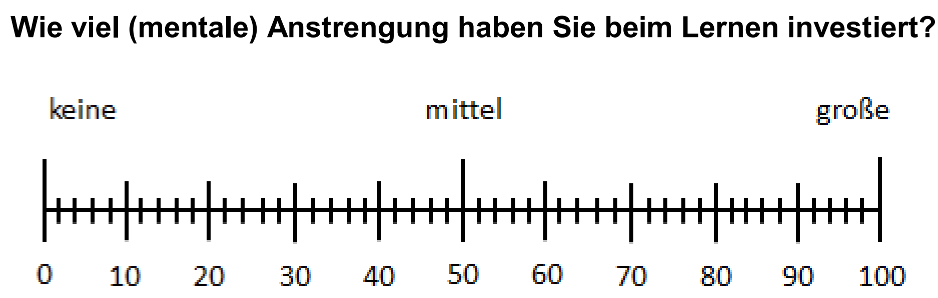

Figure E7. Measuring mental effort.

Appendix F – Procedure

Next, the procedure of the experiments is described. For each experiment, a schematic overview about the procedure is given. A more detailed description of the procedure of each experiment can be found in the papers. The experiments took about 90 minutes to 2 hours.

F.1 Procedure of Experiment 1 and 2

Figure F1 presents an overview about the procedure of Experiment 1 and 2. The acquisition process was done with the first and, afterwards, with the second text (texts were randomized). Afterwards, retrieval took place by means of a knowledge test about both texts (with the same order of the texts in the acquisition process), and students made RC judgments after each statement of the knowledge test.

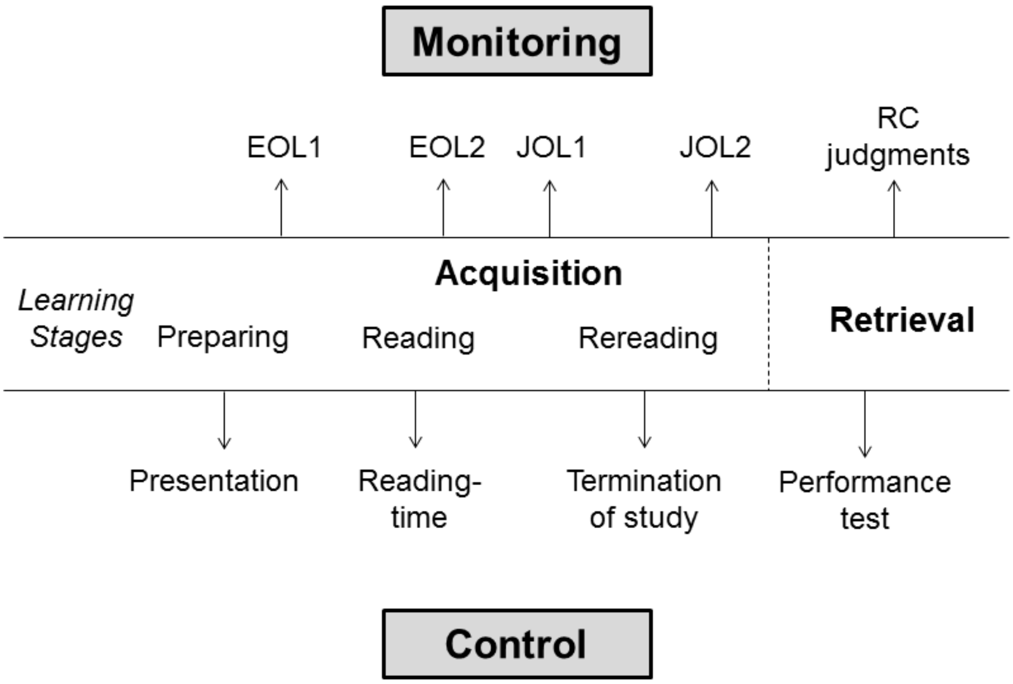

Figure F1. Procedure of Experiment 1 and 2.
EOL = ease of learning judgment; JOL = judgment of learning; RC judgment = retrospective confidence judgment.

F.2 Procedure of Experiment 3

Figure F2 presents an overview about the procedure of Experiment 3. As in Experiment 1 and 2, first, the acquisition process was done with the first and, afterwards, with the second text (texts were randomized). Afterwards, retrieval took place by a knowledge test about both texts (with the same order as the texts in the acquisition process). After each question of the knowledge test, students made RC judgments.

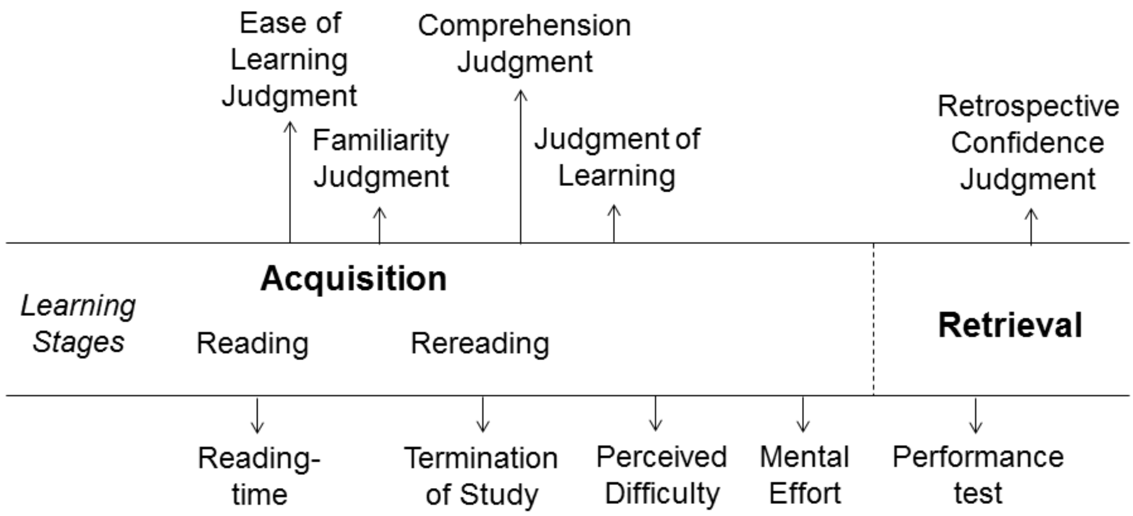

Figure F2. Procedure of Experiment 3.

F.3 Procedure of Experiment 4

Figure F3 presents an overview about the procedure of Experiment 4. First, the texts were presented for 2 seconds on the screen, and immediately after each text, an EOL judgment was made. Afterwards, students read each text once and made JOLs immediately after each text. Before and after the knowledge test, with RC judgments after each statement, students selected texts for rereading. In the second phase, students were allowed to reread the texts and made JOLs after each text. Finally the knowledge test with RC judgments was repeated. For more details and the reasoning of the procedure, see paper IV.

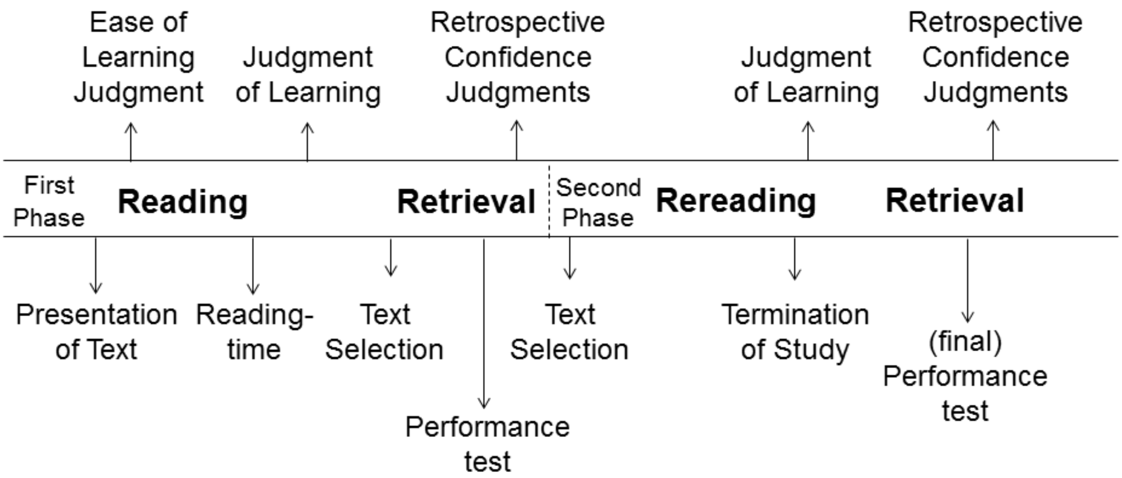

Figure F3. Procedure of Experiment 4.

Appendix G – Papers

G.1 Paper I: Fluency and Metacognition – Is Fluency Used as a Cue for Judgments During and After the Learning Process?

Pieger, E., Mengelkamp, C., & Bannert, M.. Fluency and metacognition – Is fluency used as a cue for judgments during and after the learning process?

The author Elisabeth Pieger autonomously developed the research presented in this thesis. She is further the leading author of this paper. Elisabeth Pieger had the leading role in developing, managing, analyzing and presenting the experiments. This includes the conceptualizing, preparation, management, data-analysis, discussion, and presentation of the experiment. Moreover, she was leading during publishing the paper. The co-authors Maria Bannert and Christoph Mengelkamp advised the publication process. They were discussion-partners and provided feedback on the paper.

Abstract

Metacognitive judgments play an important role in educational contexts because they can affect further learning. Previous research shows that fluency can affect students` metacognitive judgments. However, the focus has been on judgments of learning when learning with word-pairs. Thereby, perceptual fluency usually is manipulated within-person. The aim of this study is to investigate whether conceptual fluency affects different types of judgments during learning with texts, when fluency is manipulated between-person. Students learned with two intact ($N = 32$) or deleted letter texts ($N = 31$) and made different types of judgments. Results show that students judge disfluent texts as more difficult than fluent texts after reading. Furthermore, students predict lower performance for disfluent than for fluent texts, although performance was not affected by fluency. Fluency further did not affect retrospective confidence judgments. Future research should investigate how the effects of fluency on monitoring can be used to improve performance.

Keywords: metacognitive judgments; fluency; metacomprehension; learning

1. Introduction

At schools and universities, reading texts is very common, e.g., when learning for an exam. But, just reading a text is not enough to master an exam, students also have to make sure that they understand and retain what they have read (*metacomprehension*; see Maki & Berry, 1984). Therefore, students have to monitor their learning by making different types of judgments during the entire learning process (Nelson & Narens, 1990). This is important because students` judgments about the material and their own performance can affect further learning (e.g., Nelson & Leonesio, 1988; Thiede & Dunlosky, 1999; see Son & Kornell, 2008, and Son & Metcalfe, 2000, for an overview) and thus, performance (e.g., Thiede, Anderson, & Therriault, 2003). This is also supposed in theories about metacognition and self-regulated learning (e.g., Boekaerts, 1997; Nelson & Narens, 1990; Winne & Hadwin, 1998; Zimmerman, 1990). Altogether, metacognitive judgments play an important role in educational contexts. Thus, it is important to know on which cues different metacognitive judgments are based.

According to the cue utilization approach (Koriat, 1997), ease of processing (*fluency*) should be a cue for judgments. Fluency as a cue for judgments is often investigated by asking students to predict their performance when learning with words or word-pairs, although in educational contexts, students usually have to learn from texts. When learning with texts, different levels of text-processing should be monitored, and thus, students should make different types of judgments (Pieger, Mengelkamp, & Bannert, 2017). Moreover, in previous research, fluency has been found to affect judgments in within-person designs (e.g., Mueller, Tauber, & Dunlosky, 2013, Experiment 2; Yue, Castel, & Bjork, 2013, Experiment 1a, 2a, 2b, and 3). However, when learning only with fluent or disfluent word-pairs, fluency often is not used as a cue for judgments (e.g., Susser, Mulligan, & Besken, 2013; Yue et al., 2013, Experiment 1b). The aim of this study is to investigate whether fluency affects different types of judgments during the learning process when fluency is manipulated between-person and students learn with texts. We assume that letter deletion should affect metacognitive judgments when learning with texts in a between-person design (see next sections).

1.1 Fluency and Metacognition

In theories about metacognition and self-regulated learning (e.g., Boekaerts, 1997; Nelson & Narens, 1990; Winne & Hadwin, 1998; Zimmerman, 1990), monitoring and control are central components. Koriat, Ma'ayan, and Nussinson (2006) define control by the regulation of cognitive processes, whereas monitoring refers to the assessment of these processes. Monitoring and control interact (Koriat, 2012; Koriat et al., 2006): Students can use their monitoring to control their learning (*monitoring-based control*), e.g., they might invest more time when they judge the material as difficult to learn. Control can further affect monitoring (*control-based monitoring*), e.g., students might judge material as difficult to learn when processing requires a lot of time.

Fluency affects how easy (*fluent*) or difficult (*disfluent*) material can be processed (e.g., Alter & Oppenheimer, 2009; Lanska, Olds, & Westerman, 2014; Oppenheimer, 2008; Schwarz, 2010). When students learn with disfluent material, processing is slowed down, which is supported by longer reading-times for disfluent than for fluent material (e.g., Eitel & Kühl, 2016; McDaniel, 1984, Pilot of Experiment 1; McDaniel, Einstein, Dunay, & Cobb, 1986, Experiment 1 and 2; Miele & Molden, 2010, Experiment 3; Pieger, Mengelkamp, & Bannert, 2016; Sanchez & Jaeger, 2015, Experiment 1 and 2). Therefore, students control their learning by slowing down processing due to disfluency. This experience can affect students` monitoring: Schwartz and Efklides (2012; see also Efklides, 2009) suppose that students` judgments are informed by their metacognitive experiences during reading. Similarly, Koriat (1997) describes fluency as an experience-based cue for judgments. The *ease of processing heuristic* (Kornell, Rhodes, Castel, & Tauber, 2011) as well as the *ELER-heuristic* suppose that material that is "Easily Learned" is "Easily Remembered" (Koriat, 2008; Miele, Finn, & Molden, 2011). Hence, students infer learning and remembering from fluency. Disfluent material is difficult to process and students should infer difficulties in learning and remembering from encoding difficulties. Thus, disfluent material is not only difficult to process but further disfluency leads to the judgment that the learning material is difficult to learn and difficult to remember (see Figure 1).

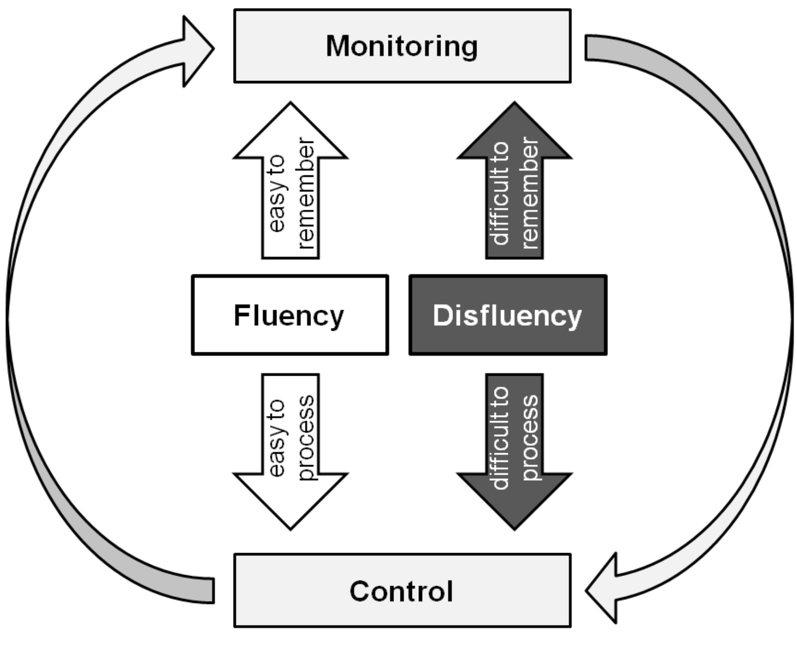

Figure 1. Interplay between fluency, monitoring, and control.

1.1.1 Fluency effects on monitoring. In previous research, there is some evidence for *fluency effects* (difference between fluent and disfluent material) on judgments. However, the focus of previous research has been on words and word-pairs, and thereby, fluency was almost varied within-person. In these studies, students made lower performance predictions for disfluent than for fluent material (e.g., Magreehan, Serra, Schwartz, & Narciss, 2016, Experiment 4 and 5; Mueller et al., 2013, Experiment 2, for word-pairs; Besken & Mulligan, 2013; Kornell et al., 2011; Susser et al., 2013, for mixed word-lists in Experiment 1 and 2; Yue et al., 2013, Experiment 1a, 2a, 2b, and 3, for words or word-lists; Rawson & Dunlosky, 2002, Experiment 4, for texts). Inversely, there are fewer studies in which fluency was manipulated between-person (e.g., Magreehan et al., 2016, Experiment 1, 2, and 3, for word-pairs; Maki, Foley, Kajer, Thompson, & Willert, 1990, Experiment 2, for texts; McDaniel et al., 1986, for texts; Susser et al., 2013, Experiment 1 and 3, for pure word-lists; Yue et al., 2013, Experiment 1b, for word-lists). In these studies, judgments were not affected by fluency.

A possible reason for these findings is that the experience of fluency is more obvious in within-person designs than in between-person designs. When fluency is manipulated within-person, students experience both fluent and disfluent processing. Therefore, fluency is obvious, and students are able to compare fluent and disfluent material when making judgments (see also Susser et al., 2013). In between-person designs, students experience only one type of fluency (fluency vs. disfluency), and therefore, fluency was not an obvious cue for judgments. But, because there is little research that investigates fluency effects on judgments in between-person designs, only a limited number of fluency-manipulations has been considered. Different types of fluency can be more or less obvious, and thus, there may be types of fluency that affect judgments, even in between-person designs (see also Susser et al., 2013, for this suggestion). For example, font size might be a fluency-manipulation that is too subtle and only obvious in within-person designs where different sizes can be compared (e.g., Susser et al., 2013, Experiment 1).

Font size is a manipulation of *perceptual fluency*. Perceptual fluency describes how easily a word can be identified (Schwarz, 2010) and is a manipulation of *processing fluency*. *Conceptual fluency* is another type of processing fluency and describes how easily the meaning of words and knowledge structures can be identified (Schwarz, 2010). Deleted letter texts are a manipulation of conceptual fluency because this fluency-manipulation stimulates propositional processing (see McDaniel & Butler, 2011, for an overview). Students can generate the meaning of a word by activating their knowledge from the context (see also Bjork & Storm, 2011; Bjork & Yue, 2016). Therefore, conceptual fluency affects the experience of fluency/disfluency in processing on higher levels than perceptual fluency does (Schwarz, 2010). Thus, conceptual fluency should be particularly relevant for monitoring learning with texts. We assume that deleted letter texts, therefore, should be an obvious cue for judgments, even on higher levels of text-processing and when students learn only with fluent texts or only with disfluent texts. Next, different types of judgments and their relation to different levels of text-processing are described.

1.1.2 Judgments and levels of processing. When learning with texts, processing takes place on different levels of text-processing. The deciphering of words takes place on the *surface level* of text-processing. Higher levels of text-processing include the acquisition of the meaning and the relation between words (*textbase level*) as well as the elaboration of the content to memorize it (*situation model*, Kintsch, 1994; see also De Bruin & van Gog, 2012; Redford, Thiede, Wiley, & Griffin, 2012). To monitor different levels of text-processing, students can make different types of judgments during the learning process (see Pieger et al., 2017).

Ease of learning judgments (EOL judgments) are usually made at the beginning of the learning process by judging the ease or difficulty to learn the text (Nelson & Narens, 1990), which is relevant when making plans for learning. In this phase, students try to get an overview of the learning material and have not yet learned the text. Therefore, ease of learning judgments can be related to the surface level of text-processing. For this reason, ease of learning judgments might have often been neglected in previous research (e.g., Jönsson & Lindström, 2010; Pieger et al., 2016).

Judgments of learning (JOLs) refer to students` judgments about their actual or future performance and were in the focus of previous research (see above). Dependent on the stage of the learning process, these judgments can refer to the textbase level or to the situation model. When making JOLs after reading a text once, JOLs should mainly refer to the textbase level. When making JOLs after learning and memorizing the text, JOLs should mainly refer to the situation model of text-processing.

Retrospective confidence judgments (RC judgments) are usually made during and after retrieval, e.g., during the knowledge test. Thus, these judgments mainly refer to students` memory and to the integration of knowledge into memory (situation model). Like for EOL judgments, there is little research on the effects of fluency on RC judgments (e.g., Pieger et al., 2016). However, Dinsmore and Parkinson (2013) found that text characteristics are most frequently used as a cue for these judgments. Therefore, conceptual fluency might be a text characteristic that is used for RC judgments.

Summing up, the aim of the current study is to investigate whether fluency is a cue for judgments in a between-person design when learning with texts. In the current study, we thereby focus on conceptual fluency because we assume that conceptual fluency affects monitoring of different levels of text-processing. Fluency should be obvious and hence, be used as a cue for different types of judgments during the learning process.

1.1.3 Deleted letters, performance, and monitoring. Conceptual fluency might not only affect monitoring of higher levels of text-processing but also improve performance (e.g., Maki et al., 1990, Experiment 1). This is useful in educational contexts (see also Pieger et al., 2016). But, to be able to investigate whether fluency affects different types of judgments, performance should not be affected by fluency. Otherwise, retrieval could be used as an additional cue for judgments that covers fluency effects on judgments (see also Maki et al., 1990, Experiment 1; Susser et al., 2013, Experiment 3). Letter deletion, however, affects performance only under some conditions (see McDaniel & Butler, 2011, for an overview).

One factor that seems to play a role for fluency effects on performance is the design (within-person vs. between-person). There is some evidence that deleted letters increase performance in within-person designs (e.g., Maki et al. 1990, Pilot and Experiment 1; McDaniel, 1984; McDaniel, Ryan, & Cunningham, 1989), but not in between-person designs (e.g., Maki et al., 1990, Experiment 2; McDaniel et al., 1986, Experiment 1). Maki et al. (1990) assume that in within-person designs, students pay more attention to deleted letter texts at the cost of intact texts. Therefore, fluency should not affect performance in a between-person design.

Another factor that determines whether deleted letters affect performance is the instruction on how to deal with this fluency-manipulation (mentally filling in vs. writing-in deleted letters; Rawson & Dunlosky, 2002, Experiment 4; see also Maki et al., 1990, Experiment 1). This factor can be subsumed under *encoding operations* in the framework of *desirable difficulties* (see McDaniel & Butler, 2011). Writing-in deleted letters seems to improve performance, whereas mentally filling in deleted letters does not seem to be a sufficient encoding operation to improve performance (Rawson & Dunlosky, 2002, Experiment 4). Thus, to test if judgments are based on fluency, letter deletion can be used when students mentally fill in deleted letters. Writing-in deleted letters vs. mentally filling in deleted letters might not only impact the fluency effect on performance. Writing-in deleted letters into texts has not been found to affect judgments in a between-person design (e.g., Maki et al., 1990, Experiment 2; McDaniel et al., 1986). This could be due to the fact that writing-in deleted letters erases the fluency effect on judgments: After writing-in deleted letters, the text is no longer a deleted letter text. Because students make judgments after they wrote-in deleted letters, fluency might no longer be an obvious cue for judgments. However, when mentally filling in deleted letters, fluency should still be obvious as a cue for judgments because letters are still deleted when making judgments.

Besides the design and the encoding operations, the material is a further determinant that affects which type of difficulty is desirable and improves performance: According to McDaniel and Butler (2011), letter deletion stimulates propositional processing and, hence, higher levels of text-processing. Because expository texts also stimulate propositional processing, letter deletion should not improve performance for expository texts (see also McDaniel & Butler, 2011). Inversely, letter deletion should improve performance for narrative texts like fairy tales, because these texts stimulate relational processing and thus, letter deletion fosters a level of processing that is not already stimulated by the text itself (see McDaniel & Butler, 2011, for an overview of evidence). Thus, when using expository texts letter deletion should not affect performance.

Summing up, to be able to test whether fluency (manipulated by letter deletion) is used as a cue for judgments, performance should not be affected. This should be the case when students mentally fill in deleted letters in expository texts and when they only learn with intact or with deleted letter texts (between-person design). Inversely, mentally filling in deleted letters should affect monitoring.

1.2 Research Questions and Hypotheses

The aim of the current study is to investigate whether students use fluency as a cue for judgments in a between-person design. Thereby, the effect of conceptual fluency (i.e., letter deletion) on monitoring of different levels of text-processing is investigated. Students mentally fill in deleted letters to ensure that performance would not be affected by fluency and thus, to ensure that fluency effects on judgments are neither due to performance differences nor covered by performance differences.

We suppose that conceptual fluency should be an obvious cue for judgments when learning with texts: Students should judge disfluent compared with fluent texts as more difficult to learn (EOL judgment), they should judge their performance as lower for disfluent than for fluent texts (JOL) and they should be less confident in the correctness of their answers in a knowledge test (RC judgments).

2. Method

2.1 Participants and Design

Data were collected from $N = 63$ students from a university in Germany (age: $M = 20.76$, $SD = 2.86$ years, 77.78% female). Form these students, $N = 52$ (82.54%) students studied media communication and $N = 11$ (17.46%) students studied human-computer-interaction (semester of studies: $M = 2.71$, $SD = 1.84$). The required sample size of $N = 60$ was computed by G*-Power (Faul, Erdfelder, Lang, & Buchner, 2007) and is fulfilled. Thereby, a Type I error of .05, a power of .80, and an effect size of $f = .30$ was assumed because we expected bigger effects for conceptual than for perceptual fluency but smaller effects than in a within-person design (see also Maki et al., 1990; Rawson & Dunlosky, 2002, Experiment 4). Fluency was manipulated between groups. Students in the *fluent group* ($N = 32$) read an intact text and students in the *disfluent* *group* ($N = 31$) read a deleted letter text. Students were randomly assigned to one of the two groups. As dependent variables, monitoring (EOL judgments, JOLs, and RC judgments), control (reading-time and termination of study), and performance were captured.

2.2 Material

As learning material, we used two expository texts. These texts were about motivational psychology (Rudolph, 2009) and depicted the rubicon model (Text A, 936 words, Flesch-Kincaid grade-level score = 18.94) or causal dimensions in attribution theory (Text B, 929 words, Flesch-Kincaid grade-level score = 20.92). Length and difficulty of these texts were comparable as scores computed by a tool by Michalke (2012) show.

For the disfluent text, letters of an intact text were deleted based on an algorithm by McDaniel (1984; see also Maki et al., 1990; Rawson & Dunlosky, 2002, Experiment 4) that we adapted for German texts. This algorithm includes deletion of all vowels in nouns, adjectives and verbs, excluding initial letters. Each deleted letter was replaced by one underscore. Between each word, five blank spaces were inserted to ensure enough space to clearly separate words (see Figure 2, for an example of a deleted letter text). After the pilot study with $N = 10$ students, all letters of four technical terms were filled in for the experiment because more than 50% of the students were not able to write-in the deleted letters of these terms.

Dies ist ein T_xt mit
f_hl_nd_n B_chst_b_n.
Alle V_k_l_ in N_m_n,
Adj_kt_v_n und V_rb_n f_hl_n.

Figure 2. Text with deleted letters.

Inverse to the pilot study, students mentally filled in deleted letters of the texts in the main study because previous research shows that mentally filling in deleted letters does not affect performance (e.g., Rawson & Dunlosky, 2002, Experiment 4; see also Maki et al., 1990), but it might affect judgments. To be able to test if fluency is a cue for judgments, it is important to ensure that fluency does not affect performance (Pieger et al., 2016). Moreover, letter deletion seems to foster performance only when students learn with both fluent and disfluent texts (e.g., Maki et al., 1990).

2.3 Instruments

We captured reading-time as a manipulation-check. To capture students` judgments, we asked students to type in integer numbers of a continuous visual scale from 0 to 100. EOL judgments captured “How easy or difficult is it to learn the text?” (scale labeles: 0 = *difficult*, 50 = *middle*, 100 = *easy*). JOLs captured “What percentage of questions about the text will you answer correctly?” (scale labels: 0 = *none*, 50 = *half*, 100 = *all*). RC judgments captured “How confident are you that your answer is correct?” (scale labels: 0 = *unconfident*, 50 = *middle*, 100 = *confident*).

The knowledge test was made-up of 23 questions on text A, and of 24 questions on text B. For each question, six statements were presented on separate screens and students typed in if the statement was true or false. They received one point when their answer was correct. Performance score was the mean of all points, weighted by texts. Because each statement had a 50% chance of being guessed correctly, the mean was corrected using the algorithm $200 \cdot x - 100$. This algorithm transforms the value of guessing to zero ($200 \cdot 0.50 - 100 = 0$). Reliability was computed by Cronbach´s Alpha, $\alpha = .85$ ($M = 46.43\%$, $SD = 12.29\%$). Figure 3 shows an example of a question with 6 statements.

Question:

Which statements are correct, regarding to the transition from the pre-actional to the actional phase?

Statements:

a) I have a concrete plan and ignore all stimuli to be able to focus on my aim.
b) I implement my realization plans in actions.
c) I have a concrete plan but still react flexible on circumstances.
d) I envision the possible consequences of my intention.
e) I focus my attention on the execution and realize my plan.
f) I cannot realize my plans, therefore I think about new alternatives.

Figure 3. Six statements of a question from the knowledge test. All statements were presented separately.

2.4 Procedure

Data were collected using *E-Prime-Software* (E-Prime Professional 2.0). The duration of the experiment was about 2 hours. To capture the judgments during the learning process, the first text was presented for 2 seconds and students had to type in their EOL judgment (EOL1) on the next screen. After students read the text once (reading-time was captured as a manipulation-check), they made another EOL judgment (EOL2) and a JOL (JOL1). Afterwards, students re-read the text for a maximum of 15 minutes, and thereby, it was possible to make notes. Then, students were asked for another JOL (JOL2). This procedure was repeated with the second text (texts were randomized) before the knowledge test was conducted. After each statement of the knowledge test, RC judgments were captured. The procedure is visualized in Figure 4.

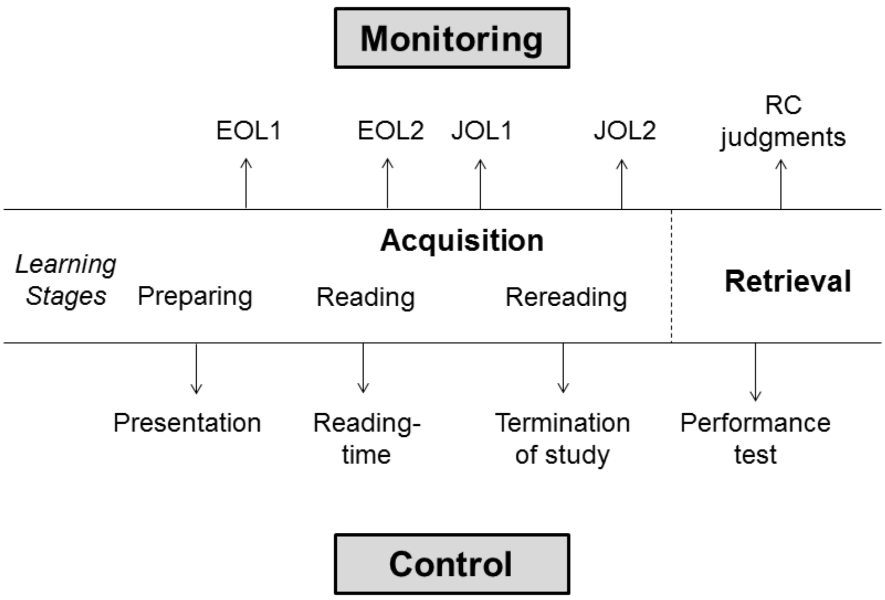

Figure 4. Procedure.

3. Results

Descriptive statistics of the dependent variables (i.e., judgments, reading-time, termination of study, and performance) in the fluent and the disfluent group can be found in Table 1. We computed a MANOVA to test if there were significant multivariate effects of fluency (fluent vs. disfluent) on dependent variables, $V = .409$, $F(8, 54) = 4.67$, $p < .001$, $\eta_p^2 = 0.41$. Correlations between dependent variables across groups can be found in Appendix A, Table A1. In the next section, we describe fluency effects on the dependent variables in more detail.

Table 1. *Descriptive Statistics of Dependent Variables Between the Fluent and the Disfluent Group*

	Fluent group (N = 32)		Disfluent group (N = 31)	
Variable	*M*	*SD*	*M*	*SD*
Control				
Reading-time	5.38	1.69	7.34	1.83
Termination of study	11.81	2.15	11.72	1.85
Monitoring				
EOL1	39.06	21.40	37.97	22.07
EOL2	51.25	13.44	41.85	20.73
JOL1	50.02	14.27	38.00	16.60
JOL2	60.42	15.42	52.16	16.04
RC judgment	72.80	13.63	70.14	14.40
Performance	49.01	12.19	45.06	12.35

Note. Judgment and performance scales ranged from 0 to 100; reading-time and termination of study in minutes.

Reading-time was significantly longer in the disfluent than in the fluent group, and thus, our manipulation was successful, $F(1, 61) = 19.53$, $p < .001$, $\eta_p^2 = 0.24$. We found no significant fluency effect on EOL1, $F(1, 61) = 0.04$, $p = .842$, $\eta_p^2 < 0.01$. After reading a text once, students in the disfluent group judged the text as more difficult than students in the fluent group, as can be seen from the results for EOL2, $F(1, 61) = 4.58$, $p = .036$, $\eta_p^2 = 0.07$. Furthermore, the students predicted lower performance after reading (JOL1) in the disfluent group compared to the fluent group, $F(1, 61) = 9.51$, $p = .003$, $\eta_p^2 = 0.13$, and students in the disfluent group judged their performance lower after rereading (JOL2) than in the fluent group, $F(1, 61) = 4.35$, $p = .041$, $\eta_p^2 = 0.07$. For RC judgments, $F(1, 61) = 0.66$, $p = .421$, $\eta_p^2 = 0.01$, for performance, $F(1, 61) = 1.63$, $p = .206$, $\eta_p^2 = 0.03$, and for termination of study, $F(1, 61) = 0.04$, $p = .849$, $\eta_p^2 < 0.01$, no fluency effects were found.

4. Discussion

The aim of this study was to investigate whether conceptual fluency is used as a cue for different types of judgments when learning only with fluent or disfluent texts. Thereby, we manipulated conceptual fluency by asking students to mentally fill in deleted letters. We assumed that this fluency-manipulation is obvious during text-processing. Results support our assumption for judgments after reading (EOL2 and JOL1) and after rereading (JOL2), but not after a text presentation for 2 seconds on the screen (EOL1) or during the test (RC judgments), which is discussed next.

When reading disfluent texts, students controlled their learning by slowing down processing. Therefore, students seem to have experienced fluency/disfluency during reading, which is supported by longer reading-times for disfluent than for fluent texts and is also consistent with prior research (e.g., McDaniel, 1984, Pilot of Experiment 1; McDaniel et al., 1986, Experiment 1 and 2). Inversely, the finding that fluency affected students` judgments after reading and rereading, although students experienced only fluency or disfluency, goes beyond the findings of previous research. In previous research, which focused on words and word-pairs, usually no fluency effects on judgments were found in between-person designs (e.g., Magreehan et al., 2016, Experiment 1, 2, and 3, for word-pairs; Yue et al., 2013, Experiment 1b, for word-lists). However, because research that investigated fluency effects in between-person designs is limited, we assumed that there are fluency-manipulations that affect judgments, even in between-person designs (see also Susser et al., 2013). This assumption is supported by our findings that letter deletion was an obvious cue for students` judgments (EOL2, JOL1, and JOL2) even in a between-person design. Thus, conceptual fluency (i.e., letter deletion) seems to be a fluency-manipulation that affects judgments after reading and rereading. Thereby, it seems to be important that students mentally fill in deleted letters. Previous research shows that writing-in deleted letters does not affect judgments in between-person designs (Maki et al., 1990, Experiment 2; McDaniel et al., 1986). When making judgments after writing-in deleted letters, however, letters are no longer deleted, and therefore, fluency seems to be not an obvious cue for judgments. This seems to be different when students mentally fill in deleted letters. In this case, the manipulation of conceptual fluency is still obvious when making judgments after reading and rereading.

However, conceptual fluency did not affect judgments after a text presentation for 2 seconds (EOL1) or during the text (RC judgments). Therefore, we assume that the fluency effect on judgments depends not only on the manipulation of fluency but also on the stage of the learning process and the type of judgment. One reason for the finding that students made equal EOL1 for the fluent and the disfluent texts is that these judgments are made at a very early learning stage, i.e., after a text presentation for 2 seconds on the screen. Therefore, students only got an impression of the text and were unable to process the text on higher levels of text-processing. Because conceptual fluency affects processing on higher levels, conceptual fluency might not have affected EOL1 that refer to surface levels of processing. Therefore, these judgments can be affected by perceptual fluency (Pieger et al., 2016) but not by conceptual fluency. However, EOL2 after reading were affected by fluency. Hence, one type of judgment might refer to different levels of processing dependent on the learning stage. Reading a text once not only provides more information about the text but also might provide a stronger experience of disfluency. Fluency is more obvious and is used as a cue for EOL2. Inversely, when presenting a text for 2 seconds, the experience of disfluency is not very strong. Thus, not only the type of fluency and its relation to text-processing but also the extent of the experience of disfluency might determine if fluency is an obvious cue for judgments in a between-person design.

This is also supported for RC judgments. Although conceptual fluency affects processing on higher levels of text-processing and RC judgments refer to higher levels of text-processing (situation model), RC judgments were not affected by conceptual fluency. This finding is consistent with the findings by Pieger et al. (2016; see also Maki et al., 1990), who also found no fluency effects on RC judgments. However, Pieger et al. (2016) used perceptual fluency that affects processing on lower levels than conceptual fluency. Maki et al., (1990) asked students to write-in deleted letters, and hence, fluency might not have been obvious. Thus, the current study extends previous research to conceptual fluency (i.e., mentally filling in deleted letters). Both, perceptual and conceptual fluency seem to be not obvious when making RC judgments. This might also be due to the fact that when making RC judgments, students did not experience disfluency because there were no performance differences between fluent and disfluent texts. Inversely, for judgments after reading and rereading, fluency was an obvious cue for judgments. Students seem to use fluency only as a cue when fluency is obvious, which can result when students experience disfluency during the task. Although Dinsmore and Parkinson (2013) found that students use text characteristics as cues for their RC judgments, fluency does not seem to be a cue students use for RC judgments. When making RC judgments, students might have used retrieval as the most obvious cue because they did not experience fluency/disfluency during the test. Because performance was equal between fluent and disfluent texts, using retrieval as a cue should not induce fluency effects on RC judgments. However, future research is required to investigate the underlying mechanisms and conditions to test when and how fluency affects judgments and which cues students might use instead of fluency (e.g., retrieval).

The finding that fluency did not affect performance was expected. We assumed that letter deletion does not affect performance when students either learn with fluent or with disfluent expository texts and when they mentally fill in deleted letters. This was necessary to investigate fluency effects on monitoring. Nevertheless, in educational settings, it would be helpful to improve performance. Asking students to write-in deleted letters might foster performance (see Rawson & Dunlosky, 2002, Experiment 4). However, this seems to be the case particularly in within-person designs (e.g., Maki et al., 1990) and not necessarily for expository texts (see McDaniel & Butler, 2011, for an overview). Hence, the interplay between different factors seems to play a role for fluency effects on performance. Although there is a lot of research on fluency effects on performance, research is still often unable to predict fluency effects on performance (especially for perceptual fluency). Thus, a systematic investigation of different factors that might affect fluency effects on performance is required. Besides the factors described above (e.g., interplay between the learning material, the design, and encoding operations), metacognition might be a further determinant for fluency effects on performance. According to theories about metacognition and self-regulated learning, adequate metacognitive processes can improve performance. Thus, knowing on which cues students base their judgments is useful to be able to affect these cues and thus, judgments. Thereby, the type of judgment, the learning stage, and the type of fluency seem to play a role. Future research is required to investigate the underlying mechanisms of how cues like letter deletion affect monitoring. Letter deletion might affect judgments because of the experience of disfluency and/or because of the belief that a disfluent text is difficult to learn and to remember (see ELER-heuristic, Koriat, 2008; Miele et al., 2011). Knowing the cues and mechanisms to affect judgments is useful because judgments have been found to affect control of cognitive processes and therefore, judgments can affect performance (e.g., Thiede et al., 2003). For example, when students judge disfluent texts as more difficult, they might invest more time and effort in the further learning process, which improves their performance. However, future research is required to test if students use judgments to control learning in a way that performance can be improved (see Pieger et al., 2016).

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. doi:10.1177/1088868309341564

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. doi:10.3758/s13421-013-0307-8

Bjork, E. L., & Storm, B. C. (2011). Retrieval experience as a modifier of future encoding: Another test effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1113–1124. doi:10.1037/a0023549

Bjork, R. A., & Yue, C. L. (2016). Commentary: Is disfluency desirable?. *Metacognition and Learning, 11,* 133–137. doi:10.1007/s11409-016-9156-8

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161–186. doi:10.1016/S0959-4752(96)00015-1

De Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*(4), 245–252. doi:10.1016/j.learninstruc.2012.01.003

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students` explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14. doi:10.1016/j.learninstruc.2012.06.001

Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema, 21*(1), 76–82.

Eitel, A., & Kühl, T. (2016). Effects of disfluency and text expectancy on learning with text. *Metacognition and Learning, 11*(1), 107–121. doi:10.1007/s11409-015-9145-3

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. doi:10.3758/BF03193146

Jönsson, F. U., & Lindström, B. R. (2010). Using a multidimensional scaling approach to investigate the underlying basis of ease of learning judgments. *Scandinavian Journal of Psychology*, *51*(2), 103–108. doi:10.1111/j.1467-9450.2009.00762.x

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294–303. doi:10.1037/0003-066X.49.4.294

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. doi:10.1037/0096-3445.126.4.349

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*(2), 416–428. doi:10.3758/MC.36.2.416

Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction*, *22*(4), 296–298. doi:10.1016/j.learninstruc.2012.01.002

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. doi:10.1037/0096-3445.135.1.36

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 22*(6), 787–794. doi:10.1177/0956797611407929

Lanska, M., Olds, J. M., & Westerman, D. L. (2014). Fluency effects in recognition memory: Are perceptual fluency and conceptual fluency interchangeable?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 1–11. doi:10.1037/a0034309

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning, 11*(1), 35–56. doi:10.1007/s11409-015-9147-1

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 663–679. doi:10.1037/0278-7393.10.4.663

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 609–616. doi:10.1037/0278-7393.16.4.609

McDaniel, M. A. (1984). The role of elaborative and schema processes in story memory. *Memory & Cognition*, *12*(1), 46–51. doi:10.3758/BF03196996

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 175–199). New York, NY: Psychology Press. doi:10.4324/9780203842539

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, *25*(6), 645–656. doi:10.1016/0749-596X(86)90041-0

McDaniel, M. A., Ryan, E. B., & Cunningham, C. J. (1989). Encoding difficulty and memory enhancement for young and older readers. *Psychology and Aging*, *4*(3), 333–338. doi:10.1037/0882-7974.4.3.333

Michalke, M. (2012). koRpus: A R package for text analysis (version 0.04-36). Available from http://reaktanz.de/?c=hacking&s=koRpus (Retrieved October 30, 2013).

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered?: It depends on your beliefs about intelligence. *Psychological Science*, *22*(3), 320–324. doi:10.1177/0956797610397954

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. doi:10.1037/a0019745

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. doi:10.3758/s13423-012-0343-6

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the 'labor-in-vain effect.'. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686. doi:10.1037/0278-7393.14.4.676

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York, NY: Academic Press. doi:10.1016/s0079-7421(08)60053-5

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. doi:10.1016/j.tics.2008.02.014

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. doi:10.1016/j.learninstruc.2016.01.012

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. doi:10.1002/acp.3326

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 69–80. doi:10.1037/0278-7393.28.1.69

Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction, 22*(4), 262–270. doi:10.1016/j.learninstruc.2011.10.007

Rudolph, U. (2009). *Motivationspsychologie kompakt [Motivational psychology compact]* (2nd ed.). Weinheim, Germany: Beltz PVU.

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,* 206–211. doi:10.3758/s13423-014-0658-6

Schwartz, B. L., & Efklides, A. (2012). Metamemory and memory efficiency: Implications for student learning. *Journal of Applied Research in Memory and Cognition, 1*(3), 145–151. doi:10.1016/j.jarmac.2012.06.002

Schwarz, N. (2010). Meaning in context: Metacognitive experiences. In B. Mesquita, L. F. Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 105–125). New York, NY: Guilford Press.

Son, L. K., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1890 to the present (and beyond). In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 333–351). New York, NY: Psychology Press.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204–221. doi:10.1037/0278-7393.26.1.204

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. doi:10.3758/s13421-013-0323-8

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. doi:10.1037/0022-0663.95.1.66

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1024–1037. doi:10.1037/0278-7393.25.4.1024

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: LEA.

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. doi:10.3758/s13421-012-0255-8

Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review, 2*(2), 173–201. doi:10.1007/BF01322178

Appendix A

Table A1. *Correlations Between Dependent Variables Across Texts and Groups (N = 63)*

Variable	1	2	3	4	5	6	7
1. EOL1							
2. Reading-time	.091						
3. EOL2	.642**	-.014					
4. JOL1	.431**	-.020	.597**				
5. Termination of study	-.132	-.061	.028	-.237			
6. JOL2	.365**	.007	.483**	.688**	-.190		
7. Performance	.091	-.114	.080	.165	.027	.346**	
8. RC judgments	.352**	-.066	.303*	.421**	-.192	.462**	.530**

* $p < .05$. ** $p < .01$.

G.2 Paper II: Judgments During the Learning Process – Does Disfluency Activate Analytic Monitoring not only for Disfluent but also for Succeeding Fluent Text?

Pieger, E., Mengelkamp, C., & Bannert, M.. Judgments during the learning process – Does disfluency activate analytic monitoring not only for disfluent but also for succeeding fluent text?

The author Elisabeth Pieger autonomously developed the research presented in this thesis. She is further the leading author of this paper. Elisabeth Pieger had the leading role in developing, managing, analyzing and presenting the experiments. This includes the conceptualizing, preparation, management, data-analysis, discussion, and presentation of the experiment. Moreover, she was leading during publishing the paper. The co-authors Maria Bannert and Christoph Mengelkamp advised the publication process. They were discussion-partners and provided feedback on the paper.

Abstract

Disfluency is supposed to initiate analytic processes and thus, to improve performance. However, disfluency often does not affect performance but, rather, monitoring. The aim of this study is to propose a specification of the effects of disfluency: We suppose that disfluency activates analytic monitoring but not necessarily analytic cognitive processing and improved performance. Once activated, analytic monitoring should remain for succeeding fluent material. To test our assumptions, students ($N = 32$) first learned with a disfluent text and then with a fluent text. Results show that fluency was not used as a cue for judgments after reading. This supports our assumption that disfluency activates analytic monitoring that remains for succeeding fluent text. When students ($N = 33$) first learned with a fluent and then with a disfluent text, fluency was used as a cue for judgments. Because fluency did not affect performance, fluency did not activate analytic cognitive processing.

Keywords: metacomprehension; disfluency; metacognitive monitoring; metacognitive control

Highlights:

- The type of contrast in fluency moderated the fluency effect on monitoring
- Fluency affected judgments when learning with a fluent and then with a disfluent text
- Analytic monitoring for disfluent text remained for succeeding fluent text
- Fluency led to analytic monitoring but not to analytic control and deep processing
- Disfluency-theory can be specified with respect to monitoring and control

1. Introduction

Monitoring one´s own learning is important in educational contexts. To monitor learning, students can make different types of judgments during the learning process (e.g., Nelson & Narens, 1990). For example, they judge how difficult it is to learn a text, or they judge how much knowledge they have acquired during learning. But, students often use cues for metacognitive judgments that are not valid for their performance. For example, students often predict high performance for texts that are easily processed, although the ease of processing (*fluency*) does not improve performance (e.g., Rawson & Dunlosky, 2002; Experiment 4). This is a problem because using valid cues for judgments is a prerequisite for effective learning and high performance (see theories of metacognition and self-regulated learning, e.g., Boekaerts, 1997; Dunlosky, Hertzog, Kennedy, & Thiede, 2005; Nelson & Narens, 1990; Winne & Hadwin, 1998; Zimmerman, 1990). Therefore, students` monitoring of the learning process should be analytic in order to make accurate judgments. *Disfluency* seems to be a way to activate analytic processes (Alter, Oppenheimer, Epley, & Eyre, 2007) that are relevant for metacognition. Analytic monitoring enables students to use more valid cues than fluency for their judgments. Moreover, once activated, more analytic processes should be found not only for disfluent but also for succeeding fluent material. The aim of this study is to investigate whether disfluency is a way to activate analytic monitoring, not only for a disfluent text itself, but also for a succeeding fluent text. In the following, we will argue for this assumption and derive our theoretical framework in more detail.

1.1 Effects of Disfluency on Metacognition: A Specification

Fluency is defined as ease of processing (Schwarz, 2010). Material that is easy to process is *fluent*, whereas material that is difficult to process is *disfluent*. There are different types of fluency, like *perceptual* and *conceptual* fluency (Schwarz, 2010). Conceptual fluency includes the ease of identifying the meaning of words and knowledge structures and can be manipulated, e.g., by letter deletion. Perceptual fluency describes the ease of identifying words and can be manipulated, e.g., by font. Conceptual fluency is especially relevant when learning with texts because it affects processing on a higher level than perceptual fluency; when learning with text, students do not only have to decipher words (*surface level*), but they also have to construct a meaningful representation of the texts (*textbase level* and *situation model*, see De Bruin & van Gog, 2012; Redford, Thiede, Wiley, & Griffin, 2012).

The disfluency-theory supposes that disfluency initiates slow and analytic processes (*System 2 processes*), whereas fluency initiates quick and surface processes (*System 1 processes*; see Alter et al., 2007). Thus, learning with disfluent material is expected to lead to better performance compared to learning with fluent material. However, previous research has shown that fluency does often not affect performance (e.g., Maki, Foley, Kajer, Thompson, & Willert, 1990, Experiment 2; McDaniel, Einstein, Dunay, & Cobb, 1986, Experiment 1; Rawson & Dunlosky, 2002, Experiment 4, for conceptual fluency; see also Eitel & Kühl, 2016; Eitel, Kühl, Scheiter, & Gerjets, 2014, Experiment 2–4; Lehmann, Goussios, & Seufert, 2016; Rummer, Schweppe, & Schwede, 2016; Strukelj, Scheiter, Nyström, & Holmqvist, 2016, for perceptual fluency).

Because the empirical evidence does often not support disfluency-theory with respect to performance, it is reasonable to more precisely specify the processes that are activated by disfluency. In accordance with Alter et al. (2007; see also Diemand-Yauman, Oppenheimer, & Vaughan, 2011), we assume that disfluency leads to more analytic metacognitive processes. However, we further state that these metacognitive processes mainly refer to metacognitive monitoring and not necessarily to metacognitive control. If disfluency would activate analytic control, performance should be better for disfluent than for fluent material: Theories of self-regulated learning and metacognition state that effective control is required to alter cognitive processes and therefore to improve performance (Boekaerts, 1997; Nelson & Narens, 1990; Winne & Hadwin, 1998; Zimmerman, 1990). However, often no performance differences between fluent and disfluent texts are found (see above). Hence, disfluency does not necessarily seem to activate analytic control but simply slows down processing of the learning material (see Figure 1). Therefore, longer reading-times for disfluent rather than for fluent texts (e.g., McDaniel, 1984; McDaniel et al., 1986, for conceptual fluency; see also Eitel & Kühl, 2016; Miele & Molden, 2010, Experiment 3; Pieger, Mengelkamp, & Bannert, 2016; Sanchez & Jaeger, 2015, Experiment 1 and 2, for perceptual fluency) might be due to less automatic processing but not necessarily due to deeper processing. Inversely, disfluency seems to be a way to activate analytic monitoring.

Our assumption that disfluency activates analytic monitoring is supported by the studies by Alter et al. (2007). For example in their Experiment 2, they found that students processed arguments of persuasion in a more analytic way when the masthead (not the arguments per se) was presented in a disfluent way than in a fluent way. Only the masthead and not the arguments were disfluent, therefore, the processing of the arguments was not directly affected by disfluency. Nevertheless, performance improved for these arguments. Thus, one explanation for this finding is that disfluency has activated analytic monitoring. This analytic monitoring even remained during reading the fluent arguments and enabled analytic processing of the arguments. Therefore, we conclude that disfluency activates analytic monitoring and that this analytic monitoring remains for the subsequent learning material, even if this material is fluent (see Figure 1).

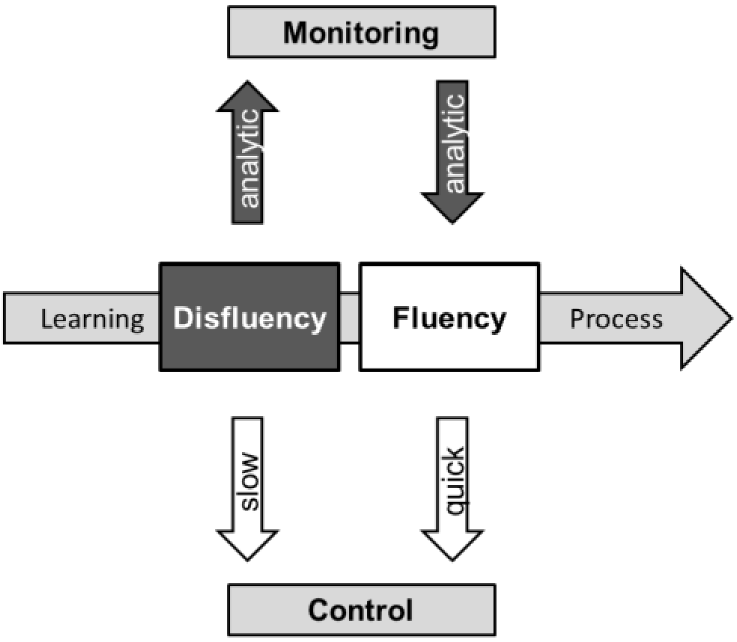

Figure 1. Activation of analytic monitoring by disfluency.

As a consequence of analytic monitoring for disfluent and succeeding fluent material, students should realize that the fluency of a text is not a valid cue for performance, and they, thus, should not predict higher performance for the fluent text than for the disfluent text. Presenting disfluent and, afterwards, fluent material (we will call this *contrast disfluent-fluent*) should reduce the *fluency effect* (difference between fluent and disfluent material) on monitoring. Inversely, when fluent material is presented before disfluent material (we will call this *contrast fluent-disfluent*), analytic monitoring is only activated for the disfluent but not for the previous fluent material. In this case, students should use fluency as a cue for their judgments (e.g., performance predictions), although, fluency does not affect performance.

Summing up, the aim of this study is to test whether disfluency activates analytic monitoring, and if this analytic monitoring remains for succeeding fluent material. We assume that students do not use fluency as a cue for their judgments when they learn with a disfluent text first and then with a fluent text because analytic monitoring is activated for both texts. Inversely, when students learn with a fluent text first and then with a disfluent text, fluency should be used as a cue for judgments. Such fluency effects on monitoring have often been found in previous research, which is described next.

1.2 Fluency Effects on Monitoring in Previous Research

In previous research, students were often asked to judge their performance when learning with word-pairs, and usually perceptual disfluency was manipulated. There is evidence that students make lower judgments for disfluent rather than for fluent word-pairs or word-lists (e.g., Kornell, Rhodes, Castel, & Tauber, 2011; Mueller, Tauber, & Dunlosky, 2013, Experiment 2; Yue, Castel, & Bjork, 2013, Experiment 1a, 2a, 2b, and 3). However, to investigate the role of fluency in educational contexts, meaningful material like texts should be used. Moreover, when learning with texts, conceptual fluency might be more relevant than perceptual fluency because it affects higher levels of processing than perceptual fluency.

When learning with meaningful text material, students should monitor different levels of text-processing during the entire learning process. Before learning, students can judge how difficult it will be to learn a text (*ease of learning judgment*, EOL judgment, see Nelson & Narens, 1990). During the early learning stages, text-processing takes place on a surface level. For example, students might have a short look at the text or read a text once. Thus, EOL judgments should refer to surface levels of text-processing. There is some evidence that students judge disfluent texts as more difficult than fluent texts (e.g., Maki et al., 1990, Experiment 1; Rawson & Dunlosky, 2002, Experiment 4, for conceptual fluency; see also Miele & Molden, 2010, Experiment 3; Pieger et al., 2016, for perceptual fluency). However, EOL judgments have often been neglected in previous research (Dunlosky & Metcalfe, 2009; Jönsson & Lindström, 2010; Pieger et al., 2016).

Research often focused on the effects of perceptual disfluency on *judgments of learning* (JOLs) when learning with word-pairs. JOLs can be made during and after studying by predicting one`s own performance on a knowledge test (Dunlosky & Metcalfe, 2009). During studying, higher levels of text-processing are in the focus (i.e., textbase level and situation model) and thus, JOLs are related to higher levels of text-processing than are EOL judgments. When learning with texts, there is also some evidence that fluency is used as a cue for JOLs (e.g., Rawson & Dunlosky, 2002, Experiment 4, for conceptual fluency; see Pieger et al., 2016, for perceptual fluency). Although fluency has been found to affect EOL judgments and JOLs, EOL judgments and JOLs represent different aspects of monitoring (Nelson & Narens, 1990), because they are only weakly correlated (Leonesio & Nelson, 1990). Moreover, Maki et al. (1990, Experiment 1) found that students used fluency as a cue when judging the difficulty of comprehending the text (which is similar to an EOL judgment). However, students that were asked about performance predictions (which are similar to JOLs) used not only fluency but also retrieval as a cue for these judgments (see also Rawson & Dunlosky, 2002, Experiment 4). Thus, when investigating the effect of fluency on monitoring, different types of judgments should be considered. Different types of judgments do not only represent different aspects of monitoring, but they can further be affected by different cues.

Besides EOL judgments and JOLs, which both are made before the knowledge test (*predictive judgments*), *retrospective confidence judgment*s (RC judgments) are also relevant in educational contexts (also called *postdictions*, see Pieschl, 2009). These judgments ask students for their confidence in the correctness of a retrieved answer during a performance test. According to the findings by Dinsmore and Parkinson (2013), text characteristics are the most frequently used cues for RC judgments (besides somewhat less frequently used cues like, e.g., item characteristics and prior knowledge). Because fluency is an obvious text characteristic, students should use fluency as a cue for RC judgments and thus fluency effects on RC judgments should be expected. However, when disfluency activates analytic monitoring that remains for a succeeding fluent text, no fluency effect on judgments should be found.

In sum, there is evidence that students often use fluency as a cue for their judgments. When learning with texts, students should make different types of judgments during the learning process (i.e., EOLs, JOLs, and RC judgments) that refer to different levels of text-processing. These different types of judgments can be affected by fluency and represent different aspects of monitoring.

1.3 Research Questions and Hypotheses

The aim of this study is to investigate whether disfluency activates analytic monitoring. We suppose that the sequence of presenting fluent and disfluent material (we will call this *type of contrast*) moderates fluency effects on judgments.

When students learn with a disfluent text first and then with a fluent text (contrast disfluent-fluent), disfluency should activate analytic monitoring, which should then remain for the succeeding fluent text. Because of analytic monitoring for the disfluent and fluent texts, fluency should not be used as a cue for judgments. Thus, no fluency effects on EOL judgments, JOLs, and RC judgments are expected.

Inversely, when students learn with a fluent text first and afterwards with a disfluent text (contrast fluent-disfluent), monitoring of the fluent text is expected to be surface level instead of analytic. Thus, students should base their judgments on the experience of fluency: They should judge disfluent, compared to fluent texts, as more difficult (EOL judgments), they should predict lower performance (JOLs) for disfluent than for fluent texts, and they should be less confident in the correctness of a retrieved answer for disfluent than for fluent texts (RC judgment).

Independent from the type of contrast in fluency, disfluency should lead to slower processing: Disfluency compared with fluency should lead to longer reading-times. Because disfluency should not lead to more analytic control of cognitive processing of the learning material, we expect no performance differences between fluent and disfluent texts. To investigate if disfluency leads to analytic processing, we used a manipulation of conceptual disfluency because this fluency-manipulation affects processing on a higher level than perceptual fluency and is, therefore, more relevant when learning with texts.

2. Method

2.1 Participants and Experimental Design

The experiment was conducted at a university in Germany. In total $N = 65$ students participated in the study (age: $M = 20.09$, $SD = 1.60$ years, 78.46% female); $N = 49$ (75.38%) students studied media communication and $N = 16$ (24.62%) students studied human-computer-interaction (semester of studies: $M = 2.68$, $SD = 1.58$). The acquired sample size fulfills the required sample size of $N = 67$ students, computed with G*-Power (Faul, Erdfelder, Lang, & Buchner, 2007) by setting the Type I error to .05, the power to .80, and assuming an effect size of $f = 0.35$ because we expected higher effects for conceptual than for perceptual fluency (see also Rawson & Dunlosky, 2002; Maki et al., 1990). Fluency was varied within persons whereas type of contrast was varied between persons. Students were randomly assigned to one of two groups that differed in the type of contrast in fluency. Students in *contrast group disfluent-fluent* ($N = 32$) read a disfluent text and afterwards a fluent text, whereas students in *contrast group fluent-disfluent* ($N = 33$) read a fluent text and afterwards a disfluent text. Monitoring (EOL judgments, JOLs, and RC judgments), control (reading-time and termination of study), and performance were captured as dependent variables.

2.2 Material

Two expository texts about motivational psychology (Rudolph, 2009) were used as learning material. These texts had been used in a previous study and were classified as comparable in length and difficulty: Text A was about the rubicon model and consisted of 936 words (Flesch-Kincaid grade-level score = 18.94); text B explained causal dimensions in attribution theory and consisted of 929 words (Flesch-Kincaid grade-level score = 20.92). Scores were computed by a tool developed by Michalke (2012).

For disfluent texts, letters were deleted using an algorithm similar to the one used by McDaniel (1984). The algorithm was adapted for German texts in order to create disfluent texts that were comparable to the experiments by Maki et al. (1990) and by Rawson and Dunlosky (2002, Experiment 4). We first deleted all vowels in every noun, adjective and verb, except for initial letters, and replaced them by one underscore for each deleted letter. To guarantee that each word was recognizable as one word, the words were spaced by five blank spaces (see Figure 2 for an example of a disfluent text). Texts were piloted to ensure that the words could still be correctly read. Therefore, students in the pilot study ($N = 10$) had to fill in each deleted letter of the two texts. If less than 50% of these students were able to fill in the letters, all letters of this word were filled in for the experiment. This was the case for four technical terms.

Dies ist ein T_xt mit
f_hl_nd_n B_chst_b_n.
Alle V_k_l_ in N_m_n,
Adj_kt_v_n und V_rb_n f_hl_n.

Figure 2. Example of a deleted letter text.

2.3 Instruments

Reading-time was used as a manipulation-check because slow processing of disfluent texts directly affects reading-time when reading a text once. Judgments were captured by asking questions that students had to answer on a continuous visual analogue scale from 0 to 100 by keyboarding integer numbers. For EOL judgments, we asked, “How easy or difficult is it to learn the text?” using a scale labeled from 0 = *difficult* (50 = *middle*) to 100 = *easy*. For JOLs, we asked “What percentage of questions about the text will you answer correctly?” using a scale labeled from 0 = *none* (50 = *half*) to 100 = *all*. For RC judgments, we asked “How confident are you that your answer is correct?” using a scale labeled from 0 = *unconfident* (50 = *middle*) to 100 = *confident*.

Performance was assessed by a knowledge test that consisted of 23 questions on text A, and of 24 questions on text B. Each question consisted of 6 statements. An example for a question with six statements on text A is presented in Figure 3. Each statement was sequentially presented on the screen, and students had to decide if it was true or false. For each statement that students correctly identified as true or false, they were awarded one point. Performance was computed as the mean of all statements. As the chance to guess the correct answer was 50%, this score was corrected for guessing using the algorithm $200 \cdot x - 100$. This algorithm transforms the value of guessing to zero ($200 \cdot 0.50 - 100 = 0$), resulting in a performance scale from 0 – 100%. Reliability of the test was Cronbach´s $\alpha = .81$ ($M = 51.07\%$, $SD = 10.38\%$).

Question:

Which statements are correct, regarding to the transition from the pre-actional to the actional phase?

Statements:

a) I have a concrete plan and ignore all stimuli to be able to focus on my aim.
b) I implement my realization plans in actions.
c) I have a concrete plan but still react flexible on circumstances.
d) I envision the possible consequences of my intention.
e) I focus my attention on the execution and realize my plan.
f) I cannot realize my plans, therefore I think about new alternatives.

Figure 3. Example of a question with six statements of the knowledge test.

2.4 Procedure

E-Prime-Software (E-Prime Professional 2.0) was used to present the materials and to collect the data. The procedure is shown in Figure 4. First, an extract of the first text was presented for 2 seconds on the screen, and afterwards, students made EOL1. Then, students were instructed to read the entire text only once on the screen (without rereading or skipping), and afterwards, they made EOL2 and JOL1 on the next screens. Next, the text was shown again, and this time, students were allowed to reread, to take notes, and to skip within the text for a maximum of 15 minutes. However, students were allowed to terminate their study before the time expired. Immediately afterwards, students made JOL2 on the next screen. This same procedure was used with the second text. Finally, the knowledge test with questions about each text was presented. The order of the texts was randomized. If text A was the first, text questions on text A were presented first, and if text B was the first, text questions on text B were presented first. Each statement of a question was separately presented on the screen, and students had to decide if this statement was true or false. Students received one point for each correct decision. Immediately after each statement, students made a RC judgment on the next screen. The experiment took approximately 2 hours.

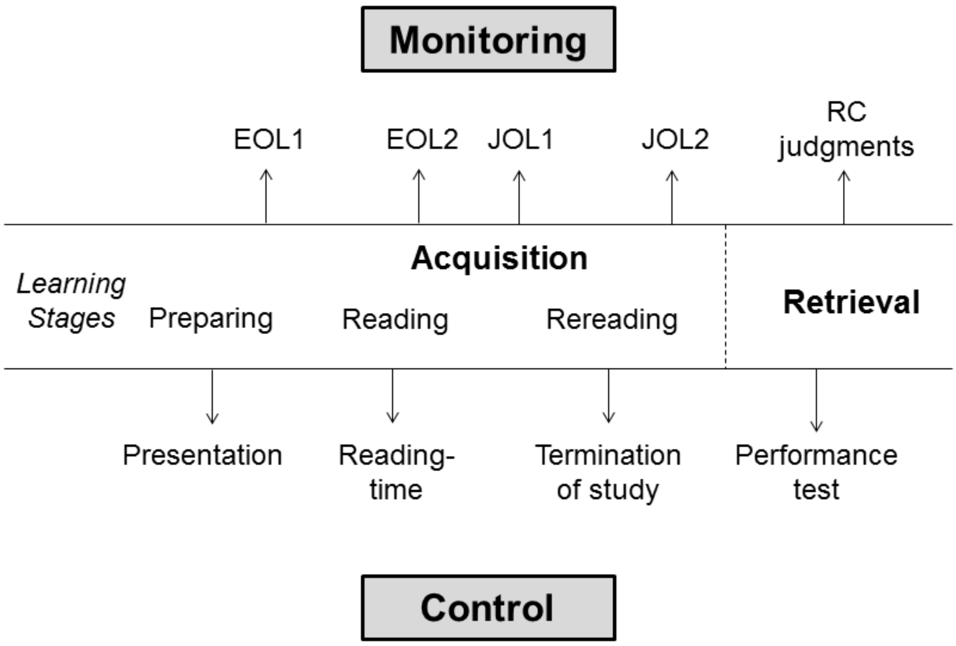

Figure 4. Procedure of the study. The procedure was the same for each of the two texts; performances were captured after acquisition of both texts.

3. Results

Descriptive statistics of dependent variables (judgments, reading-time, termination of study, and performance) in contrast group disfluent-fluent and in contrast group fluent-disfluent are presented in Table 1. We computed a mixed MANOVA to test if there was a significant multivariate effect of the within-subject factor, fluency (fluent vs. disfluent), and the between-subject factor, type of contrast (disfluent-fluent vs. fluent-disfluent), on the dependent variables that are listed in Table 1 (correlations between dependent variables can be found in Appendix A, Table A1). Results showed a significant main effect of fluency, $V = .671$, $F(8, 52) = 13.23$, $p < .001$, $\eta_p^2 = 0.67$, and a significant interaction between fluency and type of contrast, $V = .383$, $F(8, 52) = 4.04$, $p = .001$, $\eta_p^2 = 0.38$. The main effect of type of contrast was not significant, $V = .147$, $F(8, 52) = 1.12$, $p = .363$, $\eta_p^2 = 0.15$. Next, we report on which dependent variables significant effects were found.

Table 1. *Descriptive Statistics of Dependent Variables in Contrast Group Disfluent-Fluent and in Contrast Group Fluent-Disfluent*

	Contrast group disfluent-fluent (N = 32)				Contrast group fluent-disfluent (N = 33)			
	disfluent		fluent		fluent		disfluent	
Variable	M	SD	M	SD	M	SD	M	SD
EOL1[a]	36.50	23.64	44.59	21.29	47.03	21.84	35.91	31.19
Reading-time[b]	7.89	2.31	5.84	1.68	5.41	1.40	7.94	2.13
EOL2[a]	55.53	20.56	42.19	18.09	49.48	14.57	38.91	27.16
JOL1[c]	50.78	16.07	52.34	16.16	51.21	13.46	39.24	19.61
Termination of study[b]	10.38	3.29	8.89	3.89	10.79	3.01	10.04	3.15
JOL2[c]	64.35	17.55	61.87	18.02	67.00	14.57	59.39	18.02
Performance[c]	50.42	14.08	53.09	12.48	54.68	11.23	48.82	14.55
RC judgments[d]	74.70	14.67	75.47	14.95	77.29	13.06	71.27	15.14

Note. Judgment and performance scales ranged from 0% to 100%.

[a]0 = *difficult*, 100 = *easy*, [b]in minutes, [c]0 = *none*, 100 = *all*, [d]0 = *unconfident*, 100 = *confident*.

To test whether our fluency-manipulation was effective, reading-time was analyzed. We found a significant main effect of fluency on reading-time in the reading phase, $F(1, 59) = 79.96$, $p < .001$, $\eta_p^2 = 0.58$. The fluency-manipulation worked equally effectively for both types of contrast; the interaction between fluency and type of contrast was not significant, $F(1, 59) = 0.45$, $p = .506$, $\eta_p^2 = 0.01$.

Next, the judgments are analyzed in the order in which they were made throughout the experiment. We found a significant main effect of fluency on EOL1, $F(1, 59) = 10.57$, $p = .002$, $\eta_p^2 = 0.15$, but no significant interaction between fluency and type of contrast, $F(1, 59) = 0.33$, $p = .567$, $\eta_p^2 = 0.01$. The disfluent text was judged as more difficult than the fluent text in both groups and, thus, for both types of contrast.

A significant interaction between fluency and type of contrast was found for EOL2, $F(1, 59) = 14.55$, $p < .001$, $\eta_p^2 = 0.20$, indicating that the fluency effect was moderated by the type of contrast. Students in contrast group disfluent-fluent judged the fluent text compared with the disfluent text as more difficult, $F(1, 31) = 9.63$, $p = .004$, $\eta_p^2 = 0.24$, whereas students in contrast group fluent-disfluent judged the disfluent text as more difficult than the fluent text, $F(1, 32) = 5.24$, $p = .029$, $\eta_p^2 = 0.14$.

For JOL1, the fluency effect was moderated by the type of contrast because there was a significant interaction between fluency and type of contrast, $F(1, 59) = 5.63$, $p = .021$, $\eta_p^2 = 0.09$. As expected, students in contrast group disfluent-fluent predicted equal performance for the fluent and the disfluent text, $F(1, 31) = 0.29$, $p = .591$, $\eta_p^2 = 0.01$, and students in contrast group fluent-disfluent predicted lower performance for the disfluent than for the fluent text, JOL1, $F(1, 32) = 11.13$, $p = .002$, $\eta_p^2 = 0.26$.

For JOL2, after rereading the text, we found a similar pattern: The interaction between fluency and type of contrast was significant, $F(1, 59) = 5.46$, $p = .023$, $\eta_p^2 = 0.08$. Students in contrast group disfluent-fluent predicted equal performance for the fluent and the disfluent text, $F(1, 30) = 0.62$, $p = .436$, $\eta_p^2 = 0.02$, whereas the students in contrast group fluent-disfluent predicted lower performance for the disfluent compared to the fluent text, JOL2, $F(1, 32) = 4.56$, $p = .040$, $\eta_p^2 = 0.12$.

Moreover, we found a significant interaction between fluency and type of contrast on RC judgments, $F(1, 59) = 6.63$, $p = .013$, $\eta_p^2 = 0.10$. In contrast group disfluent-fluent, students made equal RC judgments for the fluent and the disfluent text, $F(1, 31) = 0.23$, $p = .636$, $\eta_p^2 = 0.01$. In contrast group fluent-disfluent, students were less confident about their performance for the disfluent than for the fluent text, $F(1, 32) = 10.11$, $p = .003$, $\eta_p^2 = 0.24$.

For termination of study in the rereading phase, a significant interaction between fluency and type of contrast indicates different fluency effects depending on the type of contrast, $F(1, 59) = 7.73$, $p = .007$, $\eta_p^2 = 0.12$. Students in contrast group disfluent-fluent studied disfluent texts significantly longer than fluent texts, $F(1, 31) = 5.67$, $p = .024$, $\eta_p^2 = 0.15$. Students in contrast group fluent-disfluent, however, did not terminate their study later for the disfluent than for the fluent text, $F(1, 32) = 2.10$, $p = .157$, $\eta_p^2 = 0.06$.

Finally, for performance, neither a significant main effect of fluency, $F(1, 59) = 3.57$, $p = .064$, $\eta_p^2 = 0.06$, nor a significant interaction between fluency and type of contrast, $F(1, 59) = 0.67$, $p = .416$, $\eta_p^2 = 0.01$, was found. Thus, for both types of contrast, performance was not significantly different between fluent and disfluent texts.

To sum up, we found different fluency effects for the two types of contrast for all judgments that were made after reading, rereading, and during the test (EOL2, JOL1, JOL2, and RC judgments). Whereas no fluency effect on these judgments was found in contrast group disfluent-fluent, a fluency effect on these judgments was found in contrast group fluent-disfluent. Additionally, the fluency effect on EOL2 was inverted in contrast group disfluent-fluent. On EOL1, a fluency effect was found for both types of contrast.

4. Discussion

The aim of this study was to investigate whether disfluency activates analytic monitoring, not only for disfluent, but also for the succeeding fluent text. We expected that the type of contrast in fluency moderates the fluency effect on judgments: When a disfluent text is presented before a fluent text (contrast disfluent-fluent), disfluency should activate analytic monitoring, which remains for the succeeding fluent text. Therefore, fluency should not be used as a cue for judgments. Inversely, when students first learn with a fluent and then with a disfluent text (contrast fluent-disfluent), monitoring of the fluent text is expected to be surface, and thus, fluency should be a cue for judgments.

Results widely support this assumption: When a fluent text was presented before a disfluent text (contrast fluent-disfluent), fluency is used as a cue for all types of judgments during the entire learning process. These results are consistent with previous findings of fluency effects on judgments (e.g., Rawson & Dunlosky, 2002, Experiment 4). Moreover, these results go beyond previous findings because we found this effect on different types of judgments during the learning process, even on RC judgments during the test. Thus, fluency affects different levels of text-processing.

Inversely, when students first learned with a disfluent and then with a fluent text (contrast disfluent-fluent), fluency was no longer used as a cue for JOL1, JOL2, and RC judgments. This was the case, although students in contrast group disfluent-fluent read and reread the disfluent text significantly longer than the fluent text. Therefore, students experienced fluency, but this experience was not used as a cue for JOLs and RC judgments due to analytic monitoring. This finding supports our hypothesis that disfluency activates analytic monitoring, and it remains activated for the succeeding fluent text.

Our assumption that disfluency activates analytic monitoring is supported by our findings, and results in both groups further support our assumption that disfluency does not necessarily lead to more analytic control or deeper processing of the learning material: Although students read the disfluent text significantly longer than the fluent text, they did not show better performance on the disfluent text. Therefore, processing of the text was slower, but not more analytic for the disfluent than for the fluent text. These findings on monitoring and control are consistent to our specification of processes that are activated by disfluency: Disfluency-theory postulates that disfluency leads to slower and to more analytic processes. We suggested that (a) these processes seem to be metacognitive (see also Alter et al., 2007) and that (b) these processes refer to analytic metacognitive monitoring but not necessarily to analytic metacognitive control or deep processing of the text.

However, based on our results, we conclude that the effects of disfluency should not only be specified with respect to monitoring and control. Moreover, the effects of disfluency on monitoring seem to depend on the type of metacognitive judgment. We found that fluency was not used as a cue for JOLs or for RC judgments in contrast group disfluent-fluent but fluency was used as a cue only for EOL1 judgments. This finding is consistent with the findings by Maki et al. (1990, Experiment 1): Fluency effects on judgments were found when students judged how difficult a text is to understand, but no fluency effects on judgments were found when students predicted performance. Whereas JOLs and RC judgments ask students about their performance, EOL judgments ask students about the difficulty of the texts. Disfluent texts are indeed more difficult to read, as can be seen in the prolonged reading-times compared to fluent texts: Mentally filling in deleted letters requires more time compared to reading an intact text. This affects the surface level of text-processing. Therefore, students judged the disfluent text as more difficult to learn than the fluent text (EOL1) even though the content was, in fact, not more difficult to learn, as can be inferred from the equal performance between fluent and disfluent texts. Inversely, because disfluency did not affect performance, students made equal JOLs and RC judgments for disfluent and fluent texts. Therefore, students did not use fluency as a cue for JOLs and RC judgments because of analytic monitoring.

Moreover, students in contrast group disfluent-fluent judged fluent texts as more difficult to learn than disfluent texts after reading the texts once (i.e., EOL2). This is somewhat surprising given the fact that they judged the disfluent text as being more difficult than the fluent text after a presentation for 2 seconds on the screen (EOL1). Therefore, not only the type of judgment but also the learning stage seems to play a role in fluency effects on EOL judgments. When reading a text once, students may experience disfluency, and they have to process the text for a time longer than 2 seconds but apparently, they did not use disfluency as a cue. However, when reading the fluent text, there might also be some text-passages for which processing is slow, e.g., due to text-difficulty. Whereas students attribute the experience of slower processing of the disfluent text to its disfluency, they might attribute the experience of slower processing of the fluent text-passages to its difficulty. This could be one reason why students in contrast group disfluent-fluent judged the fluent text as being more difficult than the disfluent text. Because EOL2 explicitly asked for the difficulty of the text, this effect was found only for EOL2 but not for JOL1 (both judgments being made after the first reading phase), which asks for performance predictions. However, further research is required to investigate this post-hoc explanation for the inverted fluency effect on EOL2.

Summing up, disfluency is a way to activate analytic monitoring. Moreover, this analytic monitoring is not only found for disfluent but also for succeeding fluent material: More analytic monitoring is activated by the disfluent text and remains for the succeeding fluent text when making JOLs and RC judgments. Inversely, when presenting first a fluent and, afterwards, a disfluent text, monitoring of the fluent text is not analytic. In this case, students base their judgments on the experience of quick processing of the fluent texts and of slow processing of the disfluent text. Thus, the type of contrast in fluency affects the fluency effect on JOLs and RC judgments.

5. Conclusion

In this study, we investigated if disfluency activates analytic monitoring not only for disfluent but also for a succeeding fluent text. Another goal was to give some empirical evidence for our specification of the effects of disfluency due to inconsistent findings in previous research. Disfluency-theory supposes that disfluency leads to slow, analytic processes whereas fluency leads to quick, surface processes (Alter et al., 2007). However, these processes need to be specified as they can refer to metacognitive monitoring, metacognitive control, or processing of the learning material. We supposed that disfluency activates analytic monitoring but that it does not necessarily activate analytic control or deeper processing of the learning material, and these assumptions are in line with our results. Moreover, based on our results and on results from previous research, we suppose that further specifications are required: Fluency effects on judgments are affected by the interplay between the type of contrast in fluency, the type of judgment students are asked for, and the stage of the learning process. To derive guidelines for educational contexts, it is important to conduct systematic research that investigates this interplay. These guidelines are needed because disfluency is often used in textbooks (e.g., italic, bold, font types, fill-in-the-blank text) without understanding the effects of disfluency and its interactions with the learning stage of students' monitoring. Based on our findings, we conclude that disfluency is more than just a cue for judgments, it is a way to activate analytic monitoring.

Acknowledgement

We would like to thank our student research assistants Jessica Hartleb and Tamara Schmidtner for their help in completing this research, as well as Emily Lubaway for the proofreading.

References

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. doi:10.1037/0096-3445.136.4.569

De Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*(4), 245–252. doi:10.1016/j.learninstruc.2012.01.003

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161–186. doi:10.1016/S0959-4752(96)00015-1

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition, 118*(1), 114–118. doi:10.1016/j.cognition.2010.09.012

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students` explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14. doi:10.1016/j.learninstruc.2012.06.001

Dunlosky, J., Hertzog, C., Kennedy, M. R. T., & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *Cognitive Technology, 9*(1), 4-11.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Los Angeles, CA: SAGE.

Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema, 21*(1), 76–82.

Eitel, A., & Kühl, T. (2016). Effects of disfluency and text expectancy on learning with text. *Metacognition and Learning, 11*(1), 107–121. doi: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. doi:10.1002/acp.3004

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. doi:10.3758/BF03193146

Jönsson, F. U., & Lindström, B. R. (2010). Using a multidimensional scaling approach to investigate the underlying basis of ease of learning judgments. *Scandinavian Journal of Psychology*, *51*(2), 103–108. doi:10.1111/j.1467-9450.2009.00762.x

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 22*(6), 787–794. doi:10.1177/0956797611407929

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning, 11*(1), 89–105. doi:10.1007/s11409-015-9149-z

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 464–470. doi:10.1037/0278-7393.16.3.464

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 609–616. doi:10.1037/0278-7393.16.4.609

McDaniel, M. A. (1984). The role of elaborative and schema processes in story memory. *Memory & Cognition*, *12*(1), 46–51. doi:10.3758/BF03196996

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, *25*(6), 645–656. doi:10.1016/0749-596X(86)90041-0

Michalke, M. (2012). koRpus: A R package for text analysis (version 0.04-36). Available from http://reaktanz.de/?c=hacking&s=koRpus (Retrieved October 30, 2013).

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. doi:10.1037/a0019745

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. doi:10.3758/s13423-012-0343-6

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York, NY: Academic Press. doi:10.1016/S0079-7421(08)60053-5

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. doi:10.1016/j.learninstruc.2016.01.012

Pieschl, S. (2009). Metacognitive calibration – An extended conceptualization and potential applications. *Metacognition and Learning*, *4*(1), 3–31. doi:10.1007/s11409-008-9030-4

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 69–80. doi:10.1037/0278-7393.28.1.69

Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction, 22*(4), 262–270. doi:10.1016/j.learninstruc.2011.10.007

Rudolph, U. (2009). *Motivationspsychologie kompakt [Motivational psychology compact]* (2nd ed.). Weinheim, Germany: Beltz PVU.

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacognition and Learning, 11*(1), 57–70. doi:10.1007/s11409-015-9151-5

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,* 206–211. doi:10.3758/s13423-014-0658-6

Schwarz, N. (2010). Meaning in context: Metacognitive experiences. In B. Mesquita, L. F. Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 105–125). New York, NY: Guilford.

Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning, 11*(1), 71–88. doi:10.1007/s11409-015-9146-2

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: LEA.

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. doi:10.3758/s13421-012-0255-8

Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review, 2*(2), 173–201. doi:10.1007/BF01322178

Appendix A

Table A1. *Correlations Between Depend Variables Across Texts and Groups (N = 65)*

Variable	1	2	3	4	5	6	7
1. EOL1							
2. Reading-time	.054						
3. EOL2	.643**	.098					
4. JOL1	.351**	.071	.272*				
5. Termination of study	-.058	.018	-.061	-.083			
6. JOL2	.372**	-.007	.297*	.533**	-.022		
7. Performance	.101	.007	-.060	.306*	.175	.335**	
8. RC judgments	.091	.153	.088	.261*	.065	.288*	.544**

* $p < .05$. ** $p < .01$.

G.3 Paper III: Fostering Analytic Metacognitive Processes and Reducing Overconfidence by Disfluency: The Role of Contrast Effects

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. doi:10.1002/acp.3326

The author Elisabeth Pieger autonomously developed the research presented in this thesis. She is further the leading author of this paper. Elisabeth Pieger had the leading role in developing, managing, analyzing and presenting the experiments. This includes the conceptualizing, preparation, management, data-analysis, discussion, and presentation of the experiment. Moreover, she was leading during publishing the paper. The co-authors Maria Bannert and Christoph Mengelkamp advised the publication process. They were discussion-partners and provided feedback on the paper.

G.4 Paper IV: Metacognitive Judgments and Disfluency – Does Disfluency Lead to more Accurate Judgments, Better Control, and Better Performance?

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. doi: 10.1016/j.learninstruc.2016.01.012

The author Elisabeth Pieger autonomously developed the research presented in this thesis. She is further the leading author of this paper. Elisabeth Pieger had the leading role in developing, managing, analyzing and presenting the experiments. This includes the conceptualizing, preparation, management, data-analysis, discussion, and presentation of the experiment. Moreover, she was leading during publishing the paper. The co-authors Maria Bannert and Christoph Mengelkamp advised the publication process. They were discussion-partners and provided feedback on the paper.