Running Head: WORDING IN MULTIPLE DOCUMENTS

The Effect of the Wording of Multiple Documents on Learning

Cornelia Schoor, Nadine Melzner and Cordula Artelt

University of Bamberg

Corresponding Author:

Cornelia Schoor

University of Bamberg

Department of Educational Research

Markusplatz 3

96047 Bamberg

Germany

Phone: +49 951 863 2064

Fax: +49 951 863 5061

cornelia.schoor@uni-bamberg.de

Acknowledgments

We want to thank Regina Jucks and Elisabeth Mayweg-Paus for providing us with their material, Wolfgang Lenhard for organizing a room for conducting the study at the University of Würzburg, and Sabrina Brunner for English language proofreading.

Post print version of the article: Schoor, C., Melzner, N., & Artelt, C. (2019). The effect of the wording of multiple documents on learning. *Zeitschrift für Pädagogische Psychologie*, *33*(3-4), 223-240. https://dx.doi.org/10.1024/1010-0652/a000246

The Effect of the Wording of Multiple Documents on Learning

Abstract

Learning from multiple documents is challenging, amongst other things, because authors often do not use the same words for the same concept. Derived from theoretical considerations and earlier research, it was assumed that different wording across multiple documents leads to higher cognitive load, but can be beneficial for the learning of highly motivated students. In the present study, N=100 university students (laypersons with regard to the topic) took part. They read two texts on the topic of depression either with the same wording or with different wording and worked on a case study. Prior knowledge, epistemic beliefs, cognitive load, motivational state during learning, and post learning knowledge were assessed. While controlling for prior knowledge, a significant interaction effect was found from wording and motivational state on post knowledge about concepts. Different wording was beneficial only for students with relatively high motivation. However, wording had no effect on cognitive load.

Keywords: Multiple document comprehension; wording; motivation; cognitive load; knowledge acquisition.

Der Einfluss der Wortwahl in multiplen Dokumenten auf das Lernen

Zusammenfassung

Lernen aus multiplen Dokumenten ist herausfordernd, unter anderem weil verschiedene Autoren oft nicht dasselbe Wort für dasselbe Konzept verwenden. Basierend auf theoretischen Überlegungen und früherer Forschung wurde angenommen, dass eine unterschiedliche Wortwahl zu höherer kognitiver Belastung führt, jedoch für hochmotivierte Lernende lernförderlich wirkt. N = 100 Studierende, die bezüglich des Themas Laien waren, nahmen an der Studie teil. Sie lasen zwei Texte zum Thema Depression entweder mit gleicher oder ungleicher Wortwahl und bearbeiteten eine Fallgeschichte. Vor- und Nachwissen sowie epistemische Überzeugungen, kognitive Belastung und Motivation wurden erfasst. Unter Kontrolle von Vorwissen wurde ein signifikanter Interaktionseffekt von Wortwahl und Motivation auf das Nachwissen gefunden. Unterschiedliche Wortwahl war nur für relativ hoch motivierte Studierende lernförderlich. Unterschiedliche Wortwahl hatte jedoch keinen Einfluss auf kognitive Belastung.

Schlüsselwörter: multiple Dokumente; Wortwahl; Motivation; kognitive Belastung; Wissenserwerb.

The Effect of the Wording of Multiple Documents on Learning

Utilizing multiple documents to learn a single topic is a key task in university studies. Students have to learn from multiple documents, for example, in order to prepare presentations in seminars, write term papers or write their thesis. The authors of these documents often do not use the same wording or terminology for the same concept. Therefore, the students themselves have to identify different wording for the same concept in order to relate information across documents. However, this might be difficult if the students do not possess sufficient prior knowledge about the topic and/or lack the motivation to elaborate on the concepts.

Research on the influence that wording has on learning with multiple documents has been scarce. In the present paper, we test whether results taken from a study on collaborative learning are also valid for individual learning. This study is worthwhile for two reasons: First, the (direct or conceptual) replication of results is an endeavor that should enjoy greater value in psychological research (e.g., Shrout & Rodgers, 2018; Ulrich et al., 2016). Conducting and publishing more replication studies could be one component needed to prevent scientific literature from becoming a distorted image of reality (Ulrich et al., 2016). Conceptual replication, in particular, can contribute to analyzing the generalizability of results (Shrout & Rodgers, 2018).

Second, collaborative learning and individual learning are different with regard to several aspects that speak in favor of the assumption that results cannot be transferred easily from one scenario to another and that a replication is equivalent to a conceptual replication. For example, collaborative learning can be more motivating than individual learning (e.g., Dolmans & Schmidt, 2006; Ginsburg-Block & Fantuzzo, 1998; Hänze & Berger, 2007; Nichols, 1996; cf. Schoor & Bannert, 2013). In collaborative learning, there are motivational mechanisms that are genuine to the group context and that have no counterpart in individual

learning (e.g., Deutsch, 1949; Johnson & Johnson, 1992). Moreover, learners can pool their cognitive resources during collaborative learning, which is not the case for individual learning (e.g., F. Kirschner, Paas, & Kirschner, 2009a). The differences between individual and collaborative learning might make it more difficult to deal with different wording in individual learning than in collaborative learning.

Therefore, the present paper aims to contribute to existing research on the influence that the wording has on learning with multiple documents by conceptually replicating results from collaborative learning for the scenario of individual learning. First, we sketch the state of research on learning from multiple documents. Then, we consider potential benefits and pitfalls of different wording, thereby also contrasting individual and collaborative learning. Afterwards, we introduce motivation as a potential moderator and explain the hypotheses for the present study. Finally, we present and discuss results concerning the impact of using the same versus different wording on individual learning from multiple documents.

1 Learning from Multiple Documents

Reading multiple documents on a single topic involves not only the comprehension of single documents but also the relations that the documents have to each other. Within the Documents Model Framework (e.g., Britt, Perfetti, Sandak, & Rouet, 1999; Britt & Rouet, 2012; Perfetti, Rouet, & Britt, 1999), it is assumed that readers not only have to integrate content across documents (integrated situation model), but also have to represent information about the sources (intertext model) and relate it to content information (documents model).

Multiple documents can be seen as a special case of multiple representations (cf. Ainsworth, 2006; Schnotz & Bannert, 2003; Seufert, 2009), which is why research on multiple documents can benefit from research on multiple representations. Multiple representations can be described according to their functional relationship: they can

complement each other, they can constrain the interpretation of each other, or they can assist in constructing a deeper understanding of the topic (Ainsworth, 1999, 2006). In order to fulfill these functions, mapping processes are necessary, that is to say the comparison and relation of information across representations or documents (Schoor & Artelt, 2015; Van Meter & Firetto, 2008; cf. Gernsbacher, 1997; Wineburg, 1991). When processing multiple representations, learners compare incoming information to information already available in working memory. If the comparison reveals that there is enough overlap ("coherence" in the terms of Gernsbacher, e.g. 1997), they integrate the incoming information with the prior information. Otherwise, the new information cannot be integrated and is built into a new structure (cf. Gernsbacher, 1997).

Research has identified several strategies that are beneficial for multiple document comprehension (MDC). Closely related to mapping is corroboration, that is the comparison of claims across documents (Wineburg, 1991). Moreover, Wineburg (1991) revealed that experts apply not only corroboration, but also contextualization (they relate the content of the documents to prior knowledge) and sourcing (they attend to information about the source). More recent research has found that the strategy of self-explanation is also related to multiple document comprehension (Wolfe & Goldman, 2005) and that better learners do more self-explanation on reliable (as compared to unreliable) websites (Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012).

In addition to strategies, there are several learner characteristics that are related to MDC (cf. Barzilai & Strømsø, 2018; List, Stephens, & Alexander, in press; Maier, Richter, Nauroth, & Gollwitzer, in press; Merkt, Werner, & Wagner, 2017; Stang Lund, Bråten, Brandmo, Brante, & Strømsø, in press), among them prior content knowledge and epistemic beliefs. The research on prior content knowledge traces back to the seminal study by Wineburg (1991) who found strategy differences between content experts and novices. But

also recent research has consistently shown a positive effect of prior content knowledge on MDC (e.g., Kobayashi, 2009; Le Bigot & Rouet, 2007; List et al., in press; Stang Lund et al., in press).

Epistemic beliefs are beliefs about the nature of knowledge and knowing (Greene, Azevedo, & Torney-Purta, 2008; Hofer & Pintrich, 1997; Schommer, 1990; Stahl & Bromme, 2007). There are several conceptualizations of epistemic beliefs (e.g., Greene et al., 2008; Hofer & Pintrich, 1997; King & Kitchener, 2004; Kuhn & Weinstock, 2002; Schommer, 1990). Hofer and Pintrich (1997), for example, differentiate the dimensions "simplicity of knowledge", "certainty of knowledge", "source of knowledge", and "justification for knowing". The relationship between epistemic beliefs and MDC is well documented (cf. Bråten, Britt, Strømsø, & Rouet, 2011; Ferguson, 2015). Bråten et al. (2011) reviewed how the four dimensions of Hofer and Pintrich (1997) are related to MDC. Based on this review, they assume that beliefs about the simplicity of knowledge and the justification of knowing are related to the integration of content across documents, while beliefs about the certainty of knowledge and the source of knowledge are more closely related to the representation of sources.

2 Wording in Multiple Documents in Individual and Collaborative Learning

There is only little research on the effects that different wording across multiple documents has on multiple document comprehension, although different authors often use different wording. There is first evidence that different wording can be beneficial for collaborative learning. Jucks and Paus (2013) gave two learning partners of a dyad each a text about depression, using either the same or different wording from the wording of their learning partner's text. Jucks and Paus (2013) were able to show that dyads who had received texts with different wording produced better answers for their collaborative task and performed better in an individual test of post-task knowledge. The positive effect was

probably due to the dyads with different wording discussing and elaborating on the meaning of central concepts more often than dyads with the same wording. However, Jucks and Paus (2013) had a different research focus and did not report whether the elaboration they observed in the dyads was mediating the effect of wording on the collaborative task solution and on individual post knowledge.

Individual and collaborative learning are two learning situations that are quite distinct in several aspects. For the present study, two aspects are especially relevant: the possibility of sharing processing demands (e.g., F. Kirschner et al., 2009a), and motivation (e.g., Dolmans & Schmidt, 2006; Ginsburg-Block & Fantuzzo, 1998; Hänze & Berger, 2007; Nichols, 1996). Therefore, it is not necessarily the case that results from collaborative learning can be replicated in individual learning. Based on theoretical considerations and empirical results, the following alternative predictions can be made about the effect of different wording across documents (as compared to the same wording) on MDC in an individual learning situation: 1. Different wording is beneficial for individual learning and for collaborative learning, 2. different wording impedes individual learning, or 3. different wording is beneficial for a subgroup of individual learners only.

2.1 Alternative Prediction 1: Different Wording is Beneficial

Based on the results of Jucks and Paus (2013), one can argue that different wording is beneficial in individual learning as well. This prediction is also supported by the fact that different wording could act as a *desirable difficulty* (R. A. Bjork, 1994). Desirable difficulties are difficulties that trigger processes that are beneficial for learning (E. L. Bjork & R. A. Bjork, 2011). For example, distributed practice (cf. Küpper-Tetzel, 2014) or tests instead of presentations (cf. Rowland, 2014) generate desirable difficulty and have positive long-term effects on retention. Measures for creating disfluency when reading a text (like an

extremely small font or a bad photocopy) have been discussed (e.g., Diemand-Yauman, Oppenheimer, & Vaughan, 2011; Eitel & Kühl, 2016; Pieger, Mengelkamp, & Bannert, 2016), but discarded (R. A. Bjork & Yue, 2016). Different wording could also act as a desirable difficulty when it triggers elaboration processes on the meaning of the different words. In the context of cognitive load theory, researchers even assume that redundancy – where the use of the same wording in this case would be redundant - impedes learning (Kalyuga & Sweller, 2014).

2.2 Alternative Prediction 2: Different Wording Impedes Learning

There is good reason to assume that in contrast to the results of Jucks and Paus (2013), in individual learning different wording impedes learning. In collaborative learning, it is possible to share processing demands with a learning partner. F. Kirschner et al. (2009a) argue that learning partners that pool their cognitive resources during collaborative learning can reduce the cognitive load for single learners (cf. P. A. Kirschner, Sweller, F. Kirschner, & Zambrano R., 2018). However, this advantage is deprecated by the costs that arise through the necessity of coordinating learning partners (cf. Ringelmann, 1913). F. Kirschner et al. (2009a) argue that the advantage exceeds the costs only in complex tasks. Indeed, they found that collaborative learners outperformed individual learners in a complex task in the domain of biology (F. Kirschner, Paas, & P. A. Kirschner, 2009b).

Different wording creates a certain processing demand, which can be shared in collaborative learning, but not in individual learning. In the case of the same wording being used, individual readers can map prior information and incoming information onto each other immediately. In the case of different wording, the readers first have to make the connection that the different words are referring to the same concept before mapping can occur. This might be an easy and almost automatic translation in the case of high prior

knowledge, but in the case of low prior knowledge, it might fail. Therefore, mapping processes and corroboration are supposed to be easier if the documents use the same wording than if they use different wording (cf. Schoor & Artelt, 2015).

There is first evidence for these considerations: Redundant information in different wording takes longer to be processed than redundant information in the same wording (Schoor & Artelt, 2015). Readers recognize information with a higher chance as conflicting across documents if the information is provided using the same wording, compared to information provided using different wording (Schoor & Artelt, 2016). With regard to single texts, Epstein, Glenberg, and Bradley (1984) found that different wording made it more difficult to recognize conflicting information.

2.3 Alternative Prediction 3: Different Wording is Beneficial For a Subgroup Only

There are hints that difficulties like different wording might be desirable, and therefore beneficial, for some learners. In the context of cognitive load theory, there is an expertise reversal effect where the addition of redundant text to a picture is beneficial for learners with low prior knowledge but detrimental for learners with high prior knowledge (Kalyuga, Chandler, & Sweller, 1998). Similarly, learners with low prior knowledge benefit from highly coherent text, while learners with high prior knowledge benefit from a text with low coherence (McNamara, Kintsch, Songer, & Kintsch, 1996; McNamara & Kintsch, 1996). However, different wording should not be difficult for learners with high prior knowledge because their prior knowledge results in an automatic mapping of the different words onto each other. Therefore, no elaboration on the terms is necessary and different wording cannot be considered a *desirable difficulty* for high prior knowledge learners either. While different wording might be harmful for low prior knowledge learners, no effect is expected for high prior knowledge learners.

Not only prior knowledge can moderate design effects, like desirable difficulties, but motivational factors as well. In her cognitive-affective theory of learning with media, Moreno (2006) suggests that next to the restrictions of the cognitive system, motivation and affect influence learning by influencing cognitive engagement. Similarly, Schnotz, Fries, and Horz (2009) argue that motivation can influence working memory capacity to the extent that a given learner has more working memory capacity available when she or he is highly motivated, compared to when he or she is only minimally motivated. Research on positive effects of seductive details (e.g., Park, Flowerday, & Brünken, 2015) seem to support these assumptions.

There is also evidence that collaborative learning itself is more motivating than individual learning (e.g., Dolmans & Schmidt, 2006; Ginsburg-Block & Fantuzzo, 1998; Hänze & Berger, 2007; Nichols, 1996). Moreover, when the learning material is distributed like in Jucks and Paus' (2013) study, each learning partner gets the role of expert for part of the content. Creating such positive interdependence (Deutsch, 1949; Johnson & Johnson, 1992) is a key feature in several motivational approaches for fostering collaborative learning (e.g., Aronson, Blaney, Stephan, Silkes, & Snapp, 1978; cf. Slavin, 1996). Therefore, collaborative learning might motivate the learners to discuss the meaning of different words and to build a common ground (Clark & Brennan, 1991; see Jucks & Paus, 2013, for further discussion). In individual learning, in contrast, the learner lacks the (potentially) motivating influence of a learning partner. In the case of low motivation, the learner might ignore the different wording, similar to how learners often ignore conflicting information (Otero & Campanario, 1990; Stadtler & Bromme, 2014), resulting in poor learning outcomes. Therefore, is an open question whether the beneficial effects of different wording in multiple documents in collaborative learning can be replicated in an individual learning situation or whether different wording is only beneficial for highly motivated individual learners.

To sum up, the literature suggests several alternative predictions on what type of effect different wording would have on learning with multiple documents in an individual learning situation. The first prediction is that different wording would act as a desirable difficulty and therefore foster learning. This prediction is also in line with earlier results on the effect of wording in collaborative learning. However, in collaborative learning, the learners can pool their cognitive resources in order to deal with high processing demands. Therefore, the second alternative prediction is that different wording in individual learning would impede mapping and comparison processes and therefore hinder learning. The third prediction is that in an individual learning situation there is an aptitude-treatment interaction such that different wording would act as a desirable difficulty only for a subgroup of learners. It was argued that this subgroup would not be characterized by high prior knowledge, because for high prior knowledge learners, different wording would not be difficult. Rather it was argued that different wording would be a desirable difficulty for the subgroup that is highly motivated.

This prediction is also in line with earlier results on the effect of wording in collaborative learning because of the differences between individual and collaborative learning with regard to motivation and the possibilities to share processing demands: In collaborative learning, motivation might be higher and cognitive load lower due to the learning situation, bringing the beneficial effects of different wording to bear. In individual learning, the learner her- or himself has to muster the necessary motivation, which might be only the case for a subgroup of learners.

3 Motivation as Moderator

We have discussed motivation in several paragraphs: as a factor that might enhance working memory capacity (Schnotz et al., 2009), as a factor that might counteract cognitive

load (Moreno, 2006), and as a factor that is influenced by the learning situation (e.g., Dolmans & Schmidt, 2006; Ginsburg-Block & Fantuzzo, 1998; Hänze & Berger, 2007; Nichols, 1996). All in all, this leads to the assumption that motivation might moderate the effects of wording on learning.

Following Moreno's (2006) line of thinking, we suppose that it is the motivational state during learning (e.g., Rheinberg, Vollmeyer, & Rollett, 2000) that is important in the present context. The motivational state during learning mediates the effects that initial motivation (before learning) has on the learning outcome (Rheinberg et al., 2000). According to the expectancy-value model by Eccles and Wigfield (2002), the expectation of success and the subjective task value influence achievement-related choices and performance. The subjective task value comprises, amongst others, interest-enjoyment value and attainment value. There is a vast amount of empirical research that supports these assumptions (e.g., Pajares, 1996; Pintrich, 2003; Schiefele, 1999; Schiefele & Schreyer, 1994; Schoor, 2016). Some of this research also provides support for the motivational state of these variables (e.g., Vollmeyer & Rheinberg, 2000, 2006). This is why we used the students' level of motivational state with regard to their expectation of success and the subjective task value for the present study.

4 The Present Study

The aim of the present study was to test whether the effect found by Jucks and Paus (2013), namely that different wording is beneficial for (collaborative) learning, can be replicated on individual learners with low prior knowledge. When learning from multiple documents, it is necessary to compare (corroborate) and relate (map) information across documents. There are several alternative predictions on how different wording could affect this: It could foster learning, impede learning, or foster learning only for a subgroup of learners. Given the earlier results on collaborative learning (Jucks & Paus, 2013) and given

that collaborative learning might lead to reduced cognitive load and higher motivation than individual learning, we expected that different wording would also be beneficial for individual learning. However, we also expected that this effect was limited to participants with high motivation in terms of motivational state during learning.

This was expected because we assumed that different wording would be a desirable difficulty for learners with low prior knowledge if the learners took the chance of elaborating on the meaning of the different words, but that different wording would also lead to a higher cognitive load than the same wording, since comparison and mapping processes are necessary. In contrast to collaborative learning, the learners would not be able to divide the cognitive load during individual learning. Moreover, the individual learning situation might not be as motivating as collaborative learning. Since motivation during learning can partly compensate for high cognitive load, we assumed that the motivational state would moderate effects of the wording.

For learners with high prior knowledge, the choice of wording was supposed to have no effect, as this type of learner automatically relates the meaning of different words. For this reason, we chose to research only learners with low prior knowledge. However, also in low prior knowledge there might be variance that could influence the results, which is why we chose to control for prior knowledge. Additionally, we chose to control for epistemic beliefs, since they are also known to influence individual MDC.

These assumptions led to the following hypotheses:

H1. Controlling for prior knowledge and epistemic beliefs, different wording (as compared to the same wording) leads to higher cognitive load.

H2. Controlling for prior knowledge and epistemic beliefs, different wording is beneficial for participants with a higher motivational state with regard to their a) performance in

the learning task, b) understanding of specialist vocabulary, and c) knowledge about concepts.

5 Method

5.1 Sample, Design, and Procedure

The participants consisted of 100 university students not majoring in psychology or medicine (in order to avoid effects of prior knowledge on the topic of the study – depression). A power analysis with G*Power 3.1.4 (Faul, Erdfelder, Lang, & Buchner, 2007) revealed that this sample size was large enough to detect a medium wording effect (as it was reported by Jucks & Paus, 2013) qualified by motivation with a power of 70%. For the sake of simplifying the power analysis, we assumed that the sample was divided with regard to their motivation by means of a median split and that a 2x2 ANOVA was calculated.

The largest proportion of the participants (32%) were enrolled in teacher education. The participants received a gift certificate of 10 € from a large online seller as reimbursement. The participants were 19 to 36 years old (M = 21.6, SD = 2.73, 47% female). On average, they were in their 5th semester of studies (M = 5.0, SD = 3.26, n = 92; 8 students did not report their semester).

The design was a two-group randomized experiment in which wording was varied (the same vs. different). The participants were randomly assigned to the conditions, resulting in 50 participants per condition. The participants in the different conditions did not differ significantly with regard to age (t_{97} = -.24, p = .81), gender (χ^2= 0, df = 1, p = 1) or semester of study (t_{90} = .38, p = .70).

The participants were tested in groups in an empty room at the university. The study was advertised on-campus and all who wished to participate were invited to come to this room, as long as they met the inclusion criteria. Testing was done until the planned sample size of 100

was reached. All material was presented on paper. The participants first completed the pre-test of prior knowledge and two questionnaires on epistemic beliefs. Then, they received two text documents. After having read them, they were provided with a learning task and additional sheets of paper to write down their answer. The participants were allowed as much time as they wanted. After they had finished the task, the text documents, the learning task and answers were collected and the participants had to fill in the questionnaires on motivational state and cognitive load, a knowledge post-test, and some questions on demographics[1]. Afterwards, they were thanked and received their reimbursement. The test session took about one to one and a half hours.

5.2 Material and Instruments

5.2.1 Knowledge Pre-Test

As a measure of prior knowledge, the participants were asked in an open-ended question to list and explain all central concepts about depression. We chose this measure in order to avoid lexical priming and prevent the participants from searching for the correct answers in the texts that were to be read during the study. The second author counted the number of different correct concepts. The first author independently counted the concepts for 11 participants. The inter-rater reliability was very high (ICC = .97). Disagreement was solved through discussion.

The participants also had to rate their prior knowledge about depression on a 5-point Likert scale, and the number of words they wrote in the open-ended question was calculated using the Microsoft Word® count function. These two measures were only used to test for

[1] In addition, they repeated the questionnaires on epistemic beliefs. This second measurement will not be analyzed in the present study.

group differences in order to check the randomization. All three measures were the same as Jucks and Paus (2013) used in their study.

5.2.2 Multiple Text Documents

The text documents used in the present study were two texts about depression taken from two popular German magazines (www.stern.de and www.welt.de), which Jucks and Paus (2013) used in their study. Jucks and Paus (2013) had adjusted these texts with regard to the wording. For each of 19 different concepts, they found two synonyms (either a Germanic or a Greco-Roman term, see Appendix A). They compiled these terms into two different wording lists for the 19 concepts balancing the Germanic or Greco-Roman origin of the term. Using either one of the wording lists, they created two different versions for each of the two texts. The texts were each about 1260 words long and were pre-tested by Jucks and Paus (2013). They report that the texts did not differ with regard to readability, emotional state, perceived text comprehensibility, familiarity, and interest for the topic. More information about the texts can be found in Appendix A.

In the present study, the participants received both texts, depending on the condition, either in the same wording or with different wording (i.e. drawing either on the same wording list or on different wording lists for creating the two texts). The wording list used for each of the texts was balanced across participants and within conditions.

5.2.3 Learning Task

As a learning task, the participants read a short case study about a man suffering from early signs of depression. In a first subtask, the participants had to write a diagnosis and justify it. In the second subtask, the participants were asked to discuss advantages and disadvantages of the different ways of treating the problem and to give advice. This was the

task that the participants in the Jucks and Paus (2013) study had to solve collaboratively. The learning task was supposed to guide the learners' learning with the texts, which is why they had the texts available while working on the learning task.

The solution for the learning task was used as an indicator of learning with multiple documents. We used both a quantity measure (number of words) and a quality measure (overall quality rating). The quantity measure was included because a longer answer often has a higher chance of receiving a better quality rating than a short answer. However, short answers can also be of high quality, which is why the quality measure was included.

The overall rating of the quality of the learning task solution was based on the quality rating by Jucks and Paus (2013). A participant only listing one or several treatment options without justification received one point. Two points were awarded for a well-grounded treatment recommendation. Three points were awarded for a mixed approach that was well-grounded. The participants could gain four points for a discussion of different treatment options with a reflection of their limitations. Appendix B provides an example for each of these ratings.

The second author rated all solutions. An additional independent rater (the first author) rated the solutions from 10 participants. The inter-rater agreement for this scoring was ICC = .97. Disagreement was solved through discussion.

5.2.4 Knowledge Post-Test

The knowledge post-test consisted of three parts: (1) the self-rating of the participants' post-knowledge about depression on a 5-point Likert scale, which was not used for the present analysis due to its subjectivity, (2) a test for the understanding of specialist vocabulary (cloze test), and (3) two open-ended questions on knowledge about concepts. All three measures were the same as in Jucks and Paus (2013). In the cloze test for

understanding of specialist vocabulary, the participants had to fill in 21 gaps in a text of about 285 words. Jucks and Paus (2013) created the gaps through the rational deletion procedure of content words. For an example, see Appendix C.

For each gap we created a code which indicated whether the answer was correct or not. The number of correctly filled gaps was summed up. The second author coded all data. The first author independently coded 10% of the data. The inter-rater agreement on this measure was ICC = .99. Disagreement was solved by discussion. We built the same scale as Jucks and Paus (2013) and therefore excluded the same 3 items that Jucks and Paus (2013) excluded (Items 3, 5, 6). Like in their study, these items were also too easy in our study (mean correct solution rate: ≥ 90%). Therefore, the maximum possible score on the cloze test was 18. The internal consistency of this scale was McDonald's ω = .85[2]. Since Jucks and Paus (2013) had used two subscales and had found differential effects, the same two subscales were built in the present study. The first subscale measured the general understanding of the topic (11 items; McDonald's ω = .84), while the second subscale (7 items; McDonald's ω = .95) was supposed to capture specific knowledge about the manipulated concepts.

Knowledge about concepts was tested by asking two open-ended questions on the depressive syndrome. The first one asked for the definition and diagnosis of depression. The second question asked participants to list and describe treatment options for depression. Note that the open-ended questions were quite similar to the questions in the learning task. However, in contrast to the learning task, the participants had to answer the knowledge post-test without having the texts available. Therefore, the open-ended questions in the post-test tested the participants' memory on the concepts they had worked with during the learning

[2] McDonald's ω is an alternative to Cronbach's α that accounts for many problems of Cronbach's α (cf. McNeish, 2017). It was calculated as ω total with the R package *psych* (Revelle, 2016).

task. The scores in the learning task and the knowledge post-test correlated with $r = .34$ ($p < .001$).

We coded the open-ended questions according to the coding scheme by Jucks and Paus (2013), which we adjusted to fit the answers the participants provided in the present study. The points the participants received are listed in table 1. The maximum score for the first question was 7 and 12 for the second question. The second author coded all data. The first author independently coded 10% of the data, and their inter-rater agreement was ICC = .95 for the first question and ICC = .88 for the second question. Disagreement was solved by discussion. For further analyses, we used the sum score of both questions.

5.2.5 Motivational State

Motivational state was measured by means of six items. Two of them captured the *expectation of success* and four of them captured the *subjective task value*. The two items for *expectation of success* stem from the on-line motivation questionnaire by Vollmeyer and Rheinberg (2003). The four items capturing the *subjective task value* were the four items that Schoor (2010) used to assess the learning partner's motivation. These items are based on the subjective task value item from Vollmeyer and Rheinberg (2003). They were significantly related to the learning partner's self-reported motivation (measured by the on-line motivation questionnaire by Vollmeyer & Rheinberg, 2003, $r = .40$, $p < .001$) and to his/her learning success ($r = .18$, $p < .01$) in Schoor's (2010) study. Appendix D displays all items.

The participants had to rate the items on a 5-point Likert scale. A confirmatory factor analysis showed that the two facets of the motivational state do not build a unidimensional factor. Since the two items for *expectation of success* did not have a high intercorrelation ($r = .37$), this facet of motivational state was dropped completely. The four items for subjective task value had a good internal consistency of McDonald's $\omega = .82$.

5.2.6 Cognitive Load

Cognitive load was measured by means of the NASA-TLX (Hart & Staveland, 1988) and the item by Paas (1992). The NASA-TLX required the rating of the task's demands on four dimensions (i.e., mental demand, temporal demand, effort, frustration). The participants had to rate the items on a 20-point Likert scale. The items had an internal consistency of McDonald's $\omega = .82$. The item by Paas (1992) measures the mental effort on a 9-point Likert scale. The NASA-TLX and the Paas (1992) item had a correlation of $r = .40$ ($p < .001$).

5.2.7 Epistemic Beliefs

In order to measure epistemic beliefs, we used both the CAEB (Stahl & Bromme, 2007) and the DEBQ (Hofer, 2000). For both measures, the participants were asked to rate items on the topic of depression (cf. Jucks & Paus, 2013). The CAEB measures connotative aspects of epistemic beliefs with 17 pairs of adjectives (e.g., dynamic-static or objective-subjective). The participants rated these pairs on a 7-point scale. The CAEB has two factors: *texture* (beliefs about the structuredness of knowledge; 10 items; McDonald's $\omega = .83$) and *variability* (beliefs about the stability of knowledge over time; 7 items; McDonald's $\omega = .74$). The items were coded such that high values refer to a high belief in the unstructuredness of knowledge (texture dimension) and the variability of knowledge (variability dimension).

The DEBQ measures four dimensions of epistemic beliefs: *certainty/simplicity* (8 items; McDonald's $\omega = .68$), *justification: personal* (4 items; McDonald's $\omega = .79$), *source of knowledge: authority* (4 items; McDonald's $\omega = .49$), and *attainability* (2 items; McDonald's $\omega = .21$). Due to the poor consistencies of the other scales, we used only the scale

justification: personal for further analyses. The items were coded such that high values refer to a high belief in the justifiability of knowledge through personal experience.

6 Results

First, we checked whether the participants in the two conditions were comparable with regard to prior knowledge and epistemic beliefs. T-tests showed that there were no significant differences between conditions in these measures (see Table 2). Descriptive data of all variables can be found in Table 2, bivariate correlations are displayed in Table 3. With regard to prior knowledge, the overall number of concepts participants were able to come up with was very low (M = 5.6; SD = 3.09; min = 0; max = 17; mode = 4; Mdn = 5). The inspection of the bivariate correlations (Table 3) revealed that there were only as many significant correlations for epistemic beliefs as could have been expected by chance. Overall, they only correlated a small amount with the other variables. Therefore, and for reasons of parsimony, we did not include epistemic beliefs as control variables, but only prior knowledge.

6.1 Effect of Wording on Cognitive Load

Next, we tested whether different wording leads to a higher cognitive load than the use of the same wording, while controlling for prior knowledge (H1). A MANOVA utilizing the two cognitive load measures (NASA-TLX and Paas item) as dependent variables, wording as the predictor, and prior knowledge (number of concepts) as the covariate revealed no significant effect of wording on cognitive load ($F_{2, 96} = 0.18$, $p > .05$, $\eta_p^2 = .00$). Therefore, Hypothesis 1 had to be rejected. The observed effect was extremely small. The power to detect a medium effect was 94% with the current sample size, 40% for a small effect, and 78% for a small to medium effect (calculated with G*Power 3.1.4.; Faul et al., 2007).

6.2 Effect of Wording and Motivational State on Performance in the Learning Task

In order to test the effect of wording and motivational state on performance in the learning task (H2a), we conducted two hierarchical regression analyses with the criterion variables 1) quality score of the learning task (quality measure) and 2) number of words written in the learning task solution (quantity measure). In each hierarchical regression, we first entered prior knowledge (number of different concepts) as a predictor. Then we added wording and motivational state separately. Finally, we included an interaction term of wording and motivational state. Since two hierarchical regression analyses were conducted, we corrected our results to account for the inflation caused by a Type 1 error (Bonferroni correction).

With regard to the overall quality rating of the learning task solution, only prior knowledge was found to be a significant predictor ($\beta = .39$, $p < .001$). Adding wording, motivational state, and their interaction did not increase the explained variance (see Table 4).

The number of words written in the task solution was predicted significantly by prior knowledge ($\beta = .30$, $p < .01$) and motivational state ($\beta = .32$, $p < .01$) but not by the wording ($\beta = .14$, $p > .05$). Adding the interaction of wording and motivational state also did not increase the explained variance (see Table 5).

Therefore, we had to reject Hypothesis 2a. The assumed interaction effect of wording and motivational state was not significant. Only prior knowledge and motivational state predicted the quantity of performance in the learning task, and only prior knowledge its quality. With the current sample size, it was possible to detect a medium effect with a power of 94% and a small effect with a power of 39%. A power of 79% was given for the detection of a small to medium effect (calculated with G*Power 3.1.4.; Faul et al., 2007).

6.3 Effect of Wording and Motivational State on Understanding of Specialist Vocabulary

In order to test the effect of wording and motivational state on the understanding of specialist vocabulary (H2b), we conducted three hierarchical regression analyses for 1) the sum score in the cloze test, 2) the score on the subscale *general*, and 3) the score on the subscale *specific* as the criterion variable. In each hierarchical regression, we first entered prior knowledge (number of different concepts) as our predictor. Then we added wording and motivational state separately. Finally, we included an interaction term of wording and motivational state. Since three hierarchical regression analyses were conducted, we corrected for the inflation of the error of Type 1 (Bonferroni correction).

The results (see Table 6) show that only prior knowledge significantly predicted the understanding of specialist vocabulary in the post test, and this only for the subscale *specific* ($\beta = .26$, $p < .05$). With the current sample size, it was possible to detect a medium effect with a power of 92% and a small effect with a power of 34%. A power of 74% was given for the detection of a small to medium effect (calculated with G*Power 3.1.4.; Faul et al., 2007).

6.4 Effect of Wording and Motivational State on Knowledge About Concepts

In order to test the effect of wording and motivational state on knowledge about concepts (H2c), we conducted a hierarchical regression analysis for the sum score on the open-ended post knowledge questions as the criterion variable. Again, we first entered prior knowledge (number of different concepts) as a predictor. Then we added wording and motivational state separately. Finally, we included an interaction term of wording and motivational state.

We found a significant interaction effect of wording and motivational state ($\beta = .35$, $p < .05$, see Table 7). This interaction is plotted in Figure 1. It can be seen that different wording

was beneficial for highly motivated students, while the wording made no difference for less motivated students. Therefore, we were able to confirm Hypothesis 2c.

7 Discussion

In the present study, we tested whether the positive effect of different wording found in a prior study for collaborative learning (Jucks & Paus, 2013) could be replicated on individual learners with low prior knowledge. Based on earlier research, three alternative predictions could be made for individual learning: 1. that different wording would be beneficial for learning, 2. that different wording would impede learning, or 3. that different wording would be beneficial for a subgroup of learners (namely the highly motivated learners) only. Assuming that different wording would make higher processing demands than the same wording in individual low prior knowledge learners, but that different wording nevertheless could be beneficial for learning if the learners invested the necessary effort, we analyzed whether different wording across multiple documents would lead to higher cognitive load. Moreover, we analyzed whether learners with higher motivational state would benefit from different wording with regard to their performance in the learning task, their understanding of specialist vocabulary, and their knowledge about concepts. We found, as expected, that different wording only enhanced knowledge about concepts for highly motivated learners. However, there was no effect found for wording on cognitive load, and no effect for wording and motivational state on the performance in the learning task and understanding of specialist vocabulary. The quality of the learning task solution was only predicted by prior knowledge, while the quantity (number of words) was predicted by prior knowledge and motivational state. Only prior knowledge was a significant predictor when considering the understanding of specialist vocabulary, but only for the subscale *specific*.

The result that the use of different wording is only beneficial for students with high motivation is consistent with the predictions that we had derived from the theory on multiple document comprehension (e.g., Britt & Rouet, 2012), cognitive load (e.g., Moreno, 2006; Schnotz et al., 2009), and collaborative learning (e.g., F. Kirschner et al., 2009a; Slavin, 1996). In order to relate information stemming from different documents, the learners have to map this information onto each other (Schoor & Artelt, 2015; Van Meter & Firetto, 2008; Wineburg, 1991). This mapping is easier when the texts use the same wording (cf. Schoor & Artelt, 2015). However, if the students invested the effort in relating differently worded information, they might benefit from the use of different wording; the latter could act as a desirable difficulty. The prediction that highly motivated students would benefit from different wording proved to be the case in the present study.

The present results largely correspond to the results of Jucks and Paus (2013) as well. Apart from the different settings of individual and collaborative learning, the main difference between the two studies was that, in the present study, the positive effect of the different wording was only present for highly motivated students. Jucks and Paus (2013) did not measure the motivation of their students. Therefore, it could be the case that most of their participants were highly motivated. This could also be due to the collaborative setting. Collaborative learning itself can be motivating for students (e.g., Dolmans & Schmidt, 2006; Ginsburg-Block & Fantuzzo, 1998; Hänze & Berger, 2007; Nichols, 1996). Moreover, the fact that both learning partners of a dyad had different texts at their disposal could have led to positive interdependence (Johnson & Johnson, 1992) and thereby to a higher motivation and a higher individual responsibility for the learning success of both partners (cf. Slavin, 1992). This could have led to a higher motivation in the collaborative setting to elaborate on the meaning of the different wording.

In slight contrast to the results by Jucks and Paus (2013), an effect caused by the wording was found only for knowledge about concepts. We could not find an effect caused by the wording for the understanding of specialist vocabulary (i.e., the cloze test and its subscales). However, Jucks and Paus (2013) found the effect of the wording on the cloze test performance only for the specific knowledge subscale and not for the general knowledge subscale. The difference in results between the two studies might be due to the fact that in the present study each participant received both texts, while in the Jucks and Paus (2013) study each participant only had one text available. In order to get all possible information, the participants in the Jucks and Paus (2013) study had to talk to their learning partner, while all the information was directly available to the participants in the present study. In the Jucks and Paus (2013) study, the learning partners in the different wording condition talked more about comprehension problems than the learning partners in the same wording condition. Therefore, they probably also clarified more unclear issues in the different wording condition than in the same wording condition. This might have led to a better understanding of the specialist vocabulary that was not initially available to both learning partners. This might have resulted in the observed difference in the performance on the *specific* subscale of the cloze test. Since the present study made all information available to all participants, no similar effect of the condition on the cloze test performance could occur.

In contrast to our expectations, we could not confirm that the same wording would be easier in terms of cognitive load. The highly motivated students who learned using texts with different wording did not perform better on the learning task either. This raises questions with regard to the assumed theoretical mechanism of how different wording leads to a beneficial learning outcome. An alternative explanation is that the use of different wording might make connecting information to the student’s prior knowledge easier, since the use of different wording provides more opportunities to relate the new information to prior

knowledge than the use of the same wording does. While students worked on the learning task, the texts were made available. During the post test, the texts were not available. The students had to rely on the information they had previously learned. In the different wording condition, the same concepts were referred to with twice as many words as in the same wording condition. For each concept, there was a double opportunity to “ring a bell” in prior knowledge and therefore to be connected to prior knowledge. When the student had to recall the information without text, this double chance might have caused the effect of the different wording (combined with the motivation to think about the words). Further research is needed in order to test this assumption.

An alternative explanation for the moderating effect found for motivation derives from the fact that motivation was measured after the learning task and not before the learning task. It could be argued that those learners who easily solved the learning task and performed well (due to their better cognitive abilities or better vocabulary, for example) were also those learners who felt more motivated after the learning task because of their performance. In this case, the role of motivation would be a different one. Instead of being a moderator of the effect caused by different wording, motivation would be an outcome in addition to high performance. That is to say, different wording would not be beneficial for highly motivated learners (motivational state as a moderator), but for high-performing learners, the different wording would lead to a higher motivation. The observed correlation of motivational state and the score in the learning task speaks in favor of such an interpretation (see Table 3). However, the lacking correlation of motivational state and the post-test knowledge about concepts and understanding of specialist vocabulary indicates that a potentially higher motivation of learners who perform well in the learning task has no effect on the acquisition of knowledge.

7.1 Limitations

Of course, the current study has its limitations. First, the results are based on the single topic of depression. It thus cannot be taken for granted that research on other content areas leads to the same results. However, we believe that the observed effects are generalizable across different topics because prior knowledge and motivation were controlled for. Differences between topics would also be reflected in differences in prior knowledge and motivation, and these two measures would probably account for much of a potential topic effect. Nevertheless, there might be variables that still could have an influence in certain topics and change the results, for example strong prior beliefs on a topic that are not reflected in a higher prior knowledge (cf. Maier & Richter, 2013, 2014).

Second, the results are limited to the population that the present sample had been drawn from. The participants were university students with low prior knowledge. It can be expected that there will not be substantial wording effects in populations with high prior knowledge because they will already possess the ability to translate easily between the different wording used for the same concept. With regard to populations other than university students, we can expect the results to be different. Both directions are plausible. The effect of different wording could be even more pronounced for non-university learners because they benefit even more from being able to connect the concepts to their prior knowledge. The effect could also be in the opposite direction as well, such that the different wording is so confusing that they are not able to relate the different words for the same concept to each other and understand only half of the texts. Further research is needed in order to clarify this.

7.2 Practical Implications

With the above-mentioned limitations, we can make some careful considerations about practical implications. Different wording across texts might be helpful for some learners, but

not for all. A practical implication would be to only provide highly motivated learners – or learners expected to become highly motivated while learning – with documents that use different wording. For less motivated students, documents that use the same wording would probably be more suitable. Further research needs to be done to determine whether talking about the use of different words for the same concept in class beforehand would make a difference for these students.

8 Conclusion

To our knowledge, the present study provided the first evidence for the effects that wording has in individual learning with multiple documents. It extended prior research on wording in collaborative learning and was able to show comparable effects for individual learning. It also showed the limitations that exist for the found effect for highly motivated students, thereby emphasizing the necessity to take learner characteristics and their interaction with document characteristics (wording) into account. As such, it is a good starting point for further research on the interaction of document and learner characteristics in multiple document comprehension.

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33*(2–3), 131-152. doi: 10.1016/S0360-1315(99)00029-9

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183-198. doi: 10.1016/j.learninstruc.2006.03.001

Aronson, E., Blaney, N., Stephan, C., Silkes, J., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage.

Barzilai, S., & Strømsø, H. I. (2018). Individual differences in multiple document comprehension. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 99-116). New York: Routledge.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York, NY, US: Worth Publishers.

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, US: The MIT Press.

Bjork, R. A., & Yue, C. L. (2016). Commentary: Is disfluency desirable? *Metacognition and Learning, 11*(1), 133-137. doi: 10.1007/s11409-016-9156-8

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist, 46*(1), 48-70. doi: 10.1080/00461520.2011.538647

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J.-F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. Van den Broek (Eds.), *Narrative, comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209-233). Mahwah, NJ: Erlbaum.

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276-314). New York: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: American Psychological Association.

Deutsch, M. (1949). A theory of co-operation and competition. *Human Relations, 2*(2), 129-152.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold and the italicized: Effects of disfluency on educational outcomes. *Cognition, 118*(1), 111-115. doi: 10.1016/j.cognition.2010.09.012

Dolmans, D. H. J. M., & Schmidt, H. G. (2006). What do we know about cognitive and motivational effects of small group tutorials in problem-based learning? *Advances in Health Sciences Education, 11*(4), 321-336.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109-132.

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

Epstein, W., Glenberg, A., & Bradley, M. (1984). Coactivation and comprehension: Contribution of text variables to the illusion of knowing. *Memory & Cognition, 12*(4), 355-360. doi: 10.3758/bf03198295

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*(2), 175-191.

Ferguson, L. E. (2015). Epistemic beliefs and their relation to multiple-text comprehension: A Norwegian program of research. *Scandinavian Journal of Educational Research, 59*(6), 731-752. doi: 10.1080/00313831.2014.971863

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes, 23*(3), 265-304. doi: 10.1080/01638539709544994

Ginsburg-Block, M., & Fantuzzo, J. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology, 90*(3), 560-569.

Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and learning from internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly, 47*(4), 356-381. doi: 10.1002/rrq.027

Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and methodological directions. *Educational Psychologist, 43*(3), 142-160. doi: 10.1080/00461520802178458

Hänze, M., & Berger, R. (2007). Cooperative learning, motivational effects, and student characteristics: An experimental study comparing cooperative learning and direct instruction in 12th grade physics classes. *Learning and Instruction, 17*(1), 29-41.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52,* 139-183. doi: 10.1016/S0166-4115(08)62386-9

Hofer, B. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology, 25,* 378-405.

Hofer, B., & Pintrich, P. R. (1997). The development of epistemological theories: beliefs about knowledge and their relation to learning. *Review of Educational Research, 67*(1), 88-140.

Johnson, D. W., & Johnson, R. T. (1992). Positive interdependence: Key to effective cooperation. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups. The theoretical anatomy of group learning* (pp. 174-199). New York: Cambridge University Press.

Jucks, R., & Paus, E. (2013). Different words for the same concept: Learning collaboratively from multiple documents. *Cognition and Instruction, 31*(2), 227-254. doi: 10.1080/07370008.2013.769993

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

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 247-262). New York, NY, US: Cambridge University Press.

King, P., & Kitchener, K. (2004). Reflective Judgement: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist, 39*(1), 5-18.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009a). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review, 21*(1), 31-42. doi: 10.1007/s10648-008-9095-2

Kirschner, F., Paas, F., & Kirschner, P. A. (2009b). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior, 25*(2), 306-314.

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R., J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning, 13*(2), 213-233. doi: 10.1007/s11412-018-9277-y

Kobayashi, K. (2009). The influence of topic knowledge, external strategy use, and college experience on students' comprehension of controversial texts. *Learning and Individual Differences, 19*(1), 130-134. doi: 10.1016/j.lindif.2008.06.001

Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 121-144). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Küpper-Tetzel, C. E. (2014). Understanding the distributed practice effect. *Zeitschrift für Psychologie, 222*(2), 71-81. doi: 10.1027/2151-2604/a000168

Le Bigot, L., & Rouet, J.-F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research, 39*(4), 445-470. doi: 10.1080/10862960701675317

List, A., Stephens, L. A., & Alexander, P. A. (in press). Examining interest throughout multiple text use. *Reading and Writing*. doi: 10.1007/s11145-018-9863-4

Maier, J., & Richter, T. (2013). Text belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction, 31*(2), 151-175. doi: 10.1080/07370008.2013.769997

Maier, J., & Richter, T. (2014). Fostering multiple text comprehension: How metacognitive strategies and motivation moderate the text-belief consistency effect. *Metacognition and Learning, 9*(1), 51-74. doi: 10.1007/s11409-013-9111-x

Maier, J., Richter, T., Nauroth, P., & Gollwitzer, M. (in press). For me or for them: how in-group identification and beliefs influence the comprehension of controversial texts. *Journal of Research in Reading*. doi: 10.1111/1467-9817.12132

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43. doi: 10.1207/s1532690xci1401_1

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*(3), 247-288. doi: 10.1080/01638539609544975

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*. doi: 10.1037/met0000144

Merkt, M., Werner, M., & Wagner, W. (2017). Historical thinking skills and mastery of multiple document tasks. *Learning and Individual Differences, 54*, 135-148. doi: 10.1016/j.lindif.2017.01.021

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149-158.

Nichols, J. D. (1996). The effects of cooperative learning on student achievement and motivation in a high school geometry class. *Contemporary Educational Psychology, 21*(4), 467-476.

Otero, J. C., & Campanario, J. M. (1990). Comprehension evaluation and regulation in learning from science texts. *Journal of Research in Science Teaching, 27*(5), 447-460. doi: 10.1002/tea.3660270505

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429-434.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578.

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior, 44*, 267-278. doi: 10.1016/j.chb.2014.10.061

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99-122). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Pintrich, P. R. (2003). Motivation and classroom learning. In W. Reynolds & G. Miller (Eds.), *Handbook of psychology: Educational psychology* (Vol. 7, pp. 103-122). Hoboken, NJ: Wiley & Sons.

Revelle, W. (2016). *psych: Procedures for Personality and Psychological Research*. Retrieved from https://CRAN.R-project.org/package=psych

Rheinberg, F., Vollmeyer, R., & Rollett, W. (2000). Motivation and action in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 503-529). San Diego: Academic Press.

Ringelmann, M. (1913). Recherches sur les moteurs animé: Travail de l'homme [Research on animate sources of power: The work of man]. *Annales de l'Institut national agronomique: administration, enseignement et recherche, 12*, 1-40.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432-1463. doi: 10.1037/a0037559

Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading, 3*(3), 257-279.

Schiefele, U., & Schreyer, I. (1994). Intrinsische Lernmotivation und Lernen: Ein Überblick zu Ergebnissen der Forschung [Intrinsic motivation to learn and learning. A review of recent research findings]. *Zeitschrift für Pädagogische Psychologie, 8*, 1-13.

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

Schnotz, W., Fries, S., & Horz, H. (2009). Motivational aspects of cognitive load theory. In M. Wosnitza, S. A. Karabenick, A. Efklides, & P. Nenniger (Eds.), *Contemporary motivation research. From global to local perspectives* (pp. 69-96). Göttingen: Hogrefe.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*(3), 498-504. doi: 10.1037/0022-0663.82.3.498

Schoor, C. (2010). *Die Bedeutung von Motivation für Wissenserwerbsprozesse beim computerunterstützten kooperativen Lernen* [The importance of motivation for

knowledge acquisition processes during computer-supported collaborative learning]. Berlin: Logos.

Schoor, C. (2016). Perceived utility of reading - predictor of reading competence? *Learning and Individual Differences, 45,* 151-158. doi: 10.1016/j.lindif.2015.11.024

Schoor, C., & Artelt, C. (2015). *Prozesse der Informationsintegration beim Lesen multipler Dokumente.* [*Processes of information integration during reading of multiple documents*]. Paper presented at the 15. Fachtagung Pädagogische Psychologie, 14.-16.09.15, Kassel, Germany.

Schoor, C., & Artelt, C. (2016). *Conflict detection in multiple documents as a function of kind of conflict, relevance, and wording.* Paper presented at the 26th Annual Meeting of the Society for Text and Discourse, 18.-20.07.2016, Kassel, Germany.

Schoor, C., & Bannert, M. (2013). Aspects of dynamics in motivation: What the cooperative situation and individual motivation contribute to motivation during computer-supported cooperative learning. *Unterrichtswissenschaft, 41*(4), 330-347.

Seufert, T. (2009). Lernen mit multiplen Repräsentationen - Gestaltungs- und Verarbeitungsstrategien [Learning with multiple representations - Design and processing strategies]. In R. Plötzner, T. Leuders, & A. Wichert (Eds.), *Lernchance Computer. Strategien für das Lernen mit digitalen Medienverbünden* (pp. 45-66). Münster: Waxmann.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology, 69*(1), 487-510. doi: 10.1146/annurev-psych-122216-011845

Slavin, R. E. (1992). When and why does cooperative learning increase achievement? Theoretical and empirical perspectives. In R. Hertz-Lazarowitz & N. Miller (Eds.),

Interaction in cooperative groups: the theoretical anatomy of group learning (pp. 145-173). Cambridge: University Press.

Slavin, R. E. (1996). Research on cooperative learning and achievement: what we know, what we need to know. *Contemporary Educational Psychology, 21*, 43-49.

Stadtler, M., & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp & J. L. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379-402). Cambridge, MA: MIT Press.

Stahl, E., & Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction, 17*(6), 773-785. doi: 10.1016/j.learninstruc.2007.09.016

Stang Lund, E., Bråten, I., Brandmo, C., Brante, E. W., & Strømsø, H. I. (in press). Direct and indirect effects of textual and individual factors on source-content integration when reading about a socio-scientific issue. *Reading and Writing*. doi: 10.1007/s11145-018-9868-z

Ulrich, R., Erdfelder, E., Deutsch, R., Strauß, B., Brüggemann, A., Hannover, B., . . . Rief, W. (2016). Inflation von falsch-positiven Befunden in der psychologischen Forschung [Inflationary Incidence of False-Positive Results in Psychological Research]. *Psychologische Rundschau, 67*(3), 163-174. doi: 10.1026/0033-3042/a000296

Van Meter, P., & Firetto, C. (2008). Intertextuality and the study of new literacies: research critique and recommendations. In J. Coiro, M. Knobel, C. Lankshear, & D. J. Leu (Eds.), *Handbook of research on new literacies* (pp. 1079-1092). New York: Erlbaum.

Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction, 10*, 293-309.

Vollmeyer, R., & Rheinberg, F. (2003). Aktuelle Motivation und Motivation im Lernverlauf [Current motivation and motivation in the course of learning]. In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von Motivation und Selbstkonzept* (pp. 281-295). Göttingen: Hogrefe.

Vollmeyer, R., & Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review*, *18*(3), 239-253.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73-87. doi: 10.1037/0022-0663.83.1.73

Wolfe, M. B. W., & Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction*, *23*(4), 467-502. doi: 10.1207/s1532690xci2304_2

Appendix A: Text Materials

Text 1 (http://www.stern.de/gesundheit/depressionen-zurueck-ins-leben-3739184.html) contained

- a short case study
- general information about the prevalence of depression and its causes within metabolism of the brain
- information about studies questioning the effectiveness of antidepressant drugs
- information about side effects of antidepressant drugs
- information about effects of psychotherapy
- information about the importance of the relationship between therapist and patient, based on the introductory case study

An extract of the text can be found in Figure A.1.

Text 2 (https://www.welt.de/gesundheit/psychologie/article2435260/Griff-zur-Pille-hilft-bei-Depressionen-nicht-immer.html) encompassed information about

- the consumption of ataractics in the society
- reasons (like money or risk of suicide) for the widespread use of antidepressant drugs
- the effectiveness of different kinds of psychotherapy
- a new psychotherapy approach (CBASP)
- psychoanalysis in comparison with cognitive behavioral therapy

An extract of the text can be found in Figure A.2.

The two texts contained mainly complementary information about depression, although the gists of the two texts with regard to antidepressant drugs were different with Text 1 being more critical about antidepressant drugs than Text 2.

The manipulated words can be found in Table A.1.

Table A.1. Manipulated words and wording lists with approximate English translation in brackets (Jucks & Paus, 2013, p. 234).

wording list 1	wording list 2
interpersonell (interpersonal)	zwischenmenschlich (between people)
tiefenpsychologisch (depth psychological)	psychoanalytisch (psychoanalytical)
Therapie (therapy)	Behandlung (treatment)
Arznei (medicine)	Medikament (drug)
Selbstmord (self-inflicted death)	Suizid (suicide)
Botenstoff (messenger)	Transmitter (transmitter)
Placebo (placebo)	Scheinheilmittel (dummy pill)
Stimulanzien (stimulants)	Aufputschmittel (pep pills)
andauernd (persistent)	chronisch (chronic)
Auseinandersetzung (face-off)	Konfrontation (confrontation)
Verhaltens- (action-related)	Behavioral (behavioral)
Trauma (trauma)	Schock (shock)
Nervenarzt (neurologist)	Psychiater (psychiatrist)
Anzeichen (sign)	Symptom (symptom)
Stimmungsaufheller (mood enhancers)	Antidepressiva (antidepressants)
Lethargie (lethargy)	Antriebsschwäche (lack of drive)
somatisch (somatic)	körperlich (physical)
Insomnie (insomnia)	Schlafstörung (sleep disorder)

Appendix B: Example of Learning Task Rating

Example for 0 points:

"Mr. M. should let his family doctor recommend him a psychiatrist. He could be suffering from a depression or at least undergo a depressive phase. Mr. M. should discuss the best treatment option with his psychiatrist and decide afterwards. The psychiatrist can give Mr. M. a more qualified recommendation than I after 10 minutes of reading two journalistic texts."

Example for 1 point:

"... If he can't find a reason for his inactivity, he should go to a doctor or psychologist. It is important that he can establish a personal relationship to his psychologist so that both of them can find a joint way to change his behavioral patterns. Since Mr. M. shows only a light to medium depression, I advise against drugs; however, he should go to a regular treatment."

Example for 2 points:

"... Mr. M. could either take ataractics or begin a therapy. In his case, ataractics probably are not reasonable and effective because it has been proven that they only help in the case of severely ill persons. Moreover, they have many negative side effects. Another possibility that Mr. M. might consider is a therapy. He could do a behavior therapy or a psychoanalysis. The psychoanalysis would only be reasonable if his problem is based deeply in his childhood or adolescence. I personally would recommend him a behavior therapy as it could be confirmed that a therapy is more successful than a medical treatment of his problem. In my opinion, he should search for a therapist he can trust. So, he can try to solve his problems without ataractics."

Example for 3 points:

"... A therapy of depression can consist of: - medication: In my opinion especially in severe cases of depression necessary in order to reduce the risk of suicide. Is paid by the health insurance. Unfortunately, there is prejudice and distrust in the society against medication, as it only treats symptoms but not the underlying deep disorder. Purely physical approach. Can have many side effects. – behavior, psychotherapies: Many variants of asomatic treatment alternatives, for example an advancement of the Freudian psychoanalysis. Tackles individually the problem; however, a good relationship between therapist and patient is necessary. Lengthy, but sustainable form of therapy. In case of severe depressions and maybe thoughts about suicide, I'd urgently recommend M. a combination of medication and acknowledged (in most cases paid by the health insurances) psycho / behavior therapy. In case of mild depressions, medication might not be necessary. I'd advise against experimental therapies (art / music therapy), the success is less secure and he would probably have to pay for the treatment costs."

Example for 4 points:

"... Basically, Mr. M. has two options: First, he can try to solve his problem with drugs. Second, he can ask for therapeutic help. The first option has the advantage that moodlifters normally work fast and in case of a diagnosed depression are paid by the health insurance. The disadvantage is that in case of light depressions the drugs are not much more effective than a placebo and that there is the danger that after stopping the intake there is a relapse. With regard to the therapeutic approach there is the cognitive behavioral therapy and the interpersonal therapy. Advantages are that the therapy helps better than drugs in case of light and medium depressions and a relapse is less probable than with drugs. Disadvantage is that the health insurance might not cover it and that a therapist has to be found first who matches

oneself, and in case of a depression you tend to have the impression that nothing is worth the pain and the illness is rather severed than improved. To conclude, I'd recommend Mr. M. to try a therapy and to take the time to find someone of who he has the impression that he can help him. If he still doesn't get it under control, moodlifters could still be used."

Appendix C: Extract of the cloze test

The cloze test in this study was the same as the one in Jucks & Paus (2013). An example sentence with two gaps was the following:

"If success is needed immediately on the _symptom_ level, _behavioral_therapy_ is appropriate." (Jucks & Paus, 2013, p. 239). The German original sentence was the following:

"Müssen rasche Erfolge auf Ebene der _Symptome_ erzielt werden, eignet sich die _Verhaltenstherapie_ besser." (Jucks & Paus, 2013, p. 239)

Appendix D: Motivational State

The following items were used to measure the motivational state (originally in German):

- I was very motivated while working on the task.
- I tried to hand in a good result.
- While working on the task I was very involved.
- I enjoyed working on the task.
- While working on the task I was sure to find the correct solution.
- I always knew how to proceed.

Table 1. Points Awarded in the Post Test on Knowledge about Concepts.

Question	Points	for …
Question 1	1	Explanation / definition of depression
	1	Causes of depression
	Up to 2	Information about the diagnosis of depression
	Up to 3	Symptoms
Question 2	1	Self-help
	1	Help by family and friends
	1	Mentioning psychoanalysis
	1	Explaining psychoanalysis
	1	Mentioning (cognitive) behavior therapy
	1	Explaining (cognitive) behavior therapy
	1	Mentioning interpersonal therapy
	1	Explaining interpersonal therapy
	1	Mentioning CBASP
	1	Explaining CBASP
	0.5	Mentioning medication
	0.5	Explaining mode of action of medication
	0.5	Mentioning side effects of medication
	0.5	Mentioning disadvantages of medication

Table 2. Means, Standard Deviations, Test Statistics for Group Differences, and Effect Sizes.

	same wording condition	different wording condition	d	t	df	p
prior knowledge						
self-rating	2.72 (0.99)	2.70 (1.02)	.02	0.10	98	n.s.
# words	35.46 (27.27)	41.84 (26.40)	-.24	-1.19	98	n.s.
# central concepts	5.29 (2.58)	5.92 (3.52)	-.20	-1.02	98	n.s.
epistemic beliefs						
CAEB texture	3.58 (0.85)	3.77 (0.96)	-.20	-1.02	98	n.s.
CAEB variability	3.11 (0.62)	3.30 (0.72)	-.28	-1.38	98	n.s.
DEBQ justification	3.36 (0.70)	3.30 (0.73)	.08	0.38	98	n.s.
learning task						
# words	188.76 (75.13)	220.66 (85.69)	-.40	-1.98	98	.05
score	2.55 (1.17)	2.89 (1.14)	-.29	-1.47	98	n.s.
post knowledge						
open-ended questions	5.83 (2.01)	5.87 (2.73)	-.02	-0.08	90.00	n.s.
cloze test	8.23 (3.00)	8.68 (3.48)	-.14	-0.69	98	n.s.
cloze test subscale general	5.84 (1.80)	6.00 (1.82)	-.09	-0.44	98	n.s.
cloze test subscale specific	2.39 (2.11)	2.68 (2.23)	-.14	-0.69	98	n.s.
motivational state	3.71 (0.61)	3.82 (0.61)	-.17	-0.86	98	n.s.

cognitive load						
NASA-TLX	9.52 (2.91)	9.32 (2.87)	.07	0.35	98	n.s.
Paas item	5.46 (1.09)	5.30 (1.47)	.12	0.62	90.32	n.s.

Note. In case of unequal variances, correction of df was applied.

Table 3. Bivariate Correlations.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
prior knowledge														
1. self-rating														
2. # words	.24*													
3. # central concepts	.25*	.63***												
epistemic beliefs														
4. CAEB texture	.12	.05	.24*											
5. CAEB variability	.15	.03	.21*	.42***										
6. DEBQ justification	-.08	-.01	-.03	-.22*	-.30**									
learning task														
7. # words	-.01	.46***	.34***	.13	.02	-.08								
8. score	-.12	.31**	.39***	-.02	-.05	.12	.55***							
post knowledge														
9. open-ended questions	-.01	.09	.19	.09	-.02	-.10	.30**	.34***						

10. cloze test overall	-.03	.07	.21*	.20*	.04	-.11	.21*	.13	.53***				
11. cloze test: general	-.02	-.03	.08	.17	.09	.01	.13	.13	.34***	.80***			
12. cloze test: specific	-.03	.13	.26*	.16	-.02	-.17	.21*	.09	.53***	.86***	.37***		
13. motivational state	.21*	.19	.07	.17	-.03	-.04	.35***	.20*	.02	.05	.02	.06	
cognitive load													
14. NASA-TLX	-.08	.05	.11	-.02	.05	.14	.08	.26**	-.06	-.15	-.07	-.17	-.06
15. Paas item	.11	.11	-.08	.00	.05	-.13	.18	.15	.09	-.07	-.08	-.04	.17	.40***

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 4. Standardized Coefficients for the Regression of Performance in the Learning Task (Overall Quality Rating of the Learning Task Solution) on Prior Knowledge (Number of Different Concepts), Wording, Motivational State, and their Interaction.

	Model 1	Model 2	Model 3	Model 4
Constant	***	***	***	***
prior knowledge (# concepts)	.39***	.38***	.37***	.37***
wording		.11	.10	.10
motivational state			.17	.16
wording*motivational state				.01
R^2	.15	.16	.19	.19
F	17.44***	9.43***	7.48***	5.56***
ΔR^2		.01	.03	.00
ΔF		1.35	3.17	0.01

Note. Wording was coded 0/1 (same/different wording). Motivational state was z-standardized. It was corrected for the inflation of Type 1 error (Bonferroni correction).

*** $p < .001$; ** $p < .01$; * $p < .05$

Table 5. Standardized Coefficients for the Regression of Performance in the Learning Task (Quantity Measure: Number of Words in the Learning Task Solution) on Prior Knowledge (Number of Different Concepts), Wording, Motivational State, and their Interaction.

	Model 1	Model 2	Model 3	Model 4
Constant	***	***	***	***
prior knowledge (# concepts)	.34**	.32**	.30**	.29**
wording		.16	.14	.14
motivational state			.32**	.22
wording*motivational state				.14
R^2	.11	.14	.24	.25
F	12.53**	7.88**	10.14***	7.91***
ΔR^2		.03	.10	.01
ΔF		2.97	12.77**	1.15

Note. Wording was coded 0/1 (same/different wording). Motivational state was z-standardized. It was corrected for the inflation of Type 1 error (Bonferroni correction).

*** $p < .001$; ** $p < .01$; * $p < .05$

Table 6. Standardized Coefficients for the Regression of the Understanding of Specialist Vocabulary in the Posttest (Cloze Test Score Overall and Subscales General and Specific) on Prior Knowledge (Number of Different Concepts), Wording, Motivational State, and their Interaction.

	Model 1			Model 2			Model 3			Model 4		
	overall	general	specific	overall	general	specific	overall	general	specific	overall	general	specific
Constant	***	***	**	***	***	**	***	***	**	***	***	**
prior knowledge (# concepts)	.21	.08	.26*	.21	.07	.26*	.21	.07	.26*	.19	.06	.26*
wording				.05	.04	.05	.05	.04	.05	.05	.04	.05
motivational state							.03	.01	.01	-.08	-.15	-.06
wording*motivational state										.16	.22	.10
R^2	.05	.01	.07	.05	.01	.07	.05	.01	.07	.06	.03	.07
F	2.10	0.57	7.11*	1.70	0.35	3.62	1.40	0.23	2.45	1.44	0.78	1.86
ΔR^2				.00	.00	.00	.00	.00	.00	.01	.03	.00
ΔF				0.24	0.13	0.19	0.10	0.01	0.18	1.25	2.40	0.14

Note. Wording was coded 0/1 (same/different wording). Motivational state was z-standardized. It was corrected for the inflation of Type 1 error (Bonferroni correction).

*** $p < .001$; ** $p < .01$; * $p < .05$

Table 7. Standardized Coefficients for the Regression of Post Knowledge about Concepts (Open-Ended Questions Score) on Prior Knowledge (Number of Different Concepts), Wording, Motivational State, and their Interaction.

	Model 1	Model 2	Model 3	Model 4
Constant	***	***	***	***
prior knowledge (# concepts)	.19	.19	.19	.16
wording		-.01	-.01	-.01
motivational state			.00	-.25
wording*motivational state				.35*
R^2	.03	.03	.03	.10
F	3.84	1.73	1.14	2.52*
ΔR^2		.00	.00	.06
ΔF		0.01	0.00	6.45*

Note. Wording was coded 0/1 (same/different wording). Motivational state was z-standardized.

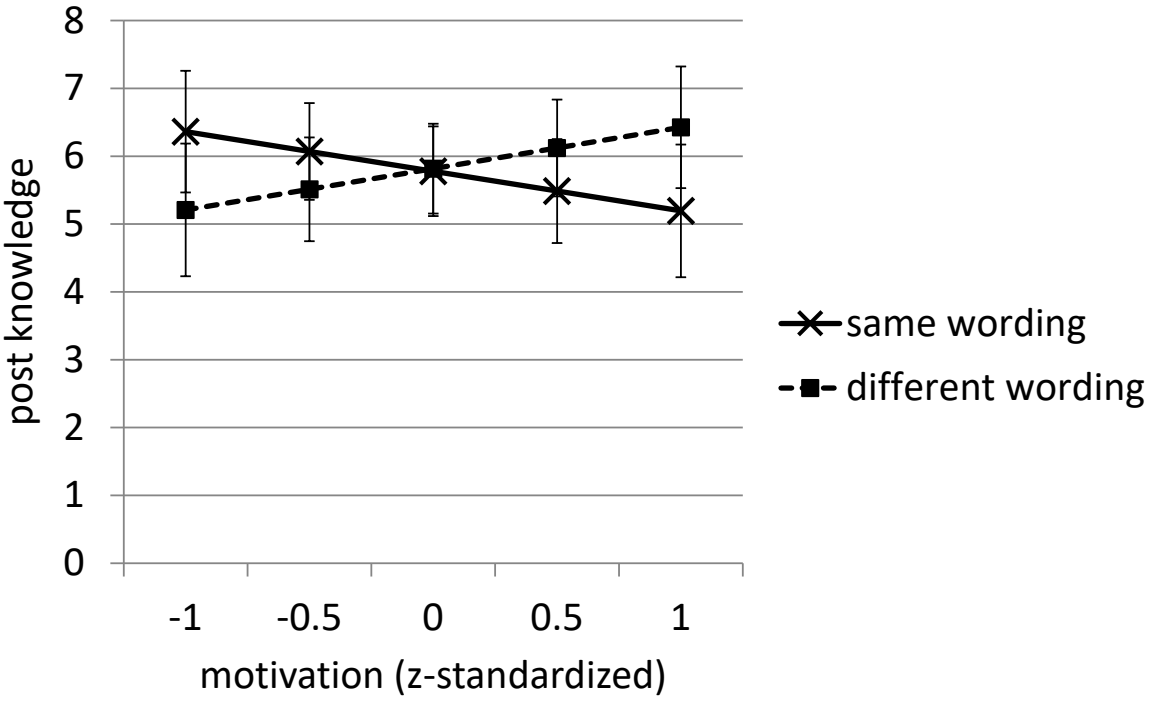
*** $p < .001$; ** $p < .01$; * $p < .05$

Figure Captions

Figure 1. Illustration of the Interaction Effect of Wording and Motivational State on Post Knowledge about Concepts (Open-Ended Questions Score) with 95% Confidence Interval.

Figure A1. Extract from Text 1 (http://www.stern.de/gesundheit/depressionen-zurueck-ins-leben-3739184.html) with Words Differing Between Wording Lists (bold, see Jucks & Paus, 2013). The participants read only one of the bold words, which they saw in normal type.

Figure A2. Extract from Text 2 (https://www.welt.de/gesundheit/psychologie/article2435260/Griff-zur-Pille-hilft-bei-Depressionen-nicht-immer.html) with Words Differing Between Wording Lists (bold, see Jucks & Paus, 2013). The participants read only one of the bold words, which they saw in normal type.


8
7
6
5
4
3
2
1
0
post knowledge
-1
-0.5
0
0.5
1
motivation (z-standardized)
same wording
different wording

For a long time, **medicaments/drugs** were considered as the best **treatment/therapy** for depressions; now the doubts on their usefulness are increasing. New hope home in on **psychological treatments/psychotherapies:** They can guide the way out of the melancholy for many patients.

(Lange galten **Medikamente/Arzneien** als beste **Behandlung/Therapie** bei Depressionen, nun mehren sich die Zweifel an ihrem Nutzen. Neue Hoffnung richtet sich auf **Psychobehandlungen/Psychotherapien**: Vielen Patienten können sie den Weg aus der Schwermut weisen.)

Most people in Germany have difficulties to deal frankly with the **diagnosis/determination** of a depression. Affected persons prefer to keep silent about mental illnesses – and about their **therapy/treatment**. For prejudice still entwines around both.

(Den meisten Menschen in Deutschland fällt es schwer, offen mit der **Diagnose/Feststellung** einer Depression umzugehen. Betroffene schweigen lieber über psychische Erkrankungen – und über ihre **Therapie/Behandlung**. Denn immer noch ranken sich Vorurteile um beide.)