Effects of the Specificity and the Format of External Representations on Students' Revisions of Fictitious Others' Texts

Andreas Lachner* & Teresa Schurer°

* University of Freiburg; University of Tübingen; Leibniz-Institut für Wissensmedien, Tübingen | Germany

° University of Halle, Halle | Germany

Abstract: University students are often challenged with the demand of providing cohesive explanatory texts. To support students in revising their explanatory texts with regard to cohesion, it could be useful to provide students with external representations as formative feedback. In this study, we provided participants with a scenario in which they were asked to review a fictitious student's draft containing several cohesion gaps. Additionally, participants received an external representation as feedback to support them during their revisions. We varied the format (concept map versus outline) and the specificity (general versus specific) of the provided external representations. We found that participants with specific concept map representations correctly noticed more cohesion gaps, and perceived less cognitive load during reviewing than participants with the specific outline representation. Students with general external representations showed the lowest performance on the noticing task and the highest amount of cognitive load. However, there were no differences among the external representations regarding the quality of students' revisions. Evidently, specific concept maps can be regarded as a useful scaffold to support students' evaluation processes. However, additional instructional support is needed, particularly for novice writers, to effectively revise expository texts for cohesion.

Keywords: reviewing, formative feedback, cohesion, concept maps

journal of
WRITING RESEARCH

Lachner A., & Schurer T., (2018). Effects of the speci ficity and the format of external representations on students' revisions of fictitious others' texts. *Journal of Writing Research, 9*(3), 333-351. doi: 10.17239/jowr-2018.09.03.04

Contact: Andreas Lachner, Leibniz-Institut für Wissensmedien, Schleichstraße. 6, Tübingen, D-72076 | Germany – a.lachner@iwm-tuebingen.de.

1 Introduction

Providing comprehensible and organized explanatory texts can be regarded as a critical skill in our knowledge society (National Commission on Writing, 2004; Rowan, 1988). Besides other text features (e.g., the validity of the provided information, syntactic complexity, or concreteness), text comprehension research emphasized the central role of cohesion to enhance the comprehensibility of a text (Graesser, Millis, & Zwaan, 1997; McNamara, Crossley, & McCarthy, 2010; McNamara & Kintsch, 1996; Swales & Feak, 2012). Cohesive devices make the relationships of neighboring sentences explicit; this helps readers establish an organized representation of the text (McNamara & Kintsch, 1996; Swales & Feak, 2012). Cohesion can be achieved by making changes on a sentence level, either by inserting syntactic ties, such as using connectives (e.g., therefore, and, because), or by using semantic ties, for instance by using common noun phrases (e.g., by reiterating concepts), using near-synonyms, or by inserting bridging information which explicitly states the semantic relation between two adjacent sentences (Halliday & Hasan, 1976; McNamara et al., 2010).

Recent research, however, has documented that students tend to face difficulties to establish cohesion within their texts during writing (Concha & Paratore, 2011; Lachner & Nückles, 2015). One potential instructional strategy to enhance students' cohesive writing could be to scaffold students' revision activities. For instance, Lachner , Burkhart and Nückles (2017a) showed that providing students with computer-generated external representations as formative feedback significantly enhanced students' revisions and helped them improve the cohesion of their writing. However, it is less clear for which revision processes the external representations were particularly beneficial, and whether these effects depended on the particular format of the external representations. Furthermore, as writing research emphasized the central role of working memory during reviewing (Kellogg & Whiteford, 2009; Kellogg, Whiteford, Turner, Cahill, & Mertens, 2013; Klein & Boscolo, 2016), the question remains how students' level of cognitive load during reviewing accounted for their reviewing performance. Therefore, the aim of the experimental study reported in this paper was to examine a) whether and how different formats of external representations affected students' reviewing activities (i.e., the noticing and revision of cohesion deficits), and b) how potential effects of the external representations depended on students' particular level of cognitive load during reviewing.

1.1 Reviewing as a Complex Skill

Reviewing texts is a complex and demanding cognitive activity (Hayes, 2012; Myhill & Jones, 2007). Reviewing texts commonly comprises two sub-processes (Hayes, 2012). First, writers need to *evaluate* their current text products and notice potential writing problems (e.g., deficits of cohesion). Therefore, writers are required to re-read and simultaneously monitor for potential writing deficits (*Chanquoy*, 2009). Additionally, when writers notice distinct comprehension problems, they have to plan and implement additional *revision activities* in order to edit the critical text passages, which may result in additional reviewing cycles (Hayes, 2012; Kellogg, 1996). Therefore, writers need to retrieve and apply both topic knowledge about the content of the writing task itself, as well as discursive knowledge about potential linguistic strategies and the respective audience from long-term memory in order to revise one's text successfully for potential problems such as the comprehensibility of a text (Hayes, 2012).

Reviewing texts, however, commonly places intensive demands on writer's working memory (Kellogg & Whiteford, 2009). Therefore, particularly, students who are learning to write in a distinct genre such as explanatory writing, experience cognitive "overload" while revising a text (De Smet et al., 2014). Cognitive load refers to students' expended amount of mental effort during performing a learning task and works as a function of students' prior knowledge (Sweller, 2010). As such, learning to revise can often result in students' cognitive "overload", as students not only have to perform a revision task but, at the same time, have to accomplish a learning-to-revise task, which additionally loads on the limited capacity of human working memory (Kellogg et al., 2013). Therefore, students, who are often novices in writing, face difficulties to review their texts for comprehension problems (e.g., cohesion gaps). As a consequence, students' revisions often result in superficial repairs, such as orthographic or grammatical issues, which do not necessarily contribute to the comprehensibility of a text (Cho & MacArthur, 2010; Myhill & Jones, 2007).

For instance, Lachner et al. (2017a) analyzed students' revision activities during writing cohesive texts. In line with previous research (e.g., Kellogg, 1994; Sharples, 1999), the authors found that when students did not receive any further instructional scaffolds during reviewing, they predominantly accomplished revision activities which did contribute to the cohesion of their texts. These findings emphasize students' need for additional instructional scaffolds during reviewing to further enhance their writing.

1.2 Formative Feedback to Scaffold Students' Reviewing Activities

A prevalent instructional method to scaffold students' reviewing activities is the provision of formative feedback (Cho & MacArthur, 2010; Graham, Harris, & Hebert, 2011). Formative feedback is given by an instructional agent such as a teacher, peer, or a computer, and provides distinct information about a students' current writing performance (Cho & MacArthur, 2010; Hattie & Timperley, 2007). Regarding the effectiveness of formative feedback on writing, Graham et al. (2011) conducted a comprehensive meta-analysis on the benefits of formative feedback for the improvement of writing skills in adolescent students (based on 16 comparisons). Graham et al. obtained an average weighted effect of 0.77 (medium to large effect) for the provision of students with formative feedback.

Despite the beneficial effects of formative feedback, however, it is not always feasible for instructors (or even peers) to provide students with formative feedback on their current writing products, as providing feedback for an entire class is very laborious and often not manageable due to the specific time constraints in classroom instruction (Cho & MacArthur, 2010).

1.3 External Representations as Formative Feedback on Students' Reviewing Activities

Recently, research on technology-enhanced writing investigated potential effects of using computer-generated external representations as formative feedback to scaffold students' writing (Lachner et al., 2017a; Nussbaum, 2008; Pirnay-Dummer & Ifenthaler, 2011; Roscoe & McNamara, 2013; Wade-Stein & Kintsch, 2004). Therefore, students received external representations in addition to their current text product, which provided students with distinct information about the quality of their text product to review their drafts. These external representations had different formats. For instance, some researchers investigated effects of graphical representations (e.g., Lachner et al., 2017a; Pirnay-Dummer & Ifenthaler, 2011), whereas others relied on additional textual representations, such as outlines (De Smet et al., 2014; Roscoe & McNamara, 2013). Furthermore, these representations differed with regard to their specificity, as some representations only provided general information about the overall quality of the text (e.g., Roscoe & McNamara, 2013; Wade-Stein & Kintsch, 2004), or specific information, about potential allocations of writing deficits (Lachner et al., 2017a; Pirnay-Dummer & Ifenthaler, 2011). However, the question remains whether and how these different formats may have differently affected students' reviewing activities. In comparison to textual formats, graphical formats (e.g., concept maps) particularly could assist students more to exploit attentional processes during reviewing (Ainsworth, 2006; Larkin & Simon, 1987), such as the noticing of potential cohesion gaps. Thus, the graphical-spatial representation of concept maps could reduce the expended cognitive load during students' evaluation processes. The freed cognitive resources while evaluating could additionally be used by students to accomplish genuine revision activities which contribute to the overall quality of the text. Similarly, the specificity of the representations could affect students' evaluation activities, as specific information may more likely direct students to particular problems (e.g., cohesion deficits) in the texts, which would result in more genuine revision activities.

Lachner, Burkhart, and Nückles (2017b) directly examined the effects of the different representational formats, and varied the format of the representation (graphical versus textual) and the level of specificity of the representation (specific versus general). The authors investigated these effects in an ecologically field-setting within a large-lecture class on didactics. For two weeks, teacher education students (N = 251) were asked to provide an explanatory text about the current topic of the lecture as a homework assignment within one week. Afterwards, students either were randomly provided with concept map feedback (graphical representation), or outline feedback (textual representation), which contained the identical specific information but in a linear textual format. A control group of students received general textual feedback, as they only received information about the number of cohesion gaps, and a prompt to review for cohesion. To generate the different representations, the students' drafts were automatically segmented in propositions by using natural-language processing technology (Schmid & Laws, 2008). Afterwards, the computer-program visualized the propositional segmentation in the different representational formats (i.e., specific concept map, specific textual outline, or general textual feedback). Additionally, the students answered a short questionnaire on their perceived cognitive load during reviewing with the feedback. The authors found that the students who reviewed with specific concept map representations perceived lower levels of cognitive load during reviewing than the students with the specific outline representations and the general textual representations. More importantly, students who received specific concept map representations improved the cohesion of their explanations more than students with general textual representations. However, in contrast to their expectations, there were no significant differences between the concept map condition and the outline condition. Hence, based on the findings by Lachner and colleagues, one may conclude that only the specificity of representations but not the format affected students' reviewing activities. However, it has to be noted, that Lachner et al. could not clearly disentangle the effects of specificity and generality in their three-group design, as the general representation was only provided in textual format. As such, the obtained effects cannot be clearly attributed to the particular level of specificity, as the general representation condition only received general textual feedback but not general graphical feedback. Furthermore, it has to be acknowledged that the study by Lachner et al. (2017b) was conducted in an ecological, but less controllable field-setting. Such experimental settings on the one hand have high ecological validity, as these experiments resemble real-world applications of technology under authentic conditions. On the other hand, in such field-experimental studies it is less possible to control for additional extraneous variables (e.g., students' engagement to perform a task, time on task, or their commitment during participation) which potentially biased the obtained results. These biases, however, make it difficult for other researchers to replicate the study which may result in lower levels of internal validity. Hence, laboratory experimental studies with higher levels of control could be useful to clearly disentangle effects of the specificity and representational format of external representations on students' reviewing performance.

1.4 Overview of the Current Study

The current study aimed to replicate previous findings by Lachner et al. (2017b) under well-controlled laboratory conditions. For that purpose, we conducted an experimental study to investigate directly the effects of the specificity and the representation format of the feedback on students' reviewing. Hence, in contrast to the study by Lachner et al. (2017b), we followed a scenario approach (see also Bolzer, Strijbos, & Fischer, 2014; Hayes et al., 1987; Strijbos, Narciss, & Dünnebier, 2010, for similar approaches) to control better for potential inter-individual differences among students' drafts.

In this scenario, we provided German university students with a fictitious student's draft about cognitive load theory, and asked them to review this text for cohesion. The text was in German and contained several cohesion gaps. All participants were asked to identify the cohesion gaps and to provide a revision proposal for the localized cohesion gaps. During the reviewing scenario, depending on the experimental condition, the participants were provided with an external representation that varied with regard to the format (concept map versus outline) and its specificity (general versus specific) as formative feedback to the draft. This scenario approach allowed us experimentally to disentangle effects of the external representations (format versus specificity) on students' evaluation and revision activities separately, while keeping the amount of information provided within the external representations, and the appearance of the external representations constant within conditions. Using a scenario approach was further motivated by research which directly tested the convergence of cognitive activities of realistic situations (e.g., reviewing one's own text) versus scenarios (e.g., reviewing a fictitious student's text) suggesting that participants tend to act nearly similarly in real situations and in scenarios (Robinson & Clore, 2001; Strijbos et al., 2010).

1.5 Hypotheses

Following the findings by Lachner et al. (2017b), we assumed that both the graphical-spatial organization, as well as the provision of specific information about cohesion gaps, should be germane for students reviewing their texts for cohesion. Hence, we hypothesized that students with specific concept maps would correctly locate more cohesion gaps in the draft and provide more correct revision proposals, as compared to students with the specific outline, as the spatial representation of the concept maps would highlight potential cohesion gaps more than the textual representations in outlines (Ainsworth, 2006; Larkin & Simon, 1987). Students with general representations (i.e., general concept map or general outline) should notice the fewest cohesion gaps in the draft and provide lower-quality revision proposals, as the students had no specific information about the allocation of the depicted cohesion gaps available (*reviewing-hypotheses*).

In the same vein, we assumed that the level of perceived cognitive load during reviewing would mediate the effect of type of external representation on students' revision of cohesion gaps (*cognitive-load-mediates-reviewing-hypotheses*).

2 Method

2.1 Participants

Fifty-eight Educational Science students participated in the experiment. Their mean age was 23.45 (*SD* = 4.01). Three quarters (74.60 %) of the students were female. The students were in their fourth semester on average (*SD* = 1.41). They reported medium prior-knowledge in the domain of learning and instruction with *M* = 3.00 (*SD* = 0.73, on a 5 point Likert scale), which was the main explanatory domain of the writing task. Their self-reported mean German grade in their final exams at secondary education was $M = 2.10$ ($SD = 0.73$, varying from 1 = excellent to 6 = insufficient), as a distal indicator for their writing skills. All students had attended an introductory course in educational psychology beforehand, in which they had been introduced to cognitive load theory, and to the use of concept maps and outlines. None of them had yet attended a writing course. The students participated in exchange for course credit.

2.2 Design

The students were randomly assigned to one of the four experimental conditions. A 2 × 2-between-subjects factorial design was used with the factors "external representation" (concept map versus outline) and "specificity" (specific versus general). Dependent variables encompassed a) the number of correctly diagnosed cohesion gaps, and b) the number of correct revision proposals. The level of perceived cognitive load was the mediating variable.

2.3 Materials

2.3.1 Draft of fictitious student.

We provided the students with a draft of a fictitious student called "Hans". The draft was an explanatory text about cognitive load theory (see Appendix). The original draft was based on an encyclopedia entry by Scheiter (2013). In a first step, we optimized the original entry by Scheiter, so that the entry would contain no cohesion gaps. Therefore, we included bridging information, added connectives, and increased the argument overlap between adjacent sentences (McNamara et al., 2010). This optimized version was used as a sample solution of a high-cohesive text (see also analysis and coding section of the reviewing proposals). In the next step, we consciously inserted eight cohesion gaps by deleting connectives or decreasing argument overlap by using abbreviations, far-synonyms and pronouns, or by deleting bridging information. For that purpose, we reviewed a representative sample of low-cohesive explanatory texts on cognitive load theory of the study by Lachner et al. (2017a), and inserted typical errors to make the fictitious student's draft as authentic as possible. The draft comprised 465 words. The average text difficulty of the draft, as indicated by the Flesch-Reading-Ease-Index, was 31, which can be regarded as rather difficult (0 = very difficult to 100 = very easy).

2.3.2 External representations.

Additionally, we randomly provided students with one of the four external representations of the draft by the fictitious student "Hans" (see Figure 1) as formative feedback.

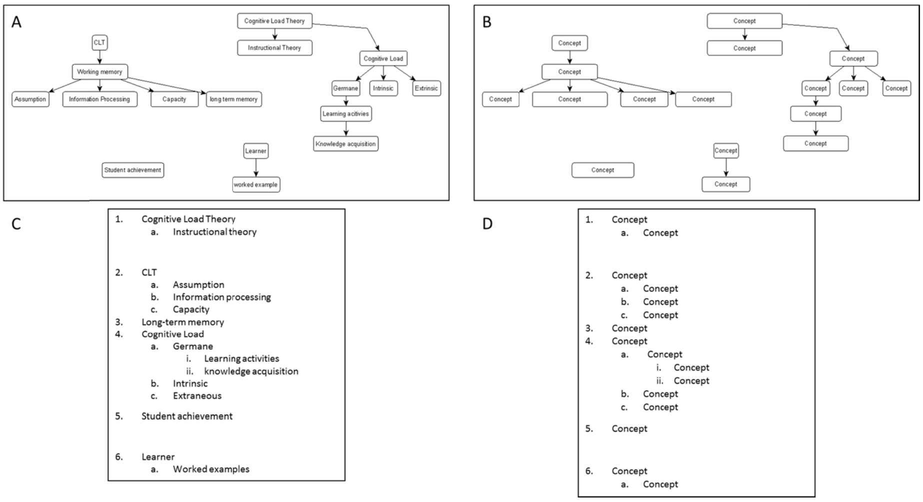

Figure 1: Excerpts of the external representations used in the experiment. 1A is an excerpt for the specific concept map condition; 1B for the general concept map condition; 1C for the specific outline condition; and 1D for the general outline condition.

2.3.3 Concept map representation.

To generate a specific concept map, we followed the procedure by Lachner et al. (2017a), and first segmented the explanations into propositions. For instance, the sentence: "Peter's father lost his wallet" would be represented as: Father → Peter; Father → Wallet. Second, we replaced all concepts that could be treated as synonyms with Open Thesaurus (http://www.openthesaurus.de/), an open source database for near-synonym suggestions. Third, we sequentially connected those propositions that had at least one identical concept (in the latter example it would be "Father"), or were related by a clear connective. If the propositions had no identical concept or no connective, no relation was drawn (Patel, Groen, & Arocha, 1990). This approach allowed us to construct a concept map of the draft (see Figure 1A). A cohesion gap within the concept map was presented as an isolated information unit that was not related to the rest of the concept map (Lachner et al., 2017a).

To obtain the general concept map, we followed the procedure by Gurlitt and Renkl (2010), and deleted the original concepts and only printed blank ovals of the concepts (see Figure 1B). Thus, we deleted the specific information about the allocation of the cohesion gaps within the concept maps, so that only the general structure of the explanation (i.e., the number of cohesion gaps, and the number of concepts and relations) was explicit in the concept map. Thus, students would need to infer the allocation of the cohesion gaps.

2.3.4 Outline representation.

To keep the information identical across conditions, for the outline representation, we similarly used the propositional information derived during the concept map construction. In contrast to the concept map condition, the concepts of each sentence were printed serially below each other and were consecutively numbered (see Lachner et al., 2017b). Cohesion gaps were marked by a blank line between the subsequent sentences (see Figure 1C).

As for the general concept map condition, for the general outline condition, we deleted the original concepts, so that only the numeration was present (see Figure 1D). Thus, also the number of concepts was present, as well as the number of cohesion gaps marked by blank lines. However, students had no information available about where the cohesion gaps were allocated in the text.

2.3.5 Perceived cognitive load.

At the end of the reviewing phase, the students assessed their perceived difficulty during reviewing the draft. We decided to choose subjective difficulty ratings instead of other subjective methods, such as mental effort ratings, or objective measures, such as reaction times, as difficulty ratings have been shown to be rather sensitive to cognitive load aspects which occur during learning (DeLeeuw, & Mayer, 2008). For that purpose, we handed out a short questionnaire used by Lachner et al. (2017b), and originally developed by Berthold and Renkl (2009). Students' perceived difficulty was assessed by four items on a 5-point rating-scale (1 = easy, 5 = difficult; e.g., "How easy or difficult was it for you to work with the feedback?" "How easy or difficult was it for you to distinguish between important and unimportant information of the feedback?"). The reliability of the questionnaire was very good, ω = .94 (MacDonalds ω). Recent research documented that, particularly for short scales, MacDonalds ω provides a less biased reliability estimate than other conventional reliability methods such as Cronbach's alpha (Dunn, Baguley, & Brunsden, 2014).

2.4 Procedure

The experimental study took place during regular classroom sessions of the first author. The students sat separately so that they could neither see the other students' solutions, nor what kind of feedback the other students received. The second author led the experimental sessions. During the entire experimental session, the students were not allowed to proceed before being signaled by the experimenter (exact time on task). At the beginning, we obtained oral consent from all participants. Next, students answered a demographic questionnaire (5 minutes).

Afterwards, we provided the students with the scenario, as well as the draft by the fictitious student Hans. They were informed that Hans' draft contained several text passages which were difficult to relate to each other. The students were asked to read Hans' draft (10 minutes). Afterwards, they randomly received one of the four external representations (specific concept map, general concept map, specific outline, or general outline). They were informed that Hans received the external representation as additional feedback to assist him during his review. They were further informed that the external representation was an alternative representation to make text passages that contain cohesion gaps explicit. Students were instructed about how to read and interpret the particular external representation properly.

Next, they marked the text passages that potentially included cohesion gaps by underlining the particular text passages (15 minutes). Finally, they were asked to provide a complete revision proposal to improve the marked cohesion gaps (25 minutes). At the end of the reviewing phase, students answered the subjective cognitive load questionnaire (5 minutes). After the study, the students were debriefed. Thus, all the students were informed about the aim of the study, as well as the experimental manipulation. All the students additionally received the sample solution of the draft as well as a short presentation about effective revision strategies in order to guarantee equal conditions among the participants. Students received course credit for their participation. The experimenter followed the APA standards for the ethical treatment of human participants.

2.5 Analysis and coding

2.5.1 Evaluation performance.

Following Strijbos et al. (2010), to obtain a measure of students' evaluation performance, a coder counted the students' number of correctly identified cohesion gaps in Hans' draft. Therefore, the coder had a transparent template at hand which could be placed over the students' answers to count the number of correctly identified cohesion gaps. By using the template during the coding process, coding errors should be reduced to a minimum. For determining the inter-rater reliability, a second coder counted a subset of 13 students' answers (23% of all students' answers). Previous research on adequate sample sizes in reliability studies indicated to use at least nine cases as a subset, when high *ICC*s are assumed, (Walter, Eliasziw, & Donner, 1998). Inter-rater agreement was assessed by calculating a two-way mixed, single-measure *ICC*. The resulting inter-rater agreement was excellent *ICC* = .94. Thus, one rater coded the rest of the students' answers.

2.5.2 Revision performance.

On the basis of the sample solution, two trained raters assessed the revision proposals. For the revision proposals, students received 1 point, when they successfully closed a cohesion gap (e.g., by using argument overlap, connectives, or bridging information), or zero points, when the cohesion gap was not closed. Please note that the consistency among the raters should be considerably lower than in the evaluation task, as there was a larger variability of possible correct solutions. Therefore, as a potential safeguard, we asked our two coders to rate the entire set of students' revision proposals to measure the consistency of our raters (Walter et al., 1998). Inter-rater agreement was good, as indicated by a two-way mixed, single-measure *ICC* = .79. Differences among the raters were resolved by discussion.

3 Results

A series of ANOVAs and χ^2 tests revealed no significant differences between the experimental conditions concerning age, $F(3, 54) = 0.48$, $p = .70$; number of enrolled semesters, $F(3, 54) = 1.06$, $p = .37$; and their subjective prior knowledge, $F(3, 54) = 1.10$, $p = .36$. Furthermore, students did not differ with regard to their amount of marked cohesion gaps, $F(3, 54) = 0.92$, $p = .44$. This finding indicates that the groups of students were comparably engaged to detect potential cohesion gaps, regardless of the particular representation they received.

3.1 Reviewing-Hypotheses

To test our reviewing-hypotheses directly, we computed a series of a-priori contrast analyses (Furr & Rosenthal, 2003). As these analyses commonly require normally-distributed data, which often does not hold true for "count data", we followed suggestions by Gardner, Mulvey, and Shaw (1995), and used negative binomial distribution models in our analyses (see also Beaujean & Morgan, 2016).

With regard to the number of correctly diagnosed cohesion gaps, we hypothesized that students with a specific concept map would correctly detect more cohesion gaps in Hans' draft as compared with students with a specific outline, whereas students with a general concept map and students with a general outline should identify the fewest cohesion gaps. Therefore, we assigned the following contrast weights to our experimental conditions and used these values as dummy-contrast: specific concept map = 2, specific outline = 1, general concept map = -1.5, general outline = -1.5. We performed a negative binomial regression analysis, and included the contrast dummy as independent variable, and the students' evaluation performance was the dependent variable. The contrast was indeed significant, Wald-χ^2 (2) = 10.14, $p = .01$, indicating that students with a specific concept map outperformed students with a specific outline, and students with general external representations (concept map and outline) showed the lowest performance on the evaluation task (see Table 1).

With regard to students' reviewing performance, we performed a negative binomial regression analysis with the contrast dummy as independent variable, and the students' number of correct reviewing proposals as the dependent variable. However, the contrast was not significant, Wald-χ^2 (2) = 2.10, $p = .35$. Despite its significant effects on noticing potential cohesion gaps, the specific concept map representation did not further contribute to assist students when providing their suggestions to improve the cohesion of Hans' draft. However, it has to be noted that the students' overall performance on the revision proposals was rather low, indicating that students had only limited knowledge about how to revise expository texts for cohesion effectively (see the descriptive values in Table 1).

Table 1. Means and Standard Deviations (in Parentheses) for the Dependent Measures

Dependent variable	Specific concept map	Specific outline	General concept map	General outline
Number of correctly diagnosed cohesion gaps	3.19 (2.20)	1.80 (1.52)	1.57 (1.60)	1.46 (0.97)
Number of correct reviewing proposals	1.53 (1.19)	1.31 (1.49)	1.09 (1.51)	1.25 (1.29)
Subjective cognitive load	3.32 (1.20)	4.68 (0.51)	4.73 (0.45)	4.73 (0.52)

Note. The number of correctly diagnosed cohesion gaps and the number of correct reviewing proposals could vary between 0 and 7. The perceived cognitive load could vary from 1 (= low) to 5 (= high).

3.2 Cognitive-Load-Mediates-Reviewing-Hypothesis

To test whether the effect of specific concept maps on the students' evaluation performance (i.e., the number of correctly identified cohesion gaps) was mediated by the student's perceived cognitive load, we calculated indirect effects by applying the bootstrapping procedure implemented in the PROCESS macro for SPSS (Hayes, 2013). The bootstrapping procedure can be regarded as statistically adequate analysis for our data, as it does not require normally distributed data (Hayes, 2013). We computed unstandardized indirect effects with 10,000 resamples with replacement to derive a 95% bias-corrected confidence interval for the indirect effects. As in the prior analyses, the type of external representation was the contrast-coded predictor (specific concept map = 2, specific outline = 1, general concept map = -1.5, general outline = -1.5), students' performance on the evaluation task the criterion, and students' perceived cognitive load was the mediating variable. The mediation analysis indicated an unstandardized indirect effect of 0.16 (κ^2 =.09, medium effect) with a 95 % confidence interval ranging from .0014 to .4068. As the confidence interval did not include zero, the indirect effect was significant. Hence, we can conclude that the specific concept map helped students better evaluate the draft, as it reduced the perceived difficulty during their evaluation.

4 Discussion

In the present experimental study we aimed to replicate previous findings by Lachner et al. (2017b) that specific concept maps better support students' reviewing activities than outline representations or general feedback. In this study, we followed a scenario approach and provided Educational Science students with a scenario in which they were asked to review a fictitious student's draft containing several cohesion gaps, to reduce inter-individual variance among students. Additionally, we used a balanced factorial design and provided the students with external representations differing in their format (textual format: outline; versus graphical format: concept map) and their level of specificity (general versus specific) to support them during reviewing.

In line with our hypotheses, we found that the specific concept map representation helped students better evaluate a draft for cohesion, while inducing lower levels of cognitive load as compared to the specific outline representation. At the same time, the general external representations were least effective to support students' evaluation processes of cohesion gaps, as they induced the highest amount of perceived cognitive load. With regard to students' quality of revision proposals, however, the type of external representations did not affect their revision performance. Thus, we can conclude that specific graphical representation of concept maps was appropriate to facilitate students' evaluation processes for cohesion. However, they were not more effective than other types of formative feedback to scaffold students' planning and implementing of appropriate revisions to close the diagnosed cohesion gaps. This finding is interesting, as Lachner et al. (2017b) found that students with specific representations were also more able to revise the cohesion of their texts compared to a group of students with general feedback. We attribute these contradictory findings to the different conditions in which the two studies were conducted. In the study by Lachner et al. students reviewed their own authentic text. In the recent study, by contrast, we provided students with experimenter-generated text material for review. Thus, although scenario approaches may trigger similar processes such as authentic writing scenarios (Bolzer et al., 2014; Strijbos et al., 2010), the fictitious text material was potentially less familiar to the students than their own text. From a cognitive perspective, the experimenter-generated material could have induced higher levels of cognitive load during students' revision, as the experimenter-generated material required more cognitive processing to build up a textual representation in order to understand the draft (Graesser et al., 1997). This, however, could have been at the expense of students' overall reviewing performance, as they potentially had less capacity available for more germane revision activities such as undertaking revisions that aimed at improving the cohesion of the draft. Comparing the students' cognitive load ratings of our study to the ones in the Lachner et al. (2017b) study, it is remarkable that, on average, the cognitive load ratings in the current study were around one standard-deviation higher, favoring our cognitive load explanation. However, this interpretation is highly speculative, as the students in the current study and in the one by Lachner et al. (2017b) dealt with different topics (text comprehension versus cognitive load), and had different time constraints to accomplish their tasks (twenty minutes versus one week), which may, alternatively, explain the students' different difficulty ratings across these studies. Therefore, future research should address this methodological question in further studies, and compare students' perceived cognitive load as well as their writing performance when reviewing one’s own text or texts by (fictitious) others.

4.1 Theoretical and Practical Contributions of our Study

With regard to effects of external representations, our findings extend previous findings by Lachner et al. (2017b), as we were able to clearly disentangle distinct reviewing activities in our study (i.e., evaluation and revision processes). As such, our findings contribute to a better understanding about the effects of external representations, as they showed that graphical external representations mainly enhanced students’ attentional processes, that is detecting cohesion gaps during reviewing, but not the actual revision processes that is to plan and edit appropriate revisions. These findings are consistent with recent findings of a qualitative study by Olmanson et al. (2016). Instead of receiving concept map representations, the authors asked students to use a concept mapping tool to re-visualize their texts. In line with our findings, the authors found that generating concept maps raised students’ attentional evaluation processes. However, students still faced problems to implement revisions that aimed at contributing to the overall quality of the text. Therefore, these findings are suggestive of ways to explore potential additional instructional methods to supplement the provision of external representations during writing (Shute, 2008). For instance, it could be a beneficial approach to provide students with prior-strategy instruction (Fidalgo, Torrance, Rijlaarsdam, van den Bergh, & Álvarez, 2015), in which students receive direct writing instruction regarding distinct revision strategies to enhance the cohesion of a text. This prior-strategy instruction could help students acquire generic revision strategies which could be applied in the subsequent revision phase.

Furthermore, our findings contribute to a better understanding of the role of cognitive load during learning to write, as we showed that the effectiveness of our external representations could be explained by students’ subjective cognitive load during revising. So far, however, only a few writing studies have included cognitive load measures in their analyses (for exceptions, see De Smet et al., 2014; Lachner et al., 2017b). Therefore, we argue for incorporating cognitive load measures in research on writing to explain better why distinct writing interventions worked or not. However, we have to note that the researchers should thoroughly consider which cognitive load measure to employ, as the different measures are differently capable of measuring different facets of cognitive load. Therefore, more objective cognitive load measures, which rely on gaze data or EEG-data (e.g., Scharinger, Kammerer, & Gerjets, 2015) could be an alternative to obtain a more valid measure of cognitive load during students’ reviewing.

4.2 Study Limitations and Further Research

Despite the valuable insights of our study, there are also some limitations that need to be addressed. We decided to use a balanced 2 × 2 factorial design, to cross the factors feedback format (i.e., concept map versus outline representation), and the level of information (i.e., general versus specific information). This procedure, on the one hand, allowed us clearly to disentangle differential effects of the format and the specificity of the external representations. On the other hand, it probably produced rather artificial and high-demanding representations, particularly for the general feedback conditions (see Figure 1C and 1D), resulting in a potential overestimation of the effect of the graphical representation on students' evaluation processes. A related limitation refers to the fact that we did not include a control group of students who did not receive a representation during the reviewing phase. Therefore, our findings cannot provide evidence whether the specific concept feedback really was more beneficial in comparison to a no-feedback condition. However, it has to be noted that Lachner et al. (2017a) recently showed that students provided more cohesive explanations with concept map feedback as compared to students without feedback. Therefore, we are still tempted to attribute our findings to the general effectiveness of the concept map representation as a method of formative feedback.

Furthermore, we have to admit that we did not assess students' reading and writing skills prior to the study. Students' reading abilities, as well as their writing abilities, could have largely influenced the effects of external representations on students' writing. We, therefore, carefully selected a rather homogeneous sample of university Educational Science students who, likely, were comparable with regard to their reading and writing skills. This procedure was also confirmed by previous analyses which showed that our experimental conditions were comparable, at least regarding their German grades, as a very distal indicator of students' writing and reading skills. Nevertheless, future research should include more specific indicators for students' reading and writing skills to control for potential differences in their studies.

A final caveat refers to the ecological validity of our findings. To control for potential intra-individual differences of drafts, we followed a laboratory scenario approach to give students an experimenter-generated draft by a fictitious student. Therefore, it is unclear how our findings relate to authentic revision activities of one's own text in realistic settings. Therefore, future studies should try to generalize our findings to authentic reviewing activities, when students are required to revise their own texts in their disciplinary courses.

All in all, the present experiment shows that specific concept map representations are a valuable scaffold for formative feedback to guide students during their evaluation activities. However, the findings also suggest that integrating further instructional methods are essential to fully exploit the potential of concept map representations as formative feedback.

References

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

Beaujean, A. A., & Morgan, G. B. (2016). Tutorial on Using Regression Models with Count Outcomes using R. *Practical Assessment, Research & Evaluation, 21*.

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101*(1), 70–87. doi:10.1037/a0013247

Bolzer, M., Strijbos, J. W., & Fischer, F. (2014). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning*. doi: 10.1111/jcal.12091

Chanquoy, L. (2009). Reviewing processes. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *The SAGE handbook of writing development.* (pp. 80-98). London: SAGE Publications Ltd. doi: http://dx.doi.org/10.4135/9780857021069.n6

Cho, K., & MacArthur, C. (2010). Student reviewing with peer and expert reviewing. *Learning and Instruction, 20*(4), 328-338. doi: 10.1016/j.learninstruc.2009.08.006

Concha, S., & Paratore, J. R. (2011). Local coherence in persuasive writing: An exploration of Chilean students' metalinguistic knowledge, writing process, and writing products. *Written Communication, 28*(1), 34-69. doi: 10.1177/0741088310383383

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*, 223-234. doi: 10.1037/0022-0663.100.1.223

De Smet, M. J. R., Brand-Gruwel, S., Leijten, M., & Kirschner, P. A. (2014). Electronic outlining as a writing strategy: Effects on students' writing products, mental effort and writing process. *Computers & Education, 78*, 352-366. doi:10.1016/j.compedu.2014.06.010

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399-412. doi: 10.1111/bjop.12046

Fidalgo, R., Torrance, M., Rijlaarsdam, G., van den Bergh, H., & Álvarez, M. L. (2015). Strategy-focused writing instruction: Just observing and reflecting on a model benefits 6th grade students. *Contemporary Educational Psychology*, *41*, 37-50. doi: 10.1016/j.cedpsych.2014.11.004

Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 2*(1), 33-67. doi: 10.1207/S15328031US0201_03

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological bulletin*, *118*(3), 392. doi: 10.1037/0033-2909.118.3.392

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*(1), 163-189. doi:10.1146/annurev.psych.48.1.163

Graham, S., Harris, K., & Hebert, M. A. (2011). *Informing writing: The benefits of formative assessment. A Carnegie Corporation Time to Act report.* Washington, DC: Alliance for Excellent Education.

Gurlitt, J., & Renkl, A. (2010). Prior knowledge activation: how different concept mapping tasks lead to substantial differences in cognitive processes, learning outcomes, and perceived self-efficacy. *Instructional Science, 38*(4), 417-433. doi:10.1007/s11251-008-9090-5

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in english. London: Longman

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112. doi: 10.3102/003465430298487

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach . New York: Guilford Press.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written communication*, *29*(3), 369-388. doi: 10.1177/0741088312451260

Hayes, J. R., Flower, L. S., Schriver, K., Stratman, J., & Carey, L. (1987). Cognitive processes in reviewing. In S. Rosenberg (Ed.), *Reading, writing, and language learning Advances in applied psycholinguistics*, Vol. II, (pp. 176-240). Cambridge: Cambridge University Press.

Kellogg, R. (1994). *The psychology of writing*. New York: Oxford University Press.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57-71). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250-266. doi: 10.1080/00461520903213600

Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M., & Mertens, A. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research, 5*, 159-190. doi: 10.17239/jowr-2013.05.02.1

Klein, P. D., & Boscolo, P. (2016). Trends in research on writing as a learning activity. *Journal of Writing Research*, *7*(3), 311-350. doi: 10.17239/jowr-2016.07.3.01

Lachner, A., Burkhart, C., & Nückles, M. (2017b). Mind the gap! Automated concept map feedback supports students in writing cohesive explanations. *Journal of Experimental Psychology: Applied, 23*(1), 29-46. http://dx.doi.org/10.1037/xap0000111

Lachner, A., Burkhart, C., & Nückles, M. (2017b). Formative computer-based feedback in the university classroom: Specific concept maps scaffold students' writing. *Computers in Human Behavior, 72*(4), 459-469. http://dx.doi.org/10.1016/j.chb.2017.03.008

Lachner, A., & Nückles, M. (2015). Bothered by abstractness or engaged by cohesion? Experts' explanations enhance novices' deep-learning. *Journal of Experimental Psychology: Applied, 21*(1), 101-115 doi: 10.1037/ /xap0000038

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*(1), 65-100. doi: 10.1111/j.1551-6708.1987.tb00863.x

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes, 22*(3), 247–287. doi:10.1080/01638539609544975

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*(4), 292-330. doi: 10.1080/0

Myhill, D., & Jones, S. (2007). More than just error correction: Student's perspectives on their reviewing processes during writing. *Written Communication, 24*(4), 323-343. doi: *10.1177/0741088307305976*

National Commission on Writing (2004, September). *Writing: A ticket to work... or a ticket out.* Available on www.collegeboard.com

Nussbaum, E. M. (2008). Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology, 100*(3), 549-565. doi: 10.1037/0022-0663.100.3.549

Olmanson, J., Kennett, K., Magnifico, A., McCarthey, S., Searsmith, D., Cope, B., & Kalantzis, M. (2016). Visualizing Reviewing: Leveraging Student-Generated Between-Draft Diagramming Data in Support of Academic Writing Development. *Technology, Knowledge and Learning*, 1-25. doi: 10.1007/s10758-015-9265-5

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition, 18*(4), 394-406. doi:10.3758/BF03197128

Pirnay-Dummer, P., & Ifenthaler, D. (2011). Text-guided automated self-assessment. A graph-based approach to help learners with ongoing writing. In D. Ifenthaler, Kinshuk, P. Isaias, D. G. Sampson & J. M. Spector (Eds.), *Multiple perspectives on problem solving and learning in the digital age* (pp. 217-225). New York: Springer.

Robinson, M. D., & Clore, G. L. (2001). Simulation, scenarios, and emotional appraisal: Testing the convergence of real and imagined reactions to emotional stimuli. *Personality and Social Psychology Bulletin*, 27(11), 1520-1532. doi: 10.1177/01461672012711012

Roscoe, R., & McNamara, D. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology, 105*, 1010-1025. doi: 10.1037/a0032340

Rowan, K. E. (1988). A contemporary theory of explanatory writing. *Written Communication*, 5(1), 23-56. doi: 10.1177/0741088388005001002

Scharinger, C., Kammerer, Y., & Gerjets, P. (2015). Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading. *PLoS ONE*, 10, e0130608. doi: 10.1371/journal.pone.0130608

Scheiter, K. (2013). Cognitive load theory. In M. A. Wirtz (Ed.), Dorsch – Lexikon der Psychologie [Dorsch – encyclopedia of psychology] (p. 328-329). Bern: Huber.

Schmid, H., & Laws, F. (2008). Estimation of conditional probabilities with decision trees and an application to fine-grained POS tagging. *Proceedings of the 22nd International Conference on Computational Linguistics*.

Sharples, M. (1999). *How we write: Writing as creative design*. London: Routledge.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189. doi: 10.3102/0034654307313795

Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing reviewing tasks: Are they critical for feedback perceptions and efficiency? *Learning and instruction, 20*(4), 291-303. doi:10.1016/j.learninstruc.2009.08.008

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills* (3rd edition). Ann Arbor, MI: University of Michigan Press.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). New York: Cambridge University Press.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction, 22*(3), 333-362. doi: 10.1207/s1532690xci2203_3

Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in medicine*, *17*(1), 101-110. doi:10.1002/(SICI)1097-0258(19980115

Appendix A:
Translated draft about cognitive load theory used in the experiment

Cognitive Load Theory

According to cognitive load theory, three types of loads are considered, extraneous load, intrinsic load, and germane load. | The CLT is based on specific assumptions of information-processing. The working memory, which is the central information-processing unit, is assumed to be limited. Contrarily, the long-term memory is perceived to be unlimited. Germane load indicates load induced by genuine learning activities that support knowledge acquisition. | The higher it is, the higher the learning-achievement. When, for instance, a learner is provided with several examples, this procedure increases one's mental effort, but at the same time also facilitates one's schema acquisition. This positive load results from the application of deep learning processes, which go beyond the simple rehearsal of information in the working memory. Intrinsic load primarily indicates the complexity of the task, which is defined by the level of element-interactivity. | It can be described by the number of components that have to be learnt. | When it is high, all elements have to be processed simultaneously. | If it is low, they can be learnt serially. An example for high intrinsic load is the learning of syntax of a foreign language, as everything has to be processed simultaneously (word meaning, sentence positions, and grammar) to understand the meaning of the syntax. | Low intrinsic effort occurs during simple vocabulary learning, as the vocabulary-pairs can be processed serially. | Main difficulty is primarily the amount of words. | Thus, the interaction is dependent on students' prior knowledge. The higher a learners' prior knowledge, the more elements can be processed simultaneously. Extraneous load is defined as the load that is caused by the instructional design itself, for instance by a sub-optimal design of learning materials. Main goal is to reduce the level of extraneous load and use the available cognitive resources for more genuine learning activities to increase germane load.

Note. Cohesion gaps are marked by "|".