CrossMark

Working memory capacity and disfluency effect: an aptitude-treatment-interaction study

Janina Lehmann[1] · Christina Goussios[1] · Tina Seufert[1]

Abstract According to Cognitive Load Theory, learning material should be designed in a way to decrease unnecessary demands on working memory (WM). However, recent research has shown that additional demands on WM caused by less legible texts lead to better learning outcomes. This so-called disfluency effect can be assumed as a metacognitive regulation process during which learners assign their cognitive resources depending on the perceived difficulty of a cognitive task. Increasing the perceived difficulty associated with a cognitive task stimulates deeper processing and a more analytic and elaborative reasoning. Yet there are studies which could not replicate the disfluency effect indicating that disfluency might be beneficial only for learners with particular learner characteristics. Additional demands on working memory caused by disfluent texts are possibly just usable by learners with a high working memory capacity. Therefore the present study investigated the aptitude-treatment-interaction between working memory capacity and disfluency. Learning outcomes were measured by means of a retention, a comprehension, and a transfer test. Moreover, the three types of cognitive load (intrinsic, extraneous, and germane) were assessed. The results revealed significant aptitude-treatment-interaction effects with respect to retention and comprehension. Working memory capacity had a significant influence only in the disfluency condition: The higher the working memory capacity, the better the retention and comprehension performance in the disfluency condition. No effects were found with respect to transfer or cognitive load. Thus, the role of metacognitive regulation and its possible effects on cognitive load need further investigation.

✉ Tina Seufert
  tina.seufert@uni-ulm.de

  Janina Lehmann
  janina.lehmann@uni-ulm.de

  Christina Goussios
  christina.goussios@uni-ulm.de

[1] Department of Learning and Instruction, Ulm University, Institute of Psychology and Education, Albert-Einstein-Allee 47, 89081 Ulm, Germany

🖄 Springer

Introduction

In different learning contexts such as school, higher education, vocational training or apprenticeships, teachers are concerned with the question of how to design instructional material in order to lead their students to the maximum of success in learning. Thereby, teachers as well as students may be guided by the assumption that learning material which speeds up and facilitates acquisition during instruction enhances long-term learning (Bjork 1994; Sweller et al. 1998, 2011). In contrast, there is empirical evidence for a better learning performance with disfluent learning material, which makes reading harder (Diemand-Yauman et al. 2011; Eitel et al. 2014; French et al. 2013; Sungkhasettee et al. 2011). It raises the question of whether the decision for one type of learning material depends on some special learner characteristics like prior knowledge or working memory capacity.

Disfluency effect

Recent research (Diemand-Yauman et al. 2011; Eitel et al. 2014 (Experiment 1); French et al. 2013; Sungkhasettee et al. 2011) has shown that less legible texts can lead to better learning outcomes. This so-called disfluency effect manipulates the perceived effort of learning by increasing the perceptual difficulty. Disfluent learning material is therefore a "desirable difficulty", because it doesn't affect the objective effort simultaneously, but manipulates the subjective effort (Bjork 2013). These difficulties cause an additional cognitive burden, in case of disfluency by using a harder-to-read font. Therefore learners have to engage themselves more during the learning process leading to a deeper processing and better learning outcomes.

According to disfluency theory, the disfluency effect can be assumed as a metacognitive regulation process during which learners assign their cognitive resources depending on the perceived difficulty of a cognitive task (Alter et al. 2007). Based on the assumptions of Tverski and Kahneman (1974; James, 1890/ 1950), there are two distinct processing systems in the working memory: System 1, which leads to a quick and effortless, more associative and intuitive processing, and System 2, which leads to a slow and effortful, more analytic and deliberate processing. Whereas perceiving information processing as easy activates System 1, perceiving information processing as difficult activates System 2. Thus, increasing the perceived difficulty associated with a cognitive task (i.e., disfluency) stimulates deeper processing and a more analytic and elaborative thinking rather than a heuristic and intuitive reasoning (Alter et al. 2007). Taking James (1890/ 1950) and Alter et al. (2007) into consideration, the beneficial effects of disfluency on learning outcomes can be explained by the fact that the subjective, metacognitive perception of the learning process as difficult leads to an activation of System 2. This goes hand in hand with a deeper processing and better learning outcomes (Eitel et al. 2014).

Overall, the disfluency effect has been shown only for text-based instructional material (Diemand-Yauman et al. 2011; Eitel et al. 2014 (Experiment 1); French et al. 2013; Sungkhasettee et al. 2011). It could not be demonstrated for either spoken texts (Kühl et al. 2014) or pictures (Eitel et al. 2014). Even with regard to text-based learning material, there are studies which could not replicate the disfluency effect. Whereas some studies found a neutral

effect of disfluency on memory performance (Eitel et al. 2014 (Experiment 2); Guenther 2012; Song and Schwarz 2008; Rhodes and Castel 2008), other studies revealed even a negative effect of disfluency (Yue et al. 2013). Eitel et al. (2014) interpreted this heterogeneous data situation in a way that disfluent instructional material does not necessarily foster learning. Therefore they questioned on the one hand stability and generalizability of the disfluency effect, and on the other hand its impact for educational practice. Overall, it seems necessary to further elaborate on theoretical as well as on empirical issues of the disfluency effect. Hence, we first want to discuss the relationship between disfluency and cognitive load and second possible constraints of the disfluency effect with respect to specific learner characteristics.

Cognitive load theory

As described, disfluency improves learning by an evoked deeper processing. This goes hand in hand with an additional cognitive load. The Cognitive Load Theory (Sweller 1994; Sweller et al. 1998, 2011) assumes three types of cognitive load (CL): intrinsic (ICL), extraneous (ECL), and germane CL (GCL).

First, ICL is caused by the inherent complexity of the learning task and therefore by the element interactivity. The more elements a learner has to keep in mind simultaneously, the higher ICL is. Hence, ICL is fixed by a given task and cannot be influenced without changing the task. This type of load also depends on the learner's prior knowledge. With more expertise a learner is able to construct meaningful chunks of information. Hence, he or she can reduce the amount of single unrelated elements in working memory and therefore will experience less ICL. Second, ECL is caused by a poorly designed instruction and therefore completely under control of the instructional designer. This kind of load is extraneous, because the learner needs cognitive resources that are not directed to the learning task itself, but to additional demands like navigating, searching etc. If this type of load is too high learning can be massively hindered. Third, GCL reflects the learner's activities which contribute to a deeper comprehension of instructional material by processing, construction, and automation of schemas. This type of load is germane to the learning process because it is, in contrast to ECL, exclusively directed to the learning task. Hence, it is desirable to increase this type of load, e.g., by activating the learner with encouraging and motivating tasks. All three types of load are additive, e.g., together they constitute the overall amount of CL a learner is experiencing during a learning task. This CL burdens the working memory whose capacity is limited (Cowan 2001; Hasselhorn and Gold 2009; Miller 1994), To prevent an overload which would inhibit learning a lot, it would be most efficient to reduce ECL and to enhance GCL to foster learning (Sweller et al. 1998, 2011).

So the disfluency effect and its additional load is contradictory to the Cognitive Load Theory (Eitel et al. 2014). Making the learning material less legible should not influence ICL or GCL, but it should affect ECL. Because of the poor instructional design, disfluent material increases ECL. On the same time, it should indeed contribute to a deeper processing of the material, because learners do have to engage themselves more. But according to Eitel et al. (2014) this does not affect GCL, because learners do not have to create any new information actively at all. According to Cognitive Load Theory, presenting disfluent material and thereby increasing ECL without increasing GCL should lead to worse learning outcomes (Eitel et al. 2014). Based on the fact that disfluent learning material causes an additional load which burdens the working memory (WM), the working memory capacity (WMC) could be a crucial factor for the success of disfluency.

Talking about the influence of disfluency on CL only makes sense if CL can be measured. Pass et al. (1994) postulate that learners are aware of their own cognitive load and that subjective ratings are therefore useful to measure mental effort in general. In further studies different researchers extended this idea to the point that all three types of cognitive load can be measured differentially (e. g., Ayres 2006; Klepsch and Seufert 2012; Paas et al. 2005).

Aptitude-treatment-interaction with working memory and prior knowledge

As already mentioned, recent research concerning disfluency showed inconsistent results. One general possibility to evaluate heterogeneous data situations is to take account of learner characteristics. It is usual that recommendations regarding multimedia design cannot be applied to all learners in the same fashion. According to the concept of aptitude-treatment-interaction (ATI; Snow 1989), instructional strategies (treatments) have different degrees of effectiveness for specific learners depending upon their individual learner characteristics (aptitudes). One important ability might be the learner's WMC. This capacity is described by the number of information which could be processed simultaneously. For deeper processing the learner needs to structure the given information and integrate information from long term memory as well as to build meaningful chunks. These chunks relieve the WM which allows to activate System 2 (Tverski and Kahneman 1974; James, 1890/ 1950) and process the information more deeply.

Additional demands on working memory caused by disfluent texts are possibly just usable by learners with a high WMC. Only learners with a high WMC might have enough capacity for the higher ECL caused by less legible texts and could engage in deeper processing and a more analytic and elaborative thinking rather than a heuristic and intuitive reasoning. WMC may work as an enhancer: *The instructional strategy of using disfluent text is only effective with sufficient WMC*. By contrast, learners with a low WMC should not be able to handle a higher ECL caused by disfluent material. Instead, the increased ECL will exceed the resources available and learners cannot allocate germane resources to the learning process. Thus, the construction of a situational model is hindered. Overall, disfluency should not be beneficial for those learners. In the case of learning with fluent material, ECL should not be increased which would result in learners with high as well as low WMC having similar learning outcomes.

Another factor which burdens the WM is the level of learning performance. Based on Blooms taxonomy for cognitive learning processes (1956) we differentiate learning outcomes that either requires learner's ability to recall, to comprehend or to apply the issue to be learned. These levels of processing also can be found in theories of text processing (e.g., Model of Text Comprehension; Van Dijk and Kintsch 1983) or in multimedia learning theories (e.g., Mayer 2005; Schnotz 2005). These approaches explain how texts are processed. First the learner constructs a mental representation of the text surface (through subsemantic processing). Second, he or she generates a propositional representation of the semantic content (through semantic processing). Third, the learner constructs a mental model of the subject matter the text deals with. Overall, these construction processes result from bottom-up as well as top-down activation of cognitive schemata.

While easier tasks like recall tasks only burden the WM little, more difficult tasks like comprehension or transfer tasks need more WMC. For higher order cognitive processes learners have not only to keep in mind single unrelated elements, but to combine them or even integrate information from long term memory, like we just described the processes activated by system 2. So WMC is the crucial factor if learners with a disfluent learning

material can also handle more difficult tasks. There is also empirical evidence for different consequences of disfluent material. Whereas Diemand-Yauman et al. (2011), French et al. (2013), and Sungkhasettee et al. (2011) showed beneficial effects of disfluency on retention, Eitel et al. (2014) demonstrated improvements in transfer. Thus, disfluency might increase learning outcomes on the lower as well as higher order levels of processing if the WMC is sufficiently high.

One more, in many studies learners' prior knowledge has been addressed as one crucial learner characteristic that moderates the effects of instructional design strategies (Kalyuga 2007; Seufert and Brünken 2004). While novice learners often benefit from an extended instructional design like providing additional pictures to a text, expert learners do not need such a support or may even suffer when additional information has to be actively disregarded and need some extra effort (expertise reversal effect; Kalyuga et al. 2003; Seufert 2003). Learners' prior knowledge works as a compensator for instructional shortcomings (Mayer and Sims 1994). Hence, ATI suggests that optimal learning results when the instruction fits exactly the learner's aptitudes.

Thus, disfluency might be beneficial only for learners with particular learner characteristics. Learners with too little prior knowledge are not able to build chunks. To process all single information burdens the WM and in addition with disfluent learning material leads to a cognitive overload. On the other side experts don't need furthermore help. They would only suffer when they have to invest additional cognitive resources related to disfluency (based on Seufert 2003). This is why we included only learners with a medium level of prior knowledge in our further analyses.

Potential confounding variables

According to the INVO-Model (*In*dividuelle *Vo*raussetzungen erfolgreichen Lernens; individual determinants of successful learning; Hasselhorn and Gold 2009), there are several determinants which play a crucial role in successful learning generally. Since the enjoyment during task performance, the interest in a task, the motivation to solve a task, and the prior knowledge have an influence on learning outcomes, these variables were assessed as potential confounding variables in the present study. Especially the motivational and affective variables could be relevant for learners' reaction on disfluent texts. Learners may decide to invest more or less mental resources based on their motivational or affective states and whether they find it motivating or frustrating to learn with such material. Nevertheless, we just controlled for these variables and did not include them as independent factors but focused on the interaction with WMC.

Research questions and hypotheses

As set out above, disfluency can lead to a better learning performance by encouraging deeper processing. This goes hand in hand with an additional extraneous cognitive load, which may only be compensated by learners with high WMC. Therefore WMC should be a crucial factor deciding if disfluency improves or inhibits learning. This leads to the question whether different results of disfluency research can be explained by an aptitude-treatment interaction between WMC and disfluency. One more it is yet unclear which levels of learning performance (retention, comprehension, or transfer; Bloom 1956) are fostered by disfluency.

So to test the enhancing effect of WMC to the different levels of learning performance, the present study investigated the ATI between WMC and disfluency. We expected an interaction between WMC and disfluency with respect to retention (*Hypothesis 1*), comprehension (*Hypothesis 2*), and transfer (*Hypothesis 3*) with stronger effects on higher levels of processing where WMC is increasingly relevant and thus can foster the construction of a situational model which is fundamental for higher test performance after learning. In the fluency condition, the WMC should not influence retention (*Hypothesis 1a*), comprehension (*Hypothesis 2a*), or transfer (*Hypothesis 3a*). In the disfluency condition, the WMC should affect learning outcomes: The higher the WMC, the better the expected retention (*Hypothesis 1b*), comprehension (*Hypothesis 2b*), and transfer performance (*Hypothesis 3b*) in the disfluency condition.

Aside from the ATI regarding learning outcomes, the present study wants to examine the theoretically expected effects of disfluency on the three types of CL empirically. According to Eitel et al. (2014), presenting disfluent material should not influence ICL or GCL, but lead to an increase of ECL. We assumed an interaction between WMC and disfluency with respect to ECL. In the disfluency condition, the expected higher ECL caused by the less legible material might be compensated by learners with a high WMC. Due to their high WMC, the increased ECL might load less on their WM compared to learners with a low WMC. In the case of learning with a fluent material, ECL should not be increased and hence learners with high as well as low WMC would experience a similar ECL.

Besides the ECL, to our knowledge, previous studies of disfluency effect have not yet used a differentiated measurement of ICL or GCL. Thus, the present study investigated the influence of disfluency on ICL and GCL. Since presenting disfluent material should not increase ICL or GCL (Eitel et al. 2014), learners with high as well as low WMC should experience a similar ICL or GCL. No main effects or interaction effects with respect to ICL or GCL are expected.

To test the effects on CL, the present study used a differentiated measurement of the three types of CL. We expected no effects regarding ICL or GCL. The fluency and disfluency condition should not differ with respect to ICL (*Hypothesis 4a*) or GCL (*Hypothesis 5a*) and there should be no interaction between WMC and disfluency regarding ICL (*Hypothesis 4b*) or GCL (*Hypothesis 5b*). Regarding ECL, we expected an interaction between WMC and disfluency (*Hypothesis 6*). In the fluency condition, the WMC should not influence ECL (*Hypothesis 6a*). In the disfluency condition, the WMC should affect ECL: The higher the WMC, the lower the expected ECL (*Hypothesis 6b*) in the disfluency condition.

Method

Participants and design

Altogether, 65 students from a German university participated for course credit and sweets in the study. As mentioned above, we excluded learners with too low (i.e., with less than 25 % of the maximum test score (=1.5 of 6 points) in the test for prior knowledge) or too high prior knowledge (i.e., with more than 75 % of the maximum test score (=4.5 of 6 points) in the test for prior knowledge). Hence, 47 subjects had a medium level of prior knowledge ($M=2.03$, $SD=1.80$) and were included in the analyses. Their mean age was 22.9 years ($SD=3.77$) and 85 % of them were females. Participants were randomly assigned to the fluency ($n=24$) and disfluency condition ($n=23$) of the first independent variable "learning material" (treatment-

factor). Their WMC served as the second independent variable (continuous aptitude-factor). As dependent variables, we measured learning performance in a retention test, a comprehension test, and a transfer test as well as ICL, ECL, and GCL.

Materials

The materials comprised a demographic questionnaire and the instructional materials. All materials were printed on sheets of paper. The text-based instructional material was adapted from a study by Schnotz and Bannert (1999). It dealt with "Time and date differences on earth" and consisted of two printed pages containing 1070 words. The text contained a table presenting eight cities from all over the world and their time differences compared to Greenwich. Text legibility was manipulated by presenting text either in easier-to-read font (Arial, 12 pt, black; legible text; see Fig. 1), or in harder-to-read font (Haettenschweiler, 12 pt, grayscale 35 %; less legible text; see Fig. 1). A similar manipulation was successfully applied in Diemand-Yauman et al. (2011) as well as in Eitel et al. (2014).

Measures

The paper-based self-developed test for *prior knowledge* consisted of six open questions about the content domain (e.g., "What are time zones?"). The open answers were compared with a predefined solution. Two points were given for each correct answer to the prior knowledge questions and the final score of the prior knowledge test was determined by adding up all points given for the prior knowledge questions. To maximize variance, three items were excluded from the analyses due to a solution probability of less than 10 % or more than 90 %. Responses ranged from 0 to 6 points.

The computer-based Numerical Memory Updating subtest of the WMC test (Oberauer et al. 2000) was used to assess *WMC*. In a 3 by 3 matrix, an increasing number of fields were activated. In the activated fields, numbers were presented one after another. Afterwards, arrows were presented which showed upwards or downwards. Arrows showing upwards were an indicator of adding one to the previously shown numbers, whereas arrows showing downwards were an indicator of subtracting one from the previously shown numbers. Up to three operations had to be performed with the initially given numbers. Participants had to memorize the initially presented numbers and their location to perform the arithmetic operations and memorize the transformed numbers. Finally, question marks were presented and subjects had to type in the overall result. After a feedback, the next turn started with new active fields and numbers. The computer-based program worked adaptively, so that the number of activated fields in the current turn depended on the performance in the previous turn. The number of correct overall results served as the score for the WMC, which could reach a maximum value

Arial, 12 pt, black	(Haettenschweiler, 12 pt, grayscale 35%
For the purpose of navigation of ships or planes we can consider the earth as a globe. The surface of this globe is subdivided by using the so called meridians. These areas are definded as time zones.	For the purpose of navigation of ships or planes we can consider the earth as a globe. The surface of this globe is subdivided by using the so called meridians. These areas are defined as time zones.

Fig. 1 Example of the learning material (translated)

of nine. Results ranged from 1 to 6 points. Even if this test deals with numbers, it does not measure mathematical abilities. For that the calculations are too easy. Difficulties arise from keeping all the different numbers and dealing with them.

The paper-based test for *learning outcomes* comprised tests for retention, comprehension, and transfer performance. The retention (e.g., "According to which principle, the time zones were classified?"), comprehension (e.g., "What time is it in Frankfurt, when it is 2 pm in Mexico City?"), and transfer tests (e.g., "Your flight starts on 12th of July from Tokyo. After an eight-hour-flight, you arrive in Bangkok. Which date and which time is it in Bangkok?") each consisted of five open questions about the content domain. For answering the questions in the comprehension and transfer tests, participants used a table presenting eight cities from all over the world and their time differences compared to Greenwich. The open answers were compared with a predefined solution. Two points were given for each correct answer to the retention, comprehension, and transfer questions and the final score of the retention, comprehension, and transfer tests were determined by adding up all points given for the corresponding questions. For retention, responses ranged from 1.5 to 10 points, for comprehension, they ranged from 4 to 10 points and for transfer, they ranged from 0 to 10 points.

The paper-based Cognitive Load Questionnaire (Klepsch and Seufert 2012) was used to assess ICL, ECL, and GCL. Three items assessed ICL (e.g., "For this task many things needed to be kept in mind simultaneously."), three items assessed ECL (e.g., "The design of this task was very inappropriate to really learn something."), and three items assessed GCL (e.g., "While solving this task, I had the goal to completely understand the subject."). Each item had to be rated on a seven-point Likert scale (ICL—responses after learning: $Min=1$; $Max=7$; ICL—responses after assessing learning outcomes: $Min=2$; $Max=7$; ECL - responses after learning and after assessing learning outcomes: $Min=1$; $Max=6$; GCL - responses after learning: $Min=1$; $Max=7$; GCL—responses after assessing learning outcomes: $Min=2$; $Max=7$).

Three paper-based self-developed items were used to assess the potential confounding variables *enjoyment* during task performance, *interest* in the task, and *motivation* to solve the task. Enjoyment during task performance was assessed by the item "How much did you enjoy the task performance?". Interest in the task was assessed by the item "I was interested in the tasks". Motivation to solve the task was assessed by the item "I was motivated to solve the task". Each item had to be rated on a seven-point Likert scale (Enjoyment ranged from 1 to 6 points after learning as well as after assessing learning outcomes; interest and motivation ranged both from 1 to 7 points).

Procedure

The study was conducted in one session, lasting about 45 min. The participants were tested in groups. After filling in the demographic questionnaire and completing the test of prior knowledge, the learning phase began for all participants simultaneously. Subjects were then asked to deal individually with the learning material. Afterwards, participants had to fill in the Cognitive Load Questionnaire and rate the enjoyment, interest, and motivation they experienced during learning by responding to the respective items. Thereafter, students were asked to fill in the tests for learning outcomes without any time restrictions. At the end, they had to fill in the Cognitive Load Questionnaire and rate the enjoyment, interest, and motivation they experienced during the tests for learning outcomes by responding to the respective items. In a prior study, we already conducted all relevant individual determinants of our subjects such as

WMC. The corresponding data set could be linked because the same code system had been used to identify the participants.

Results

To test our hypotheses we set up regression analyses. Descriptive data for all variables per condition can be seen in Table 1.

Control variables

We analyzed if the potential confounding variables differ between the two groups and if they correlate with any dependent variable. In case of group differences or a significant correlation, we controlled them in further analyses (retention performance correlated with motivation during assessing learning outcomes ($r=.32$, $p=.03$), comprehension correlated with prior knowledge ($r=.37$, $p=.01$)).

Learning outcomes

Regression analyses were applied for retention, comprehension, and transfer as dependent variables with the following predictors (entered simultaneously): learning material (fluent,

Table 1 Descriptive data for all variables per condition

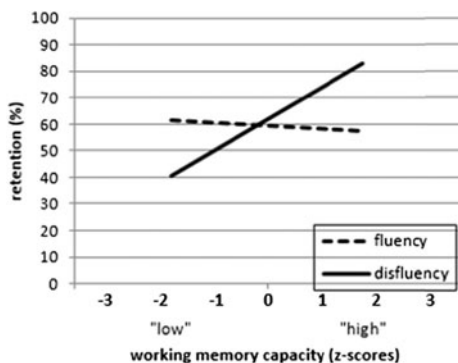|	Conditions			
	Fluency (n=24)		Disfluency (n=23)	
	M	*SD*	*M*	*SD*
Working memory capacity	3.96	1.23	4.09	.95
Prior knowledge	1.98	1.57	2.09	2.04
Retention (%)	59.79	17.35	61.52	18.12
Comprehension (%)	73.54	22.63	71.96	15.28
Transfer (%)	53.59	29.76	49.78	32.88
Intrinsic cognitive load after learning (%)	67.66	19.91	62.73	20.34
Extraneous cognitive load after learning (%)	50.20	20.20	54.45	17.34
Germane cognitive load after learning (%)	82.14	13.92	77.81	16.39
Enjoyment during learning	3.21	1.56	3.30	1.40
Interest during learning	3.83	1.76	4.09	1.70
Motivation during learning	4.75	1.29	4.70	1.58
Intrinsic load after assessing learning outcomes (%)	77.71	18.78	76.60	17.40
Extraneous load after assessing learning outcomes (%)	45.67	19.53	52.17	18.75
Germane load after assessing learning outcomes (%)	77.49	14.56	71.43	15.53
Enjoyment during assessing learning outcomes	3.36	1.79	3.30	1.64
Interest during assessing learning outcomes	3.95	1.84	3.87	1.89
Motivation during assessing learning outcomes	5.27	1.52	4.70	1.79

disfluent), WMC, interaction term learning material × WMC, and respective significant control variables. In a first step, learning material was coded with 0 for the fluency condition and 1 for the disfluency condition. In a second step, learning material was recoded (fluency = 1, disfluency = 0) and the regression model was conducted again. This method of "re-centering", which was proposed by Aiken and West (1991), enables analyzing the specific impact of WMC for the condition which is coded with 0. The WMC as well as the control variables were z-standardized. The dependent variables were transformed in percentages.

For *retention performance*, the regression model was significant ($F(4, 44)=3.56$, $p=.01$, adjusted $R^2=.19$). The learning material was no significant predictor of retention ($\beta=2.42$, $t(44)=.53$, n.s.), indicating that the two experimental groups did not differ with respect to retention. As predicted in Hypothesis 1, the interaction term was significant in the prediction of retention performance ($\beta=11.66$, $t(44)=2.46$, $p=.02$). As predicted in Hypothesis 1a, for the fluency condition the WMC was not a significant predictor for retention ($\beta=-1.01$, $t(44)=-.35$, n.s.). As predicted in Hypothesis 1b, for the disfluency condition the WMC was a significant predictor for retention ($\beta=10.66$, $t(44)=2.86$, $p=.01$): The higher the WMC, the better the retention performance with the disfluent text. The interaction pattern is depicted in Fig. 2. The control variable "motivation during assessing learning outcomes" had a significant impact on retention performance ($\beta=5.59$, $t(44)=2.38$, $p=.02$) and had been controlled therefore.

For *comprehension performance*, the regression model was significant ($F(4, 46)=3.33$, $p=.02$, adjusted $R^2=.17$). The learning material was no significant predictor for comprehension ($\beta=-2.36$, $t(46)=-.46$, n.s.), indicating that the two experimental groups did not differ with respect to comprehension. As predicted in Hypothesis 2, the interaction term was significant in the prediction of comprehension performance ($\beta=12.62$, $t(46)=2.33$, $p=.03$). As predicted in Hypothesis 2a, for the fluency condition the WMC was not a significant predictor for comprehension ($\beta=-3.90$, $t(46)=-1.20$, n.s.). As predicted in Hypothesis 2b, for the disfluency condition the WMC was a significant predictor for comprehension ($\beta=10.66$, $t(46)=2.86$, $p<.05$): The higher the WMC, the better the comprehension performance with the disfluent text. The interaction pattern is depicted in Fig. 3. The control variable "prior knowledge" had a significant impact on comprehension ($\beta=7.90$, $t(46)=3.04$, $p<.01$).

For *transfer performance*, the regression model was not significant ($F<1$, n.s., adjusted R^2 <.01). The learning material was no significant predictor for transfer ($\beta=-4.26$, $t(45)=-.46$, n.s.), indicating that the two experimental groups did not differ with respect to transfer. In

Fig. 2 Interaction between condition (fluency, disfluency) and working memory capacity for retention (controlled for "motivation during assessing learning outcomes")

Fig. 3 Interaction between condition (fluency, disfluency) and working memory capacity for comprehension (controlled for "prior knowledge")
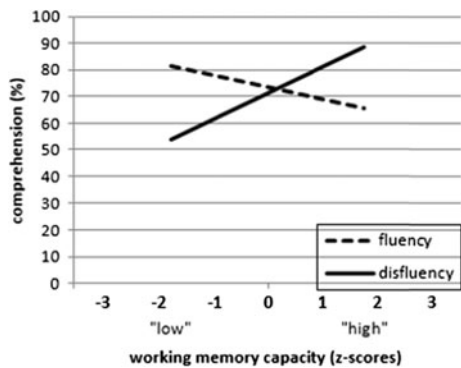
contrast to Hypothesis 3, the interaction term was not significant in the prediction of transfer performance ($\beta=-2.27$, $t(45)=-.23$, n.s.).

Cognitive load

To test the hypotheses, regression analyses were applied for ICL, ECL, and GCL (after learning and after testing for learning outcomes, respectively) as dependent variables with the following predictors (entered simultaneously): learning material (fluent, disfluent), WMC, interaction term learning material×WMC, and respective significant control variables (ICL after assessing learning outcomes correlated with enjoyment during assessing learning outcomes ($r=-.36$, $p=.02$); ECL after learning correlated with interest during learning ($r=-.31$, $p=.04$) as well as with motivation during learning ($r=-.41$, $p<.01$);), ECL after assessing learning outcomes correlated with enjoyment during assessing learning outcomes ($r=-.34$, $p=.02$), GCL after learning correlated with motivation during learning ($r=.51$, $p<.001$ and GCL after assessing learning outcomes correlated with motivation during assessing learning outcomes ($r=.44$, $p<.01$)) . As in the analyses for learning outcomes, we used the "re-centering" method. The WMC as well as the control variables were z-standardized.

For *ICL* (after learning and after and assessing learning outcomes), the regression model was not significant (after learning: $F<1$, n.s., adjusted $R^2<.01$; after assessing learning outcomes: $F(4, 44)=1.64$, n.s., adjusted $R^2=.06$). As predicted in Hypothesis 4a, the learning material was no significant predictor for ICL (after learning: $\beta=-5.44$, $t(46)=-.93$, n.s.; after assessing learning outcomes: $\beta=-1.55$, $t(44)=-.30$, n.s.), indicating that the two experimental groups did not differ with respect to ICL. As predicted in Hypothesis 4b, the interaction term was not significant in the prediction of ICL (after learning: $\beta=3.80$, $t(46)=.61$, n.s.; after assessing learning outcomes: $\beta=2.07$, $t(44)=.38$, n.s.). The control variable "enjoyment during assessing learning outcomes" had a significant impact on ICL after assessing learning outcomes ($\beta=-6.14$, $t(44)=-2.31$, $p=.03$).

For *GCL* (after learning and after assessing learning outcomes), the regression model was significant (after learning: $F(4, 46)=4.34$, $p<.01$, adjusted $R^2=.23$; after assessing learning outcomes: $F(4, 44)=2.71$, $p=.04$, adjusted $R^2=.14$). As predicted in Hypothesis 5a, the learning material was no significant predictor for GCL (after learning: $\beta=-4.92$, $t(46)=-1.25$, n.s.; after assessing learning outcomes: $\beta=-3.92$, $t(44)=-0.92$, n.s.), indicating that the two experimental groups did not differ with respect to GCL. As predicted in Hypothesis 5b, the interaction term was not significant in the prediction of GCL (after learning: $\beta=-0.79$,

$t(46)=-0.19$, n.s.; after assessing learning outcomes: $\beta=2.39$, $t(44)=.54$, n.s.). The control variable "motivation after learning" had a significant impact on GCL after learning ($\beta=7.62$, $t(46)=-3.82$, $p<.001$). The control variable "motivation during assessing learning outcomes" had a significant impact on GCL after assessing learning outcomes ($\beta=6.17$, $t(44)=-2.82$, $p<.01$).

For *ECL* (after learning and after assessing learning outcomes), the regression model was marginally significant (after learning: $F(5, 46)=2.32$, $p=.06$, adjusted $R^2=.13$; after assessing learning outcomes: $F(4, 44)=1.93$, $p=.06$, adjusted $R^2=.13$). The learning material was no significant predictor for ECL (after learning: $\beta=4.31$, $t(46)=-.84$, n.s.; after assessing learning outcomes: $\beta=5.95$, $t(44)=1.08$, n.s.), indicating that the two experimental groups did not differ with respect to ECL. In contrast to Hypothesis 6, the interaction term was not significant in the prediction of ECL (after learning: $\beta=-5.11$, $t(46)=-.93$, n.s.; after assessing learning outcomes: $\beta=.51$, $t(44)=.09$, n.s.). Whereas the control variable "interest during learning" had no significant impact on ECL after learning ($\beta=-2.85$, $t(46)=-.92$, n.s.), the "motivation during learning" had a significant impact ($\beta=-6.07$, $t(46)=-2.01$, $p<.05$). The control variable "enjoyment during assessing learning outcomes" had a significant impact on ECL after assessing learning outcomes ($\beta=-5.98$, $t(44)=-2.12$, $p=.04$).

Discussion

Overall, in the present study, we investigated the ATI between WMC and disfluency with respect to retention, comprehension and transfer. We found the expected enhancing effect of WMC on retention and comprehension performance: The higher the WMC, the better the retention and comprehension performance in the disfluency condition. In the fluency condition, the WMC did not influence the learning outcomes. Thus, disfluency only paid off when learners had sufficient WMC. Only with sufficient cognitive resources learners were able to use the stimulation, to intensify their learning process to a deeper level (System 2; Tverski and Kahneman 1974; James, 1890/ 1950). Without taking the WMC into account, we could not have shown the disfluency effect. Hence, a possible explanation for the heterogeneous data situation regarding the disfluency effect is that learner characteristics like the WMC have not been taken into account. Moreover, in contrast to Eitel et al. (2014) who demonstrated the disfluency effect with respect to transfer performance, we had no evidence for the disfluency effect for transfer performance in the present study—neither as a main effect nor as an ATI effect between WMC and disfluency. However, our results are partly in line with Diemand-Yauman et al. (2011), French et al. (2013), and Sungkhasettee et al. (2011) who showed the beneficial effects of disfluency on lower order processes like retention. Probably learners' WMC was again the critical factor. While disfluent material already burdens the WM, there is not that much capacity left for difficult tasks like transfer. System 2 (Tverski and Kahneman 1974; James, 1890/ 1950) could not be activated. So disfluency in addition with high cognitive load related tasks lead to a cognitive overload and therefore not to an advantage of disfluency. Finally, we could not find a general disfluency effect but only for learners with high WMC.

Further research is needed to examine these discrepancies. Additionally, it needs to be approved that one could find the same results concerning the ATI between WMC and disfluency with another measurement of WM. A subject might get a better result in the Numerical Memory Updating subtest (Oberauer et al. 2000) if he or she has an affinity towards numbers. The same property could also have influenced the results of the post test.

So maybe our measurement was confounded by this similarity. One more it should be investigated whether one could find the same results using learning material with a less mathematical topic.

Besides learning outcomes, we investigated the effects of disfluency on the three types of CL. To our knowledge, we were the first to use a differentiated measurement of the three types of CL in disfluency research. As expected, neither disfluency nor the interaction term disfluency × WMC affected ICL or GCL. Thus, our assumptions regarding ICL and GCL were supported. However, our hypotheses regarding GCL were based on the assumptions of Eitel et al. (2014). According to Eitel et al. (2014), disfluency should not affect GCL, because learners do not have to actively generate new information.

But considering the Model of Text Comprehension (Van Dijk and Kintsch 1983), one could argue that disfluency does increase GCL. According to Eitel et al. (2014), subjects learning with a disfluent text would not be forced to actively generate new information and thus there would be no increase in GCL. But GCL may not only be related to the generation of new information. Considering the text processing models, one could argue that learners receiving a disfluent text would be forced to actively invest more effort in the subsemantic processing of the text and the construction of the mental representation of the text surface. Hence, disfluency could increase GCL by intensifying subsemantic processing.

Although the present study showed that disfluency did not affect GCL, this does not necessarily have to be evidence against the assumption of disfluency increasing GCL. Eventually the Cognitive Load Questionnaire we used to assess GCL did not explicitly refer to these subsemantic processes. One more, the questionnaire only measures subjective ratings of cognitive load. Although learners are aware of their cognitive burdens, this does not mean that the ratings are conformed to the objective load coincidentally. This includes an additional metacognitive step of self-monitoring. Brünken et al. (2003) argue that objective ratings should be preferred therefore. So our results are only representative for the subjectively perceived level of cognitive load.

So, future research should investigate GCL with a questionnaire that measures GCL associated with subsemantic processing and additional objective load measures. In addition to Alter et al. (2007) who attributed the beneficial effects of disfluency to the stimulation of a deeper processing and a more analytic and elaborative reasoning, an increased GCL could explain the positive effects of disfluency, too. Another focus of prospective research is to monitor the metacognitive skill which is necessary to report your cognitive load approximately objective.

Moreover, since the majority of the studies (Diemand-Yauman et al. 2011; French et al. 2013; Sungkhasettee et al. 2011) showed the beneficial effects of disfluency only on lower order processes, this can be regarded as evidence that disfluency increases GCL by intensifying subsemantic processing. In the light of the above, the results of the present study could be explained, too. Eventually, we could not demonstrate the disfluency effect with respect to transfer, because transfer represents a higher order process.

However, we investigated the effects of disfluency on ECL. In contrast to our hypotheses, neither disfluency nor the interaction term disfluency × WMC were significant predictors for ECL. Eitel et al. (2014) could not show an increased ECL when learning with the disfluent material, too. Hence, the role of metacognitive regulation and its possible effects on ECL need further investigation. Possibly, the items which assessed ECL do not cover the different features of ECL. Disfluency might have influenced other facets of ECL, which are not included in the questionnaire we used. Moreover, the question arises, how well the subjects were able to estimate ECL. The

estimation of CL depends on one's capability to introspection which was not assessed in the present study. Finally, the present study was no real exam situation, so that participants were not under pressure to perform as well as in a real exam situation. Thus, they might not consider ECL as particularly high when learning with disfluent material.

Altogether, the present study has some more limitations. First, our sample is rather small and not representative. Most of our subjects were females who were young students with—due to our numerous clauses—best results in their high school diploma and therefore probably great learning skills. Consequently, the results cannot be necessarily expected to be generalizable for other learning types. Second, we did not use any manipulation check items to evaluate the fluency or disfluency of our instructional material. The manipulation of text legibility was similar to the successful manipulation applied in Diemand-Yauman et al. (2011) as well as in Eitel et al. (2014). But there is no systematic review in which text legibility has been evaluated depending on different fonts, font colors, or font sizes. Hence, the question arises which features of the font manipulation are responsible for the disfluency effect. Since the font, the font color, and font size were manipulated in the present study, this question cannot be answered at this point and should be investigated systematically in further research. One more the role of metacognition is not clear, yet. It is thinkable that metacognitive skills like the awareness of learning with a disfluent font or monitoring cognitive load while learning impact learning outcomes. It is possible that the learner reacts rather negative by realizing that they have to invest more effort. Moreover, one crucial factor for the metacognitive decision to invest more effort due to disfluency is based on learners' sensitivity towards their own cognitive resources and their experiences with different learning materials. Only when learners realize that the material is "difficult" for them—and when they feel able to enhance their effort based on their available resources, disfluency may cause positive effects. This sensitivity towards the task properties and ones own cognitive system is the product of several learning experiences that are metacognitively monitored and evaluated. Many trainings on learning strategies address these issues and foster metacognitive awareness especially of young learners (see the meta-analysis of Donker et al. 2014).

Since the present study could show that learner characteristics like the WMC should be taken into account when investigating the disfluency effect, future research should identify other aptitudes, besides WMC, which may interact with disfluency. Especially learners' prior knowledge could be a relevant moderating variable, as has been often proved in ATI studies (Kalyuga et al. 2003; Seufert 2003). One could argue that prior knowledge also relieves learners working memory capacity due to meaningful chunks in working memory and therefore a reduced amount of intrinsic cognitive load. Hence, the effects should be the same as in the present study and disfluency should be more effective with increasing prior knowledge. Moreover, as Diemand-Yauman et al. (2011) stated, the point at which a text can be considered as disfluent but not yet as illegible should be examined. Only disfluent texts can improve learning performance—with sufficient WMC and with respect to specific learning goals—whereas illegible texts should hinder learning. But in contrast to Diemand-Yauman et al. (2011), we do not believe that teachers can integrate disfluent material so easily in their lessons. If disfluency only pays off when learners have a medium level of prior knowledge and sufficient WMC, how shall teachers identify these learners in a quick and cheap way? How can they deal with the problem that only a—possibly very small—part of their learners can profit from less legible texts? Consequently, the question on the practical application of disfluent material arises. Nevertheless we think that disfluent fonts can pay off in special learning environments, for example classes with highly talented students in the middle of a learning

process of one special topic. One more one must mention that fonts are a surface characteristic which is quite easy and cheap to manipulate. In this context, future research is necessary to investigate whether the disfluency effect is only a so-called novelty effect (Tulving and Kroll 1995; Rummer et al. this issue). This would mean that the disfluency effect only occurs at the beginning when the design of the instructional material is considered new and unusual and attracts the learner's attention. Later, when learning repeatedly with the less legible texts, one might get used to this kind of texts. Possibly, the disfluent material might not seem new or unusual over time and the beneficial effects caused by disfluency might disappear. One interesting practical conclusion could be to train students' metacognitive skills by using texts with varying fonts and hence with varying fluency and by reflecting these learning experiences. Thus, learners can strengthen their metacognitive knowledge about difficulties and affordances of tasks and learn more about their way to deal with these affordances.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.

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*, 569–576. doi:10.1037/0096-3445.136.4.569.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*, 389–400.

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

Bjork, R. A. (2013). Desirable difficulties perspective on learning. In H. Pashler (Ed.), *Encyclopedia of the mind* (Vol. 4, pp. 134–146). Thousand Oaks: Sage Publication.

Bloom, B. S. (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York: David McKay.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53–61.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185. doi:10.1017/S0140525X01003922.

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*, 111–115. doi:10.1016/j.cognition.2010.09.012.

Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: a meta-analysis. *Educational Research Review, 11*, 1–26. doi:10.1016/j.edurev.2013.11.002.

Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: does harder-to-read mean better-to-understand? *Applied Cognitive Psychology, 28*, 488–501. doi:10.1002/acp.3004.

French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, A., Heritage, J., Poland, R. L., Reece, S., & Tabor, J. (2013). Changing fonts in education: how the benefits vary with ability and dyslexia. *The Journal of Educational Research, 106*, 301–304. doi:10.1080/00220671.2012.736430.

Guenther, R. K. (2012). Does the processing fluency of a syllabus affect the forecasted grade and course difficulty? *Psychological Reports, 110*, 946–954. doi:10.2466/01.11.28.PR0.110.3.946-954.

Hasselhorn, M., & Gold, A. (2009). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren* [Educational psychology: Successful learning and teaching]. Stuttgart: Kohlhammer.

James, W. (1950). *The principles of psychology*. New York: Dover (**Original work published 1890**).

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539. doi:10.1007/s10648-007-9054-3.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23–31.

Klepsch, M., & Seufert, T. (2012). Subjective differentiated measurement of cognitive load. *Paper presented at the 5th International Cognitive Load Theory Conference, Tallahassee (USA).*

Kühl, T., Eitel, A., Damnik, G., & Körndle, H. (2014). The impact of disfluency, pacing, and students' need for cognition on learning with multimedia. *Computers in Human Behavior, 35*, 189–198. doi:10.1016/j.chb.2014.03.004.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401. doi:10.1037/0022-0663.86.3.389.

Miller, G. A. (1994). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 101*, 343–352. doi:10.1037/0033-295X.101.2.343.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017–1045. doi:10.1016/S0191-8869(99)00251-2.

Paas, F., Tuovinen, J. E., van Merriënboer, J. J., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research & Development, 53*, 25–34.

Pass, F. G., van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419–430.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*, 615–625. doi:10.1037/a0013684.

Rummer, R., Schweppe, J., & Schwede, A. (this issue). Fortune is fickle: null-effects of disfluency on learning outcomes. *Metacognition and Learning.*

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–69). Cambridge: Cambridge University Press.

Schnotz, W., & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen [Influence of the type of visualization on the construction of mental models during picture and text comprehension]. *Zeitschrift für Experimentelle Psychologie, 46*, 217–236. doi:10.1026//0949-3964.46.3.217.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction, 13*, 227–237. doi:10.1016/S0959-4752(02)00022-1.

Seufert, T., & Brünken, R. (2004). Supporting coherence formation in multimedia learning. In H. Niegemann, R. Brünken, & D. Leutner (Eds.), *Instructional design for multimedia learning. Proceedings of the EARLI SIG 6 biannual workshop 2002 in Erfurt* (pp. 138–147). Waxmann: Münster.

Snow, R. E. (1989). Aptitude-treatment interaction as a framework of research in individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 13–59). New York: Freeman.

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do. *Psychological Science, 19*, 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 and Review, 18*, 973–978. doi:10.3758/s13423-011-0114-9.

Sweller, J. (1994). Cognitve load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312.

Sweller, J., van Merrienboër, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. doi:10.1023/A:1022193728205.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer.

Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review, 2*, 387–390. doi:10.3758/BF03210977.

Tverski, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124–1131.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

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 and Cognition, 41*, 229–241. doi:10.3758/s13421-012-0255-8.