Learning and Instruction 74 (2021) 101438

Contents lists available at ScienceDirect

Learning and Instruction

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

Learning by writing explanations: Is explaining to a fictitious student more effective than self-explaining?

Andreas Lachner [a,b,*], Leonie Jacob [b], Vincent Hoogerheide [c]

[a] *Department of Education, University of Tübingen, Germany*
[b] *Leibniz-Institut für Wissensmedien, Tübingen, Germany*
[c] *Department of Education, Utrecht University, the Netherlands*

A R T I C L E I N F O

Keywords:
Learning by explaining
Self-explaining
Retrieval practice
Generative learning

A B S T R A C T

Research has demonstrated that oral explaining to a fictitious student improves learning. Whether these findings replicate, when students are writing explanations, and whether instructional explaining is more effective than other explaining strategies, such as self-explaining, is unclear. In two experiments, we compared written instructional explaining to written self-explaining, and also included written retrieval and a baseline control condition. In Experiment 1 ($N = 147$, between-participants-design, laboratory experiment), we obtained no effect of explaining. In Experiment 2 ($N = 50$, within-participants-design, field-experiment), only self-explaining was more effective than our control conditions for attaining transfer. Self-explaining was more effective than instructional explaining. A cumulating meta-analysis on students' learning revealed a small effect of instructional explaining on conceptual knowledge ($g = 0.22$), which was moderated by the modality of explaining (oral explaining > written explaining). These findings indicate that students who write explanations are better off self-explaining than explaining to a fictitious student.

1. Introduction

Providing explanations is commonly regarded as a beneficial strategy to enhance students' learning (e.g., Fiorella & Mayer, 2014; Palincsar & Brown, 1984; Plötzner, Dillenbourg, Preier, & Traum, 1999; Roscoe, 2014; Roscoe & Chi, 2008). In early learning-by-explaining research, explaining as a learning activity was predominantly applied in *interactive settings* in which students provided instructional explanations of the content with the explicit intention to teach peer-students who were interactive and physically present (e.g., Palincsar & Brown, 1984; Plötzner et al., 1999; Renkl, 1995; Roscoe, 2014; Roscoe & Chi, 2008; Webb, Troper, & Fall 1995). However, even without interacting with a peer, providing instructional explanations has shown to be a beneficial instructional activity, as demonstrated by recent empirical research in which students provided instructional explanations to a *fictitious* and non-present other student by means of video-based oral explanations (Fiorella & Mayer, 2013, 2014; Hoogerheide, Loyens, & van Gog, 2014; Hoogerheide, Renkl, Fiorella, Paas, & van Gog, 2019; Hoogerheide, Visee, Lachner, & van Gog, 2019). To differentiate among different explaining activities, for the purposes of this article, we use the term instructional explaining to refer to an explaining situation, in which students act as teachers, and provide an explanation about the previously learnt contents to a mostly less knowledgeable student.

From a practical perspective, asking students to provide oral instructional explanations is often not feasible in the classroom, as it requires the availability of distinct technologies and infra-structure to generate the explanations. It is an open question, however, whether the findings of oral explaining would replicate in more parsimonious contexts with lower amounts of technical infrastructure, such as writing explanations (e.g., Lachner & Neuburg, 2019; Okita & Schwartz, 2013). On the one hand, writing offers students the opportunity to externalize their ideas and organize their thoughts (Klein, Boscolo, Kirkpatrick, & Gelati, 2014). On the other hand, writing explanations may impose additional cognitive demands, as students have to instantiate a particular rhetorical structure during writing, which could impair students' learning (Lachner & Neuburg, 2019; Sperling, 1996).

Against this background, we conducted two experiments both in a laboratory setting (Experiment 1) and in a field-setting (Experiment 2). The aims of the experiments were twofold: First, we investigated, whether the findings of explaining on students' learning would replicate, when students provide instructional explanations in written form. Second, we examined, whether the potential findings depend on the

* Corresponding author. Department of Education, Haußerstraße 43, D-72076, Tübingen, Germany.
E-mail address: andreas.lachner@uni-tuebingen.de (A. Lachner).

https://doi.org/10.1016/j.learninstruc.2020.101438
Received 27 January 2020; Received in revised form 16 November 2020; Accepted 10 December 2020
Available online 25 December 2020
0959-4752/

induced social context during explaining, as during instructional explaining students explain the content to fictitious students. To obtain robust findings regarding the effectiveness of writing instructional explanations, we compared writing instructional explanations to related yet distinct control conditions (i.e., retrieval practice, self-explaining) which did not have a social component (retrieval practice, self-explaining), or involve lower levels of generative activities as compared to instructional explaining (retrieval practice), as well as a baseline condition. Additionally, we provide updated estimates of the effectiveness of instructional explaining by means of a continuously cumulating meta-analysis (based on a recent meta-analysis by Kobayashi, 2018).

1.1. *Learning-by-explaining to a fictitious other student*

Several studies demonstrated that explaining the contents of learning materials to a fictitious (and less knowledgeable) other student is a beneficial activity for learning, and more effective than simply restudying the learning material (e.g., Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014). In line with generative learning theory (Wittrock, 2010), explaining allows students to build new knowledge by engaging in deep-level cognitive processes (e.g., organization and integration of information, see Fiorella & Mayer, 2014, 2016). For instance, Fiorella and Mayer (2014, Experiment 2) investigated the effects of preparing to explain (i.e., explaining-expectancy only) versus preparing and explaining on students' learning (i.e., explaining expectancy and instructional explaining). Students first read a text about the Doppler Effect either with the intention to be tested or to provide an oral instructional explanation about the learning contents to a fictitious student. Next, students either explained the learning contents or simply received additional study time. The authors demonstrated that explaining was more effective than restudying for students' acquisition of conceptual knowledge. In addition, they showed that students who were engaged in explaining outperformed students who only prepared to explain the learning materials (see also Hoogerheide et al., 2014).

Using videos as recording device during instructional explaining allows to capture both verbal and visual representations (e.g., gestures or visualizations, see Bobek & Tversky, 2016), which may additionally be conducive to learning. However, it has been shown that the effects of video-based explaining were particularly due to the verbalization during instructional explaining, as recent studies did not find any significant differences between video- and audio-based explaining regarding learning (Waldeyer, Moning, Heitmann, Hoogerheide, & Roelle, 2020; Wassenburg, de Koning, Koedinger, & Paas, 2020). An exception is provided by Fiorella and Kuhlmann (2020), as they found that the effectiveness of oral explaining improved when students were explicitly prompted to additionally generate visual representations. The benefits of providing instructional explanations to a fictitious student were also demonstrated in the meta-analytic review by Kobayashi (2018), who obtained a significant medium effect of instructional explaining $g = 0.48$.

An additional benefit of explaining is that it can help elicit metacognitive processes, which are conducive to enact effective cognitive strategies, as students externalize their knowledge which might allow them to monitor their current level of comprehension (see metacomprehension research: Fukaya, 2013; Lachner, Backfisch, Hoogerheide, van Gog, & Renkl, 2020). For instance, Fukaya (2013) showed that students who explained to a fictitious student showed higher levels of metacomprehension accuracy than students who only expected to explain or students who only produced keywords of the learning material (see also Jacob, Lachner, & Scheiter, 2020).

It has to be noted that research on instructional explaining has mostly used conceptual materials (e.g., expository texts), where the primary aim was conceptual understanding (e.g., Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014; Lachner et al., 2020). Explaining might particularly lend itself to conceptual learning (Rittle-Johnson and Loehr, 2017; Rittle-Johnson, Loehr, & Durkin, 2017). The primary aim of conceptual learning is to build a rich conceptual network by acquiring distinct concepts as well as by relating these concepts to each other and to previously acquired principles (Anderson, 2010; de Jong & Ferguson-Hessler, 1996). As explaining predominantly may trigger generating inferences and elaborations, it could help students integrate new concepts with their prior knowledge, and organize these concepts in a coherent knowledge representation (Fiorella & Kuhlmann, 2020; Lachner, Ly, & Nückles, 2018, see also Section 1.1.1).

1.1.1. *What drives the instructional explaining effect?*

Yet it remains an open question which underlying mechanism drives the instructional explaining effect. In the literature, there are three different views (Fiorella & Mayer, 2016; Hoogerheide, Visee, et al., 2019; Lachner et al., 2020). These views are not mutually exclusive, but rather provide different perspectives on the benefits of instructional explaining. The *retrieval hypothesis* postulates that the main effect of instructional explaining primarily occurs because a considerable amount of time during explaining is dedicated to retrieving the contents of the previously learned material from memory (Koh, Lee, & Lim, 2018; Lachner et al., 2020). Retrieving information from memory may foster learning through a consolidation function (Waldeyer et al., 2020), as retrieval intensifies potential retrieval cues (Rowland, 2014) and helps build up new retrieval cues as a function of spreading activation (Carpenter, 2009; Endres, Carpenter, Martin, & Renkl, 2017; Rowland, 2014).

The *generative hypothesis* postulates that explaining has benefits beyond mere retrieval because explaining additionally triggers students' inference-making processes and therefore leads to higher levels of generative processing (Fiorella & Mayer, 2016; Roscoe & Chi, 2008). For instance, explaining may incline students to monitor their current understanding (Fukaya, 2013; Lachner et al., 2020) and to elaborate on the material, which would help to actively make sense of the to-be-learned information (Fiorella & Mayer, 2016; Lachner et al., 2018; Ozuru, Briner, Best, & McNamara, 2010). Thus, the generative view claims that explaining activities may expand upon mere retrieval processes (Fiorella & Mayer, 2016), as the task to provide an explanation can trigger a restructuring of the content by drawing connections between concepts, or by connecting the material to prior knowledge by means of elaborations to a more pronounced extent than retrieval activities, which only require the student to retrieve the contents (see Endres et al., 2017, for enhancing retrieval practice by means of elaborative prompts).

The *social presence hypothesis* expands upon the generative view by stating that instructional explaining has additional benefits relative to self-explaining. Self-explanations commonly induce self-referential processing, as the student is required to explain the content to oneself. Contrarily, during instructional explaining, students have a fictitious communication partner in mind to whom they direct their explanations (Schober & Brennan, 2003), which might trigger distinct adaption processes (Clark & Brennan, 1991). For instance, students have to anticipate what the recipient of the explanation knows to adapt their explanation (Nickerson, 1999). In cases of instructional explaining to fictitious others, these anticipation processes work as a function of community co-membership (Schober & Brennan, 2003), as the recipient is not directly present and students have to infer the recipients' (likely lower) prior knowledge based on the explaining situations. These anticipation processes could engage students in specific audience-adjustments, and for instance lead students generate additional elaborations, in cases of lower anticipations of the recipients (see Wittwer, Nückles, Landmann, & Renkl, 2010, for empirical evidence). As such, instructional explaining may additionally contribute to students' learning, as compared to ego-centric self-explanations.

Empirical evidence for these different hypotheses is scarce, as explaining to fictitious others has mainly been compared to baseline conditions (e.g., restudying, see Fiorella & Mayer, 2013, 2014), but not to stronger control conditions that additionally involve retrieval (Koh et al., 2018; Lachner et al., 2020) or generative processes (e.g., Ainsworth & Loizou, 2003; Bisra, Liu, Nesbit, Salimi, & Winne, 2018; Roelle & Renkl, 2019).

One exception is the study by Rittle-Johnson, Saylor, and Swygert (2008). The authors investigated the effects of different explaining activities. After learning how to solve classification problems on mathematical patterns, children generated an explanation of the correct solution either to themselves (i.e., self-explanation) or to their mothers, or repeated the learning material out loud with the problem and solution still visible to them. The authors found that both explaining conditions outperformed children who had engaged in repetition on problems analogous to the learning phase ($d = 0.58$) and transfer problems ($d = 0.97$). With regards to the two explanation conditions, the authors found no performance difference on analogous problems ($d = 0.11$), but directing explanations to their mothers boosted transfer performance compared to self-explaining ($d = 0.70$). These findings provide evidence for the idea that explaining to someone else (even without interaction) is more effective than explaining to oneself, likely because the audience component triggered higher amounts of elaborative processes by distinct audience adjustments.

In a related study, Roscoe and Chi (2008) compared explaining to a fictitious peer student to self-explaining and to the interactive explanation activity of peer tutoring. The authors found that self-explaining was more beneficial for learning than explaining to fictitious others ($d = 1.86$). Self-explaining was also as effective as peer tutoring ($d = 0.39$). Additional content analyses of the provided explanations revealed that the self-explanations contained more elaborations than the instructional explanations provided to a fictitious student. Apparently, the higher levels of social presence during explaining to fictitious others (as compared to the self-explaining condition) did not necessarily result in more elaborated instructional explanations. However, this interpretation has to be treated with some caution, as the findings of Roscoe and Chi (2008) were potentially confounded by the timing of the explanations. That is, the self-explainers were told to continuously self-explain while studying the learning material, whereas the instructional explainers only had one opportunity to provide explanations at the end of the study phase (see also Lachner et al., 2020). Therefore, a potential explanation is that students in the self-explaining conditions had simply more time for explaining.

1.1.2. Is writing instructional explanations also an effective instructional strategy?

From a practical perspective, however, implementing oral explaining can be rather challenging, particularly for instructors which must assure a functioning technology environment to engage students in instructional explaining activities. Therefore, it is an open question whether providing instructional explanations to a fictitious student would also constitute an effective instructional strategy when done in writing. Positive evidence for writing explanations in general, can be found in the self-explaining literature, as several studies demonstrated positive effects of writing self-explanations on students' learning outcomes (e.g., Berthold & Renkl, 2009; Rau, Aleven, & Rummel, 2015; Roelle & Berthold, 2017; Roelle & Renkl, 2019, Rittle-Johnson et al., 2017; see Rittle-Johnson and Loehr, 2017 for a critical review). Therefore, drawing on the self-explaining literature, one may speculate that writing instructional explanations would also be an effective learning strategy. Empirical evidence can be found in the study by Larsen, Butler, and Roediger (2013). Using a within-subjects design, medical students participated in a teaching session comprising four different topics. In the subsequent learning sessions, students performed one of four written learning activities per topic crossing two factors (restudy versus retrieval, no-explaining, self-explaining). The authors obtained a main effect of retrieval ($\eta^2 = 0.33$) and self-explaining ($\eta^2 = 0.08$). Additional pairwise comparisons revealed that the self-explaining condition yielded better learning performance when combined with retrieval ($d = 0.70$), suggesting that self-explaining and retrieval may have additive effects regarding students' learning.

Contrarily to the literature on self-explaining, there is preliminary evidence that explaining to a fictitious student is not as effective when done in writing. For instance, Hoogerheide, Deijkers, Loyens, Heijltjes, and van Gog (2016) compared writing an instructional explanation to restudying learning material. Instructional explaining did not enhance learning outcomes compared to restudy. In Experiment 2, the authors directly compared written explaining, video-based explaining, and restudy. The authors found that video-based explaining ($d = 0.43$) was more effective than restudy, yet written explaining did not improve learning outcomes compared to restudy ($d = 0.19$). However, the authors did not find direct significant differences between written and video-based explaining.

Potential reasons why writing instructional explanations might not be as conducive to learning are mainly attributed to differences between generating oral and written explanations. First, writing instructional explanations may be regarded as a demanding activity that requires students to realize specific audience adjustments to make the explanations comprehensible for potentially less knowledgeable peer-students. Such audience adjustments may overload students, particularly in scenarios in which they are required to learn by writing, because explaining in writing typically places a high demand on our limited working memory resources (Lachner & Neuburg, 2019; Lachner & Nückles, 2015; Nückles, Hübner, & Renkl, 2009). Alternatively, from a perspective of pragmatic linguistics, deficits of writing instructional explanations could emerge due to differences of media constrains (Akinnaso, 1985; Clark & Brennan, 1991; Sperling, 1996). Writing, in contrast to speaking, is a non-spontaneous medium (Lakoff, 1982; Sindoni, 2014), which on the one hand allows for externalization of ideas and carefully reflecting upon one's thoughts (Klein et al., 2014; Lachner et al., 2018). On the other hand, due to the asynchronous character, writing instructional explanations evokes weaker feelings of social presence than oral discourse (Chafe, 1982; Chen, Park, & Hand, 2016; Sindoni, 2014). Indeed, several studies documented that oral explanations contained fewer personal references (1st and 2nd-personal pronouns), which are commonly associated with the perceived social presence during explaining (see Jacob et al., 2020). The lower levels of social presence may decrease the level of specific adaptions, such as elaborations during explaining, and at the same time decrease the effectiveness of writing instructional explanations (see Jacob et al., 2020; Lachner et al., 2018, for empirical evidence). These findings suggest that writing explanations is only beneficial when directed at oneself (i.e., self-explaining), not when directed at someone else.

1.2. The present study: writing self-explanations versus instructional explanations

Against this background, we conducted two experiments to examine the effects of instructional explaining to a fictitious student versus self-explaining and retrieval practice on students' learning in the context of learning-by-writing. On the one hand, it can be assumed that writing instructional explanations would be more effective than writing self-explanations and (written) retrieval practice, as additional audience adjustments may trigger additional generative processing (e.g., elaboration) which may be conducive to learning (see Rittle-Johnson et al., 2008, for empirical evidence on oral explaining). On the other hand, recent empirical research provided evidence that writing instructional explanations was not more effective than the rather poor control condition of restudy (Hoogerheide et al., 2016). Contrarily, such benefits have been demonstrated with the activity of self-explaining. Based on the available evidence, one might assume that instructional explaining would not be as advantageous as self-explaining.

To address these open research questions, in the two experiments, we realized a rigorous study design by comparing two written explaining conditions that varied in their social presence (i.e., instructional explaining to a fictitious other student versus self-explaining) to a retrieval practice condition, in which students were asked to recall the contents of the learning materials in written form (see Carpenter, 2009; Endres et al., 2017; Koh et al. for similar approaches). During these generative learning activities, the students had no learning material at hand, and therefore were required to retrieve the contents from memory. As an additional baseline condition, a fourth group of students completed a study-irrelevant puzzle task (Experiment 1) or did not receive an additional learning activity (Experiment 2). To draw legitimate recommendations for educational practice, in the present study, we combined well-controlled laboratory experimental between-participants approaches (Experiment 1) with field-experimental within-participants approaches (Experiments 2) to generalize our findings on writing explanations across contexts and domains. Additionally, to synthesize our findings with prior experimental research, we provide updated estimates of the effectiveness of instructional explaining by means of a continuously cumulating meta-analysis (CCMA, see Braver, Thoemmes, & Rosenthal, 2014; Morehead, Dunlosky, & Rawson, 2019).

2. Experiment 1

Experiment 1 was a laboratory study, in which we asked non-medical university students to learn from a medical text on the pathophysiology of bacterial endocarditis (an inflammation of the inner layer of the heart). Afterwards, students were randomly required to either a) provide a written explanation to a fictitious student (i.e., instructional explanation, Hoogerheide et al., 2016; Lachner et al., 2018), b) provide a written self-explanation (Rau et al., 2015; Roelle & Renkl, 2019), or c) recall the learning material in written form (Carpenter, 2009; Endres et al., 2017); d) a fourth group of students did not engage in a learning activity, but worked on a puzzle which was not related to the learning contents to keep the amount of tasks constant across conditions. Additionally, we explored distinct characteristics of the learning activities as well as mental effort ratings to draw inferences about the cognitive processes underlying the learning activities. Therefore, we followed recent research on oral explaining, and counted the number of elaborations and the level of completeness (as indicators for the level of generative processes; see Hoogerheide, Renkl, et al., 2019; Lachner et al., 2018), as well as the number of personal references within the explanations (as an indicator for the level of social presence, see Chafe, 1982; Hoogerheide et al., 2016; Lachner et al., 2018). Furthermore, metacomprehension ratings were applied to measure students' monitoring accuracy, as an indicator for the metacognitive processes (Fukaya, 2013; Jacob et al., 2020).

2.1. Method

2.1.1. Participants

University students ($N = 149$) from non-medical study programs of a German university participated in this study. We had to exclude two duplicates of students due to technical problems during the study (i.e., system crash). The average age of the tested sample ($N = 147$) was 24.22 ($SD = 5.70$). Thirty-three students were male. The students were in their eighth semester on average ($SD = 4.30$). All the students had very good German language skills. The high language proficiency was also reflected in students' reading skills (LGVT 6–12; see Schneider, Schlagmüller, & Ennemoser, 2007): $M = 20.45$; $SD = 8.16$, which corresponds to reading skills clearly above average (for more details, see materials section). Students received 12 euros for participating. We computed an a-priori power analysis for conducting an ANCOVA (4 conditions, 1 covariate) before running the study. The α-error was set to 0.05, and power to .80. Additionally, we assumed a medium effect of $\eta^2 = 0.075$, as empirical studies on oral explaining showed medium effects of explaining to a fictitious student versus self-explaining (Rittle-Johnson et al., 2008; Roscoe & Chi, 2008), and explaining to restudy on students' learning (Fiorella & Mayer, 2014; Hoogerheide, Visee, et al., 2019; Kobayashi, 2018). The power analysis suggested a minimum sample size of $N = 139$. Thus, the acquired sample size of 147 was good.

2.1.2. Design

The experiment had a one-factorial between-subjects design. Students were randomly assigned to one of four experimental conditions: 1) instructional explaining ($n = 38$) in which students explained the content of the materials to a fictitious female student named Martina, 2) self-explaining ($n = 37$) in which students explained the contents to themselves, 3) retrieval practice ($n = 29$) in which students were asked to recall the previously learned information, or 4) a control condition ($n = 43$) in which students solved a puzzle to keep the number of tasks constant across conditions.

2.1.3. Materials

The entire experiment was presented in the Qualtrics online survey tool (https://www.qualtrics.com).

2.1.3.1. Study text. The study text was an adapted German Wikipedia article on endocarditis, a cardiac disease which results due to an inflammation of the inner layer of the heart, and commonly involves the heart valves (https://de.wikipedia.org/wiki/Endokarditis). To gain a proper conceptual understanding of bacterial endocarditis, the text dealt with general information regarding the cardiovascular system, specifically about the construction of the endocardium. More importantly, the text covered common symptoms of endocarditis (e.g., fever, heart murmur) and potential causes of endocarditis (e.g., infection with bacteria). As such, the text comprised a complex medical topic, given that our participants were novices to this topic. The text comprised 648 words. To scaffold students' medical reasoning while learning from the text, the text included one schematic graphic to illustrate the anatomical structure and functions of the heart system.

2.1.3.2. Conceptual knowledge pretest and posttest. We used the conceptual knowledge test by Lachner and Nückles (2015) as pre- and posttest to measure students' conceptual knowledge regarding endocarditis. The test comprised eight multiple choice items (e.g., "What is the main cause of endocarditis"; "What is a possible prophylaxis for endocarditis?") with four answer possibilities and one correct solution (for more details, see Lachner & Nückles, 2015) which assessed conceptual understanding of bacterial endocarditis. To reduce the guess rate, we additionally introduced an answer option "I do not know" per question. The items were not confounded by ceiling effects, as the average item difficulty (i.e., the percentage of participants which correctly solved an item) was low both in the pretest and the posttest (pretest: 1.75%; posttest: 45.24%).

2.1.3.3. Transfer test. We used an adapted version of the transfer test by Lachner and Nückles (2015). The transfer test comprised three open questions (e.g., "Can endocarditis be the cause of a stroke?"; "Can endocarditis cause a cardiogenic shock?"), which required students to predict and explain possible consequences of endocarditis regarding possible related medical phenomena (i.e., co-morbid diseases). To assist the students' reasoning, they had a short definition of the possible comorbidities at hand. For each question, students could receive 7 points, resulting in a maximum total score of 21.20% of the transfer tasks were scored independently by two trained raters who were blind to the experimental conditions. Inter-rater reliability was good, $ICC = 0.86$ (Wirtz & Caspar, 2002).

2.1.3.4. Reading skills. To control for students' reading skills, we used the parallel version of the German reading and speed comprehension test (LGVT 6–12; Schneider et al., 2007). The test is conceptualized as a speed test and comprises a reading task with 25 gaps for which students had to decide which of three pre-given words had to be filled in. Students' reading skill was measured as the number of correctly answered gaps.

2.1.3.5. Mental effort. Students self-reported how much mental effort they had invested in studying the text, in the learning activity (depending on assigned condition), and in answering the posttest. The students rated their invested mental effort on a 9-point rating scale from 1 (very low effort) to 9 (very high effort, see Paas, 1992).

2.1.3.6. Metacomprehension accuracy. To investigate students' metacomprehension accuracy, students made prospective judgments of learning (after the learning activity) about their expected performance on the conceptual knowledge posttest, and retrospective judgments of learning after answering the posttest (e.g., Golke, Hagen, & Wittwer, 2018; Pierce & Smith, 2001). To obtain a baseline for students' overall metacomprehension skills, we additionally asked students to judge their expected performance after the reading assignment (i.e., before the intervention, see also Hertzog, Hines, & Touron, 2013, for similar approaches). Students estimated how many points they would achieve on the conceptual posttest (8 questions, one point), resulting in a scale from 0 to 8. Note that students already had a reference point (i.e., the pretest) upon which they could base their judgment (see also Kant, Scheiter, & Oschatz, 2017, for similar approaches).

We operationalized students' metacomprehension accuracy in terms of bias (see Baars, van Gog, de Bruin, & Paas, 2017; Prinz, Golke, & Wittwer, 2018, for recent applications). Bias refers to the signed difference between students' estimated number of correct answers and the actual number of correct answers (i.e., $X_{Judgment} - X_{Posttest}$). This approach allows for measuring students' over- and underestimation of their judged test performance. Positive values indicate that students overestimated their performance, negative values indicate an underestimation, and values of zero reflect accurate judgments.

2.1.4. Procedure

The students were tested in small groups in our laboratory (maximum: $n = 6$). The entire study was self-paced. At the beginning of the study, the students were informed that they would take part in a study on learning from medical introductory texts. They were instructed that after the study phase, they would engage in different learning activities which should help them understand the medical text. After providing written consent, the students were randomly assigned to the experimental conditions (i.e., instructional explaining, self-explaining, retrieval practice, baseline condition). Afterwards, all the students completed the pretest. Then, they studied the medical text. After studying the text, the students were required to indicate their invested mental effort and to give a judgment of learning. Subsequently, depending on experimental condition, they randomly completed one of the three different learning activities or the puzzle (baseline control condition). During the learning activities, they did not have access to the previous learning materials. For the instructional explaining condition, we used the following instruction which was frequently applied in previous studies on oral explaining to fictitious others (e.g., Hoogerheide et al., 2016, 2019b; Lachner et al., 2018, 2020):

> "Martina would like to train as a nurse in the local heart center. However, she has not yet dealt with cardiological diseases (such as endocarditis). Since Martina would like to know more about endocarditis, she asks you to write her an explanation about the central contents of the topic Endocarditis. Make sure to explain the content clearly and in sufficient detail so that Martina can understand your explanation well without using other materials. Enter your explanation into the free field."

The instruction in the instructional explaining condition required the students to explain the central contents in sufficient detail to a fictitious other student Martina. To raise the social awareness during explaining in line with previous studies, we added a small social scenario (see also Jacob et al., 2020; Lachner et al., 2020), and included specific information about Martina's professional background and her prior knowledge (see also Wittwer, Nückles, & Renkl, 2010, for related approaches in expert-novice communication). Besides this specific information, in line with previous studies, we increased students' distinct audience adjustments, by telling them to provide a comprehensible explanation.

The instruction in the self-explaining condition contained the identical requirements regarding the explaining task, but lacked the social component (see also Fiorella, Stull, Kuhlmann, & Mayer, 2020; Roelle & Nückles, 2019):

> Please write an explanation on the central contents of the topic Endocarditis. Make sure to explain yourself the content clearly and in sufficient detail. Enter your self-explanation into the free field.

The instruction in the retrieval condition contrarily lacked the explaining component, and required the students to recall the entire information of the text (see also Endres et al., 2017; Lachner et al., 2020). Additionally, as common recall tasks are generally non-guided and retrieval practice works as a function of concept activation irrespective of the judged importance of the repeated information (Endres et al., 2017), we did not prompt students to recall the *central* information. To increase the mere amount of recalled information the students were required to simply note down the recalled information to make the instruction distinct to the explaining conditions which commonly require more stylistic writing adjustments.

> Please recall the content of the text in written form. Write down everything you can remember from the text. Style and form do not matter. Enter your recall into the free field.

The baseline control condition contrarily received a non-related filler task to investigate the overall benefit of receiving additional learning activities. Therefore, the students were required to answer four small puzzle tasks (e.g., "The runner with the starting number 10 overtakes the competitor who is currently in 3rd place in an 800 m run. On which place is the runner with the number 10 after overtaking?").

After the learning activity, students judged their mental effort and provided a judgment of learning. Last, they answered the posttest (i.e., conceptual knowledge and transfer test), rated their invested mental effort, and provided a final judgment of learning (see Table 1. During the study, we collected students' time-on-task (e.g., during the learning activity) to explore potential differences of invested time during the learning tasks.

2.1.5. Analysis and coding

For the analyses of the quality of students' written learning activities (i.e., instructional explanation, self-explanation, retrieval practice), we rated their quality on three dimensions. First, we counted the number of *elaborations* (see Lachner et al., 2018). We determined an elaboration as a statement in which a student linked previous information of the study

Table 1
Conditions and materials used in experiment 1.

Instructional explaining	Self-explaining	Retrieval practice	Control condition
Pretest	Pretest	Pretest	Pretest
Study text	Study text	Study text	Study text
Mental effort/JoL	Mental effort/JoL	Mental effort/JoL	Mental effort/JoL
Explaining to a fictitious student	**Self-explaining**	**Retrieval practice**	**Puzzle**
Mental effort/JoL	Mental effort/JoL	Mental effort/JoL	Mental effort/JoL
Posttest	Posttest	Posttest	Posttest
Mental effort/JoL	Mental effort/JoL	Mental effort/JoL	Mental effort/JoL

Note. Bold items varied across experimental conditions.

text to her or his prior knowledge, for instance by including examples which were not present in the text, reporting one's own experiences, or making analogies (e.g., "For instance, cat bites can result in bacterial endocarditis, as bacteria can enter the blood stream"; "My uncle also had problems with the heart valves", "You have to imagine the heart as the motor of the body"). Again, 20% of the written learning activities were rated independently by two trained raters who were blind to the experimental conditions. Inter-rater reliability was good, $ICC = 0.78$ (Wirtz & Caspar, 2002). Thus, only one rater coded the rest of the explanations.

Additionally, we rated the *completeness* of the learning activities by a coding scheme, counting how many of the 10 concepts of the study text were covered in the learning artifacts (see also Hoogerheide, Renkl, et al., 2019; Lachner & Nückles, 2016, for related procedures). Again, 20% of the learning activities were rated independently by two trained raters. Inter-rater reliability was good, $ICC = 0.86$ (Wirtz & Caspar, 2002). Thus, only one rater coded the rest of the explanations.

Finally, as an indicator of the perceived social presence during explaining, we rated the number of *personal references*, that are first person pronouns (e.g., "I"; "my"; "we") and second person pronouns (e.g., "you", "your" "yours") in the explanations and retrieval protocols (see Hoogerheide et al., 2016; Jacob et al., 2020; Lachner et al., 2018). The number of personal references has been demonstrated to be a valid indicator of social presence, as it was significantly related to participants' judgements of social presence (Jacob et al., 2020). To systematically count the number of personal references, we used a computer script implemented in *R*, which automatically detected the number of personal references (Jacob et al., 2020).

2.2. Results

We applied an alpha level of 0.05 for all statistical analyses. We used partial η^2 (η_p^2) as an effect size measure, interpreting values $< .06$ as a small effect, values in the range between 0.06 and 0.14 as a medium effect, and values > 0.14 as a large effect (see Cohen, 1988). Table 2 provides the descriptive results of the study.

2.2.1. Preliminary analyses

ANOVAs showed no significant differences among experimental conditions regarding students' average prior knowledge, $F(3, 143) = 1.14$, $p = .337$, $\eta_p^2 = 0.023$, and their reading skills, $F(3, 143) = 0.04$, $p = .987$, $\eta_p^2 = 0.001$. Additional box-plot-analyses indicated that the dependent measures (i.e., conceptual knowledge, transfer) were not confounded by extreme outliers (as indicated by an asterisk, see Appendix A).

Table 2
Means and standard deviations of experiment 1.

Dependent Variable	Control	Retrieval	Self-explaining	Instructional explaining
Learning outcome				
Prior knowledge[a]	.01 (.04)	.01 (.03)	.02 (.03)	.02 (.06)
Conceptual knowledge[a]	.46 (.16)	.44 (.13)	.45 (.14)	.45 (.13)
Transfer[a]	.31 (.12)	.33 (.11)	.34 (.14)	.28 (.14)
Metacomprehension accuracy				
Bias (Prediction)[a]	.07 (.26)	.10 (.19)	.14 (.18)	.09 (.20)
Bias (Postdiction)[a]	-.04 (.17)	.03 (.17)	.05 (.15)	.01 (.17)
Mental effort				
during learning activity	6.75 (1.75)	6.64 (1.74)	6.03 (1.88)	6.61 (1.52)
during testing	6.39 (1.54)	6.72 (1.67)	6.41 (1.66)	6.58 (1.41)
Characteristics of the learning activities				
Completeness[a]	[b]	.37 (.15)	.34 (.14)	.35 (.15)
Elaborations	[b]	.03 (.17)	.05 (.23)	.05 (.23)
Personal references	[b]	.00 (.00)	.03 (.16)	.13 (.66)

[a] Values were transformed to proportions.
[b] The control condition was engaged in a task-irrelevant puzzle. Therefore, there are no values for the completeness and elaboration.

As time-on-task during the learning activity was not kept constant across conditions to provide a more natural learning setting, time-on-task differed across conditions, $F(3, 143) = 7.63$, $p < .001$, $\eta_p^2 = 0.138$. Additional post-hoc comparisons (Bonferroni) revealed that the two explaining conditions invested more time in the learning activity as compared to the baseline control condition (instructional explaining: $p < .001$; self-explaining: $p = .005$). None of the other comparisons were significant ($.086 < p < .999$).

2.2.2. Learning outcome

To test for potential differences in students' learning outcome, we computed two separate ANCOVAs with students' learning outcomes (i.e., conceptual knowledge, transfer) as dependent variables, and experimental conditions (i.e., instructional explaining, self-explaining, retrieval practice, control condition) as independent variable. Additionally, we controlled for students' prior knowledge. Regarding students' conceptual knowledge, contrary to our expectations, the ANCOVA was not significant, $F(3, 142) = 0.13$, $p = .940$, $\eta_p^2 = 0.003$, indicating that students did not differ regarding their conceptual knowledge across experimental conditions (see Table 2). Similarly, regarding students' transfer, the ANCOVA did not reach significance, $F(3, 142) = 1.95$, $p = .124$, $\eta_p^2 = 0.049$. These findings indicate that, although students in the explaining conditions invested more time during the learning activities, this additional time investment did not pay-off in higher learning outcomes, suggesting lower levels of efficiency.

2.2.3. Explorative analyses

2.2.3.1. Metacomprehension accuracy. To explore for differences between experimental conditions regarding students' metacomprehension accuracy, we conducted a repeated measures ANOVA with students' bias scores as dependent measure, test moment as the within-participants factor (i.e.; after the learning activity: prediction, and after the posttest: postdiction), and experimental condition as between-participants factor. We additionally controlled for students' initial bias scores after the reading phase by using them as covariates (see Hertzog et al., 2013; Lachner et al., 2020, for similar approaches). There was a main effect of test moment, $F(1, 142) = 6.18$, $p = .014$, $\eta_p^2 = 0.042$, and no interaction between experimental condition and test moment, $F(3, 142) = 0.40$, $p = .752$, $\eta_p^2 = 0.008$, indicating that generally students' metacomprehension accuracy increased between the learning activity and the knowledge test (see Table 2). Additionally, we found a main effect of experimental condition, $F(3, 142) = 2.75$, $p = .045$, $\eta_p^2 = 0.055$. However, although the descriptive findings of Table 2 indicated that particularly students in the instructional explaining condition achieved the most accurate metacomprehension judgments (see Table 2), none of the follow-up post-hoc tests (Bonferroni) approached significance ($p > .174$), likely due to the reduced test power of post-hoc comparisons.

2.2.3.2. Mental effort. We similarly proceeded for students' mental effort ratings and conducted a repeated measures ANOVA with students' reported mental effort ratings as dependent measure, test moment as the within-participants factor (i.e.; after the learning activity, and after the posttest), and experimental condition as between-participants factor. Students' perceived mental effort after the reading phase was taken as covariate. Neither the effect of test moment, $F(1, 142) = 2.41$, $p = .123$, $\eta_p^2 = 0.017$, nor the interaction between experimental condition and test moment, $F(3, 142) = 1.96$, $p = .123$, $\eta_p^2 = 0.040$ were significant (see

Table 2). Additionally, the main effect of experimental condition was not significant, F (3, 142) = 0.85, p = .470, $\eta_p^2 = 0.018$, indicating that the students invested comparable amounts of mental effort across conditions and across test moments.

2.2.3.3. Quality of the explanations. Finally, we tested for potential differences regarding the quality of the explanations and the retrieval protocols. As the non-significant findings regarding students' learning outcomes indicated, separate ANOVAs also revealed that there were no significant differences among experimental conditions, neither for the level of completeness, F (2, 108) = 0.59, p = .555, $\eta_p^2 = 0.011$, the number of personal references, F (2, 108) = 1.13, p = .329, $\eta_p^2 = 0.020$, nor for the level of elaboration, F (2, 108) = 0.18, p = .835, $\eta_p^2 = 0.003$.

2.3. Discussion

Contrarily to our hypotheses, we did not find significant differences among experimental conditions on students' learning outcomes, which was also reflected in the absence of differences on the quality features of the learning activities. This finding suggests that engaging students in additional explaining activities is not more effective than retrieval practice or than being engaged in a filler task unrelated to the learning content. We also found no differences among conditions on the reported effort invested in the experimental activities or on monitoring accuracy.

It is important to be cautious when making big claims on Experiment 1 alone, however. First, we have to note that we conducted a laboratory experiment with lay students, which had hardly any prior knowledge regarding the contents of the learning materials. Therefore, the students could have been overwhelmed by the task to learn from a medical text which was also reflected in the relatively high mental effort ratings and the low test-performance. Second, the variances regarding students' test performance within the experimental conditions were relatively high, as we selected non-medical students with different study backgrounds. This sampling procedure could unnecessarily have increased the inter-individual noise in our sample and could have reduced the chance to find an effect of our learning activities on learning. Therefore, the question remains whether the findings would change under different circumstances, for instance when students are more familiar with the study contents, when the materials are part of students' actual study programs, and when the interpersonal variance is lower (e.g., by means of within-participants comparisons, in which students serve as their own control).

3. Experiment 2

To address these issues, we conducted a field-experiment in an authentic pre-service teacher education course. The main topic of the course was educational technology. The course was a block course (full-time for two weeks). In preparation for the course, the students had to complete four reading assignments as homework before the block course started. In contrast to Experiment 1, we used a within-participants design. Thus, students randomly completed all of the four different learning activities (i.e., no-activity, retrieval practice, self-explaining, instructional explaining) spread over the different reading assignments which reduced the effects of potential inter-individual differences and likewise increased test power (see also Larsen et al., 2013, for related within-participants-studies in self-explaining). To avoid carry-over effects, the students were provided with a counterbalanced set of the four different learning activities by using the Latin square method (see also Lachner, Weinhuber, & Nückles, 2019; Wittwer & Ihme, 2014). As in Experiment 1, the entire experiment was conducted in a self-paced manner in the individual preparation phase of the block course. This procedure allowed us to test potential effects of instructional explaining in authentic individual learning contexts, such as homework assignments (see also Hoogerheide, Visee, et al., 2019, for related approaches). Because of the within-participants design, we obtained a nested data structure in which learning activities were naturally nested within students. To take the multi-level structure into account, we applied random coefficient models (Hox, 2010).

3.1. Method

3.1.1. Participants

An a-priori multilevel simulation study suggested sufficient power of 81% with 48 participants for finding medium effect sizes (see Appendix B). At the beginning of the course, 67 pre-service teacher students applied for the course and provided written consent to participate in the study. Fourteen students dropped the course during the study (as they did not hand in any assignment), which is rather common in pre-service teacher education. Three students did not hand in all assignments. Thus, the analyses are based on 50 students. The average age of the students was 25.10 (SD = 3.10). The students were in their ninth semester on average (SD = 2.66). In contrast to Experiment 1, the students had substantial prior knowledge, as they achieved on average 20% correct on the prior knowledge test (SD = 0.11). All the students were German native speakers.

3.1.2. Design

The within-participants design comprised students' learning outcomes (conceptual knowledge, transfer) as dependent variable, and the type of learning activity as within-participants factor: 1) instructional explaining, 2) self-explaining, 3) retrieval practice, 4) no-activity. Additionally, we controlled for students' prior knowledge, by including it as covariate. Again, we explored for differences regarding students' metacomprehension accuracy and their invested mental effort during the study activities.

3.1.3. Materials

The entire experiment was presented in the Qualtrics online survey tool (https://www.qualtrics.com). The students worked individually on the entire tasks at home.

3.1.3.1. Reading assignments. The students were required to read four different study texts, dealing with different conceptual topics in the domain of educational technology, as pre-class activity to prepare for the classroom sessions. All the study texts were instructional texts and written in German. On average the study texts comprised 17 pages (SD = 2.99). The first reading assignment was a book chapter by Niegemann et al. (2013) on cognitive load theory and the theory of multimedia learning. The main aim was to understand the central principles of cognitive load theory (e.g., information processing, limited working memory capacity assumption, type of loads, see Sweller, 2010), as well as the central concepts of the theory of multimedia learning (e.g., dual channel assumption, SOI-model, see Mayer, 2005). The second reading assignment (also taken from Niegemann et al., 2013) described the practical aspects of needs assessment for implementing educational technology (e.g., concept of problem analysis, methods of needs assessment, methods of knowledge analysis). The third reading assignment dealt with distinct learning technologies (Scheiter, 2015) and potential didactical implementations via the flipped classroom method (e.g., potential and risks of educational technology for learning, introduction of technology at schools). The fourth reading assignment was about the design of testing technologies for supporting formative assessment in the classroom (functions of assessment, types and formats of assessments, criteria for question design, see Niegemann et al., 2013).

3.1.3.2. Comprehensibility of the reading assignments. To test whether the perceived comprehensibility of the reading assignments was comparable across experimental conditions during reading the different study texts, we administered the questionnaire by Bromme, Jucks, and

Runde (2005, see also Lachner & Nückles, 2015, for recent applications). The students rated the reading assignments on a 5-point rating-scale ranging from 1 (= completely disagree) to 5 (= completely agree). The entire questionnaire comprised 21 items and four different scales. Intelligibility assessed to what extent common words were used and to what extent complex and long sentences were avoided (e.g., "The text contains familiar words"; "In the text, technical terms are explained"; Cronbach's α: 0.67). Organization assessed the quality of text cohesion of the reading assignments (e.g., "The text is structured"; "The text contains a red line"; Cronbach's α: 0.82). Shortness assessed the conciseness of the text (e.g., "The text is concise"; "The text is reduced to the actual message"; Cronbach's α: 0.49). Interestingness assessed students' affective value of the reading assignment (e.g., "The text is interesting"; "The text is appealing"; Cronbach's α: 0.88).

3.1.3.3. Prior knowledge test. A prior knowledge test was designed that comprised ten open-answer questions to measure students' prior knowledge across the four topics (e.g., "What is the theoretical assumption about human memory within cognitive load theory?"; "Describe the general design of flipped classroom instruction"; "Which disadvantages may open answer questions have?). A rater who was blind to the experimental conditions rated the students' answers to the open questions with the help of a standardized manual. For each answer, students could receive two points. A second rater coded 20% of the prior knowledge test, suggesting very good inter-rater reliability ($ICC = 0.94$).

3.1.3.4. Posttest. Four short open answer posttests were developed, one per reading assignment, to test students' conceptual knowledge with two open items (e.g., "What are the didactical functions of test-questions?"; "Which potentials can tablets have for teaching"). For each answer, students could receive two points. The transfer test comprised two open questions per reading assignment (e.g., "How can tablets be used to link in- and out-off school learning activities?"; "A peer-teacher plans a short presentation on evolution theory. During the presentation, she shows a funny picture of Charles Darwin. According to cognitive load theory, which consequences could the addition of that picture have for students' learning?"). All open questions on the transfer test required students to predict and explain possible consequences for potential teaching practices. For each answer, students could receive two points. 20% of the posttest tasks were scored independently by two trained raters who were blind to the experimental conditions, indicating very good inter-rater reliability ($ICC = 0.97$).

3.1.3.5. Mental effort. Like in Experiment 1, students rated how much mental effort they had invested in studying the texts and in the learning activity (Paas, 1992).

3.1.3.6. Metacomprehension accuracy. Additionally, we asked the students to make prospective judgments (after the study phase) and retrospective judgments of learning (after the posttest). Again, we operationalized students' metacomprehension accuracy in terms of bias. To keep the amount of judgments as parsimonious as possible, unlike to Experiment 1, students did not provide a judgment after the reading activity.

3.1.4. Procedure

The study consisted of one face-to-face session and four homework assignments. The beginning of the study started with a face-to-face session in which students were informed about the scope of the study and the study procedure. Students were informed that they were required to complete four reading assignments before attending the block course. Furthermore, they were informed that they would receive different learning activities to deepen their knowledge about their reading assignments. Afterwards, they provided written consent and answered the prior knowledge test. The reading assignments, the learning activities, and the knowledge tests were completed in individual homework sessions (see also Hoogerheide, Visee, et al., 2019, for related approaches). The homework assignments were completed in the Qualtrics online survey tool (https://www.qualtrics.com). To provide students with ample information about the assignment procedure, the students individually received information about how to accomplish the reading assignments plus learning activities and the knowledge tests:

> To prepare for the course, you need to complete four reading assignments. For each reading assignment, you will study a text and additionally engage in a learning task that will help you to elaborate on your knowledge. After each assignment, you will be provided with an open-answer knowledge test. You can plan your time independently. However, the assignments must be completed the evening before the first face-to-face session of the block course.

To access the homework tasks, the students received an individual link per homework assignment via the learning management system. The students were required to first complete the reading activity and then to start the additional homework assignment. First, students rated the comprehensibility of the reading assignment. Afterwards, they randomly received one of the four learning activities (i.e., no-learning activity, retrieval practice, self-explaining, instructional explaining). We used the identical instructions as in Experiment 1 (see Table 1).[1] Afterwards, students were required to report their invested mental effort and to provide a judgment of learning (i.e., prediction). Afterwards, they answered the posttest (i.e., conceptual questions, transfer questions), and provided a rating on their invested mental effort and a judgment of learning (i.e., postdiction). On average, the students spent 91.61 min ($SD = 279.91$) to complete the study. The type of learning activity was counterbalanced across the four reading assignments by using the Latin-Square method, so that the students accomplished one of the four learning activities randomly across the four reading assignments (see also Lachner et al., 2019).

3.2. Results

3.2.1. Preliminary analyses

There were no significant effects of the order of the learning activities on students' learning outcomes ($F < 1$). Similarly, students perceived the comprehensibility of the reading assignments to be comparable across conditions (intelligibility: $t\,(142.53) = 1.06$, $p = .293$; organization: $t\,(143.71) = 0.07$, $p = .944$; shortness: $t\,(141.79) = 0.87$, $p = .388$, interestingness: $t\,(142.31) = 0.10$, $p = .921$). Interestingly, duration times did not differ across learning activities ($0.238 < p < .999$). Additional box-plot-analyses indicated that the dependent measures (i.e., conceptual knowledge, transfer) were not confounded by extreme outliers (as indicated by an asterisk, see Appendix C).

3.2.2. Learning outcome

The descriptives can be seen in Table 3. As we conducted a within-participants design, the type of learning activity was nested within students. Therefore, we followed suggestions by Hox (2010) and applied random coefficient models to take the multi-level structure into account. We used the lme4-package in *R* and applied a varying-slope model to account for the nested data structure of our data (Hox, 2010). The

[1] As the topics between Experiment 1 and Experiment 2 differed, we used a slightly adapted instruction for the instructional explaining condition: Martina is a peer-student in the first term of her pre-service teacher studies. She is interested in the contents of the course, however, she could not enroll in the course. Since Martina would like to know more about the contents of this reading assignment, she asks you to write her an explanation of the central contents. Make sure to explain the content clearly and in sufficient detail, so that Martina can understand your explanation well without using other materials. Enter your explanation into the free field.

Table 3
Means and standard deviations of experiment 2.

Dependent Variable	Control	Retrieval	Self-explaining	Explaining to fictitious other
Learning outcome				
Conceptual knowledge[a]	.52 (.23)	.51 (.25)	.55 (.21)	.44 (.22)
Transfer[a]	.47 (.27)	.42 (.26)	.62 (.21)	.47 (.20)
Metacomprehension accuracy				
Bias (Prediction)[a]	.05 (.19)	.13 (.23)	.00 (.20)	.11 (.22)
Bias (Postdiction)[a]	.05 (.19)	.07 (.21)	.03 (.18)	.10 (.21)
Mental effort				
during learning activity	[b]	3.96 (1.41)	4.26 (1.71)	3.77 (1.80)
during testing	4.07 (1.75)	3.87 (1.60)	3.77 (1.49)	3.92 (1.57)

[a] Values were transformed to proportions.
[b]As the control condition was not provided with an additional learning activity, they did not rate their perceived mental effort during the learning activity.
[a] values were transformed to proportions.

models considered learning activities to be nested within students, so 'learning activity' represented Level 1 and 'students' represented Level 2. Learning activity and topic were included as fixed dummy-coded factors. The dependent variable comprised students' learning outcomes (i.e., conceptual knowledge and transfer). Additionally, we controlled for students' prior knowledge. As prior knowledge could vary across students, we allowed the slope of prior knowledge to vary by student, which finally resulted in the following equation: learning outcome = learning activity + topic + (1 + prior knowledge | student). To compare potential differences between the different learning activities, we used the emmeans-package in *R*. To counteract potential alpha-inflation, we used the Tukey-method in our pair-wise comparisons.

Regarding students' conceptual knowledge, in line with Experiment 1, none of the comparisons approached significance ($0.285 < p < .999$), indicating that students did not differ regarding their conceptual knowledge across learning activities (see Table 3). Regarding students' transfer, contrarily to Experiment 1, we found that self-explaining was more effective than our two control conditions: baseline control condition, t (145[2]) = 2.87, $p = .024$, $d = 0.59$; retrieval practice, t (143) = 4.26, $p < .001$, $d = 0.88$. More importantly, in line with Roscoe & Chi, 2008, self-explaining was more effective than instructional explaining, t (147) = 2.89, $p = .023$, $d = 0.59$. None of the other comparisons were significant ($0.489 < p < .999$). Therefore, in Experiment 2, self-explaining supported students' transfer, yet instructional explaining did not.

3.2.3. Explorative analyses

3.2.3.1. Metacomprehension accuracy. Regarding students' metacomprehension accuracy before the posttest (prediction accuracy), we found that none of the learning activities were more accurate than the control condition (retrieval versus control: t (136) = 2.53, $p = .059$; remaining comparisons: $p > .383$). Within the learning activities, we found that self-explaining contributed to more accurate judgments than instructional explaining, t (138) = 2.74, $p = .035$, $d = 0.57$, and than retrieval practice, t (136) = 3.67, $p = .002$, $d = 0.76$. Relatedly, regarding students' metacomprehension accuracy after the posttest (postdiction accuracy), we did not find significant differences among the learning activities ($0.443 < p < .969$). These findings suggest that, as in Experiment 1, instructional explaining did not contribute to students' metacomprehension accuracy.

3.2.3.2. Mental effort. As Table 3 indicated, there were no significant differences among the learning activities on the effort students reported that they had invested in the learning activity ($0.103 < p < .680$), or the knowledge test ($0.643 < p < .993$).

3.3. Discussion

The main finding of our field experiment was that self-explaining the learning contents was significantly better for students' transfer (but not conceptual knowledge) than providing a written explanation to a fictitious student. This finding confirms previous evidence by Roscoe and Chi (2008) on oral explanations, suggesting that providing written explanations may be most beneficial when students are required to provide self-explanations, but not when they are required to generate instructional explanations. Self-explaining was also more effective than our two control conditions (i.e., retrieval practice, no-activity). Apparently, in settings in which students possess higher levels of prior knowledge, self-explaining benefitted students' learning. Again, we did not obtain significant differences between the experimental conditions and the base-line condition regarding students' metacomprehension accuracy and their reported mental effort, suggesting that the effectiveness of self-explaining could not be explained by higher effort investments or more accurate metacomprehension judgments.

4. Continuously cumulating meta-analysis on instructional explaining

The obtained findings of our two experiments suggest that, contrarily to self-explaining, instructional explaining is not necessarily the optimal educational choice for supporting students' learning, at least when students are required to provide a written instructional explanation. Contrarily, in Experiment 2 self-explaining has been shown to be effective, likely because students possessed substantial prior knowledge. Given that our findings were not in accordance with previous evidence on instructional explaining (which was mainly realized as oral explanations), we performed a continuously cumulating meta-analysis (CCMA, see Braver et al., 2014; Morehead et al., 2019) to combine the evidence of the current and previous studies, on the effectiveness of explaining to fictitious students. Therefore, we entered all the studies (published, peer-reviewed, English) we were aware of which compared instructional explaining to a fictitious student after a study phase to a control condition, such as restudy or retrieval (based on a recent meta-analyses by Kobayashi, 2018, and a PsycInfo database search of the publication years 2013–2019[3]). During the review process, we additionally updated our CCMA with one study. We included a total of 12 articles comprising 18 experimental studies (see Table 4). As dependent variables, we encompassed students' conceptual knowledge and transfer. Following Borenstein, Hedges, Higgins, and Rothstein (2011), we used one standardized metric (g) based on the provided means and standard deviations of the single studies, to combine the different effect sizes of the studies. For between-subjects designs, we used the standardized mean difference (*SMD*) as outcome measure, as the studies included different types of knowledge assessments. For within-subjects-designs, based on Morris and DeShon (2002), we used the standardized mean change (SMCR) to compute the effect sizes. The included studies yielded 23 possible comparisons (6 on written explanations, 17 on oral explanations) regarding students' conceptual knowledge. For students' transfer, we could include 13 comparisons, as not all studies covered students' transfer in their experiments. To

[2] Degrees of freedom could slightly vary across learning activities due to missing values per topic.

[3] 2013 was taken as start date, as in that year the first study on instructional explaining to fictitious students was published by Fiorella and Mayer (2013).

Table 4
Effects of instructional explaining on students' conceptual knowledge.

Author	Standardized mean difference	95% CI lower limit	95% CI upper limit
Present Exp. 1 (written)	0.08	−0.41	0.56
Present Exp. 2 (written)	−0.31	−0.59	−0.03
Fiorella & Mayer, 2013, Exp. 1 (oral)	0.81	0.28	1.33
Fiorella & Mayer, 2014, Exp. 2 (oral)	0.55	−0.02	1.11
Fiorella, van Gog, Hoogerheide, & Mayer, 2017, Exp. 2 (oral)[a]	−0.13	−0.64	0.37
Fiorella et al., 2017, Exp. 2 (oral)[b]	−0.15	−0.66	0.36
Fiorella & Kuhlmann, 2020 (oral)	0.45	−0.07	0.96
Fukaya, 2013, Exp. 1 (oral)	0.93	0.12	1.74
Fukaya, 2013, Exp. 2 (oral)	0.57	−0.16	1.30
Hoogerheide et al., 2014, Exp. 1 (oral)	0.42	−0.13	0.98
Hoogerheide et al., 2014, Exp. 2 (oral)	0.87	0.36	1.38
Hoogerheide et al., 2016, Exp. 1 (written)	−0.06	−0.56	0.44
Hoogerheide et al., 2016, Exp. 2 (oral)	0.62	0.19	1.05
Hoogerheide et al., 2016, Exp. 2 (written)	0.39	−0.04	0.81
Hoogerheide, Renkl, et al., 2019 (oral)	0.43	−0.08	0.94
Hoogerheide, Visee, et al., 2019 (oral)	0.71	0.27	1.14
Jacob et al., 2020 (oral, easy text)	0.11	−0.34	0.56
Jacob et al., 2020 (oral, difficult text)	−0.07	−0.52	0.38
Jacob et al., 2020 (written, easy text)	−0.16	−0.61	0.30
Jacob et al., 2020 (written, difficult text)	−0.21	−0.67	0.24
Koh et al., 2018 (oral)	−0.11	−0.61	0.39
Lachner et al., 2020, Exp. 1 (oral)	−0.04	−0.55	0.47
Lachner et al., 2020, Exp. 2 (oral)	0.13	−0.37	0.62

Note.
[a] Before oral explaining/restudy, students watched a first-person perspective instructional video.
[b] Before oral explaining/restudy, students watched a third-person perspective instructional video.

conduct a random-effects meta-analysis, we used the *metafor*-package implemented in *R*. As the effects could have been affected by the modality of the instructional explaining activity, we additionally computed moderation analyses with the modality of explaining (written versus oral explaining).

Regarding students' conceptual knowledge, the meta-analysis resulted in a combined estimate based on 755 students in the explaining condition and 738 students in the control condition. The combined effect of explaining on students' conceptual knowledge was small, yet significant: $g = 0.222$, 95% CI [0.064, 0.380], $p = .006$ (see Table 4, for the single effect sizes). The heterogeneity index was significant, Q (22) $= 56.13$, $p < .001$, indicating that there was considerable heterogeneity among the studies. The moderation effect of explaining modality on students' conceptual knowledge was also significant, $QM(1) = 6.42$, $p =$ .011, indicating that the effectiveness of instructional explaining depended on the modality of explaining. Separate meta-analyses indicated a significant effect of oral explaining, $g = 0.336$, 95% CI [0.158, 0.513], $p < .001$, but no significant effect for writing explanations, $g = -0.070$, 95% CI [-0.292, 0.152], $p = .535$, indicating that instructional explaining was only superior to a baseline condition when it was given in oral form (see Table 4 for the effect sizes).

Regarding students' transfer, the meta-analysis resulted in a combined estimate based on 483 students in the explaining condition and 460 students in the control condition. Again, the combined effect of explaining on students' transfer was small, but not significant, $g = 0.155$, 95% CI [-0.026, 0.335], $p = .093$ (see Table 5, for the single effect sizes). The heterogeneity index was significant, Q (12) $= 24.51$, $p = .017$, indicating that the samples were rather heterogeneous regarding their transfer performance in our meta-analysis. The moderation effect of explaining modality on students' transfer was not significant, $QM(1) = -1.461$, $p = .227$, indicating that the effect of instructional explaining on students' transfer did not depend on the modality of explaining.

5. General discussion

We conducted two experiments to examine the effects of instructional explaining to a fictitious student versus self-explaining and retrieval practice on students' learning in the context of writing explanations. In Experiment 1, there were no significant differences among experimental conditions on learning outcomes. Additionally, there were no differences regarding the characteristics of the different learning activities (i.e., personal references, completeness, elaboration) among conditions. Apparently, our explaining manipulation did not inevitably evoke higher levels of generative processing or social presence, suggesting that the explaining manipulation did not have the intended effect. We attribute the non-significant findings to the fact that students had insufficient prior knowledge to provide high-quality explanations. This effect could have been increased by the fact that students provided written and not oral explanations, which likely triggered lower levels of social presence. In Experiment 2, contrarily, we found that only self-explaining, but not explaining to a fictitious student or written retrieval practice, enhanced students' transfer (but not conceptual knowledge) compared to a control condition. Apparently, instructional explaining did not contribute to students' understanding, whereas self-explaining did. One potential explanation of the somewhat differential findings of Experiment 1 and Experiment 2 is the distinct experimental settings in which we realized our study. Experiment 1 was a laboratory study with lay students having hardly any prior knowledge about the learning contents. Contrarily, in Experiment 2, the students had more prior knowledge, which could have been necessary to effectively use self-explaining as a generative activity to integrate the new information

Table 5
Effects of instructional explaining on students' transfer.

Author	Standardized mean difference	95% CI lower limit	95% CI upper limit
Present Exp. 1 (written)	−0.39	−0.87	0.10
Present Exp. 2 (written)	0.25	−0.03	0.52
Fiorella & Kuhlmann, 2020 (oral)	0.75	0.23	1.28
Hoogerheide et al., 2016, Exp. 1 (written)	0.21	−0.29	0.71
Hoogerheide et al., 2016, Exp. 2 (written)	0.13	−0.29	0.56
Hoogerheide et al., 2016, Exp. 2 (oral)	0.31	−0.12	0.74
Hoogerheide, Renkl, et al., 2019 (oral)	0.51	0.00	1.02
Jacob et al., 2020 (oral, easy text)	−0.20	−0.65	0.25
Jacob et al., 2020 (oral, difficult text)	0.51	0.06	0.97
Jacob et al., 2020 (written, easy text)	−0.45	−0.91	0.01
Jacob et al., 2020 (written, difficult text)	0.19	−0.26	0.65
Lachner et al., 2020, Exp. 1 (oral)	0.00	−0.51	0.51
Lachner et al., 2020, Exp. 2 (oral)	0.20	−0.30	0.70

of the learning materials within their prior understanding (see Fiorella & Mayer, 2016). This interpretation however requires more direct examinations, for instance by including prior knowledge as moderator (e.g., Hoogerheide, Renkl, et al., 2019), as the study design between the two experiments also differed (within-versus between-subjects-design). Interestingly, neither retrieval practice nor instructional explaining outperformed the control condition. A possible explanation might be that we used an immediate posttest, while generative activities might be most conducive to learning outcomes after a delay (Fiorella & Mayer, 2016; Rowland, 2014). While this is true for retrieval practice, this delayed effect is typically not the case for (instructional) explaining (see Fiorella & Mayer, 2016). For example, various studies found beneficial effects of providing (oral) explanations to a fictitious fellow student on an immediate posttest *and* a delayed posttest (e.g., Fiorella & Mayer, 2014; Hoogerheide et al., 2014).

The finding that written self-explaining was more effective for transfer (though not for conceptual knowledge) than written instructional explaining, however, deserves more attention. At first glance, it is surprising that teaching a fictitious peer student was so ineffective (also compared to the weak control condition), because many prior studies did find beneficial effects of instructional explaining (e.g., Hoogerheide et al., 2016; 2019a). A likely explanation is that writing explanations might simply not be as beneficial for students' understanding as oral explaining. This idea was explored via a cumulating meta-analysis, which showed an overall small, yet significant effect of instructional explaining on conceptual knowledge ($g = 0.22$). Additional moderation analyses indeed suggested that the effect of instructional explaining was only significant when the studies were included that had an oral explaining condition. For the written explaining studies, there was no overall effect on conceptual knowledge or transfer. Together, these findings indicate that instructional explaining does not necessarily contribute to students' understanding when it is realized as a writing task. Furthermore, they suggest that in written contexts, asking students to self-explain the learning contents may be more effective to attain conceptual understanding (see also Rittle-Johnson et al., 2017; for related findings).

So what do our findings say about the theoretical underpinnings of explaining to a fictitious student? Our finding of Experiment 2, that retrieval practice was not more effective than a baseline control condition, but self-explaining was, provides evidence for the generative hypothesis, as explanation effects are not only due to the retrieval practice that is often inherent to explanation activities, but mainly due to generative processing. These findings are in line with recent suggestions by Waldeyer, Heitmann, Moning, and Roelle (2020) who proposed that retrieval practice may rather constitute a consolidation activity to strengthen existing knowledge, but not to construct new knowledge by active knowledge integration. However, it has to be noted that the explanations had also been generated without the content being available. Therefore, it cannot be ruled out that the findings were solely due to generative processing. A potential remedy would be to analyze the characteristics of the products of the learning activities (i.e., retrieval, self-explaining, instructional explaining) in Experiment 2. We consciously decided against analyzing the learning activities, as we realized a within-participants-design comprising learning materials of four relatively heterogeneous topics. Therefore, our materials mirrored a representative but less controlled set of different instructional texts, making it difficult to draw legitimate conclusions about the characteristics of the single explanations (see also Larsen et al., 2013, for related considerations regarding within-participants designs). Future studies should therefore replicate our findings in laboratory settings with a between-participants-design (such as in Experiment 1), but with students that possess more prior knowledge about the topic.

Secondly, the finding that explaining to a (fictitious) audience impaired transfer relative to explaining to oneself points towards adjustments of the social presence hypothesis. Although the focus in the literature has been on the positive effects that feelings of social presence elicited by the audience might have (e.g., Hoogerheide et al., 2016), it is not unimaginable that an (imagined) audience could be detrimental to learning outcomes. The additional cognitive (e.g., making specific audience adjustments; Lachner & Neuburg, 2019) and affective demands (e.g., arousal and worrying thoughts; Hoogerheide, Renkl, et al., 2019) of addressing a (fictitious) audience could overload students' working memory resources, particularly when students' knowledge before explaining is still limited. In social psychology research, it is well-established that the mere presence of an actual or imaginary audience could foster task performance when expertise is high and hinder performance when expertise is low (see social facilitation research; Park & Catrambone, 2007; Wolf, Bazargani, Kilford, Dumontheil, & Blakemore, 2015). A caveat to this interpretation is that the hypothesized higher level of extraneous processing was not directly reflected in students' subjective ratings of mental effort, which is commonly considered as a coarse proxy of students' cognitive load during learning (Paas, 1992; Hoogerheide et al., 2014, 2016). Nevertheless, our findings may inform research testing the social presence hypothesis, as they may be suggestive of ways that the effects of social presence while explaining may be moderated by additional variables, such as prior knowledge or the modality of the explanations.

5.1. Limitations and future research

An important strength of our study is the combination of a laboratory-oriented and a field-oriented experiment with different learning materials, which allowed us to generalize our findings on instructional explaining across contexts and domains and make potential recommendations for educational practice (Renkl, 2013). Another important strength is the inclusion of a continuously cumulating meta-analysis, which allowed us to explore the overall effect of instructional explaining and its dependency on the modality in which the explanations were provided.

There are also some limitations to address. A critical caveat refers to the generalizability of our study. First, our findings only hold true for situations where the learning activities take place after an initial study phase. Given that the effectiveness of explaining to oneself or to a fictitious student might increase when students provide the explanations continuously during the study phase or earlier on in the learning phase (Bisra et al., 2018; Lachner et al., 2020), future research should investigate whether our findings replicate when the timing of explaining is different. For instance, beneficial effects of continuous explaining on learning could be assumed, as students would have more opportunities to explain and continuously elaborate their explanations (see Rau, Aleven, & Rummel, 2013; Rohrer, Dedrick, Hartwig, & Cheung, 2019). At the same time, the opportunity of continuous explaining would reduce the cognitive demands during explaining, as students would only be required to retrieve and explain distinct passages of the learning material. Besides the timing, it has also to be noted that our findings are restricted to generic explaining activities, as we used rather distal prompts to induce our explaining activities (e.g., "Please write an explanation on the central contents of the topic Endocarditis"), which may have impaired the effectiveness of learning-by-explaining (Rittle-Johnson and Loehr, 2017). Distal prompts may be less effective in eliciting content-specific explaining strategies than specific prompts, which are directly aligned with the content (e.g., generate self-explanations about the underlying principles of the worked-out solution steps; Atkinson, Renkl, & Merrill, 2003; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Rittle-Johnson et al., 2017). That said, it should be noted that Bisra et al. (2018) demonstrated in their meta-analysis that the effectiveness of self-explaining was not moderated by the specificity of the explaining prompt. Nevertheless, future research should replicate our findings with more specific prompts that are adapted to the content of the learning material, which is common in self-explaining research to help students go beyond the topic of the text by generating elaborations and inferences. Interestingly, in most instructional explaining research, only more general/distal prompts have been used.

Finally, because we only had written explanation conditions, it is unclear whether the results regarding self-explaining vs. instructional explaining would replicate when students provide the explanations orally. There are various reasons why one might expect different results. For instance, oral explaining may require fewer cognitive resources yet elicit higher levels of social presence than writing explanations (Hoogerheide et al., 2016). At the same time oral explaining could trigger higher levels of motivation which enables students to engage more in their explanations (Hoogerheide, Visee, et al., 2019). Therefore, future studies should test whether our findings, particularly regarding the differences between self-explaining and instructional explaining, would remain stable or even diminish when students are required to provide oral instead of written explanations.

5.2. Conclusion

All in all, our findings provide a promising starting point for further research on the effects of writing explanations on students' learning. Although writing explanations is a frequent method in educational practice, as it is a feasible generative activity in classroom settings, our findings question the assumed advantages of instructional explaining in writing. Nevertheless, our findings suggest that if written explaining activities are implemented, self-explaining may be the better alternative than instructional explaining.

Author statement

Andreas Lachner: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Leonie Jacob: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Vincent Hoogerheide: Conceptualization, Methodology, Writing – original draft, Writing – review & editing

Acknowledgements

Data and analysis scripts can be viewed under doi: 10.17605/OSF.IO/CZHTN. We would like to thank Louisa Döderlein, Eleonora Dolderer, and Anna Rosenträger for their assistance with many practical aspects during conducting the experiments. The research reported in this article was supported by the Federal Ministry of Education and Research in Germany (BMBF) under contract number 01JA1611.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2020.101438.

References

Ainsworth, S., & Loizou, A. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*(4), 669–681. https://doi.org/10.1207/s15516709cog2704_5

Akinnaso, F. N. (1985). On the similarities between spoken and written language. *Language and Speech, 28*(4), 323–359. https://doi.org/10.1177/002383098502800401

Anderson, J. R. (2010). *Cognitive psychology and its implications* (7th ed.). New York: Worth Publishing.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Baars, M., van Gog, T., de Bruin, A., & Paas, F. (2017). Effects of problem solving after worked example study on secondary school children's monitoring accuracy. *Educational Psychology, 37*(7), 810–834. https://doi.org/10.1080/01443410.2016.1150419

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

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review, 30*, 703–725. https://doi.org/10.1007/s10648-018-9434-x

Bobek, E., & Tversky, B. (2016). Creating visual explanations improves learning. *Cognitive Research: Principles and Implications, 1*, 27. https://doi.org/10.1186/s41235-016-0031-6

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Chichester, GB: John Wiley & Sons.

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9*(3), 333–342. https://doi.org/10.1177/1745691614529796

Bromme, R., Jucks, R., & Runde, A. (2005). Barriers and biases in computer-mediated expert-layperson-communication: An overview and insights into the field of medical advice. In R. Bromme, F. W. Hesse, & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication – and how they may be overcome* (pp. 89–118). New York: Springer.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563–1569. https://doi.org/10.1037/a0017021

Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35–53). Norwood, NJ: Ablex.

Chen, Y. C., Park, S., & Hand, B. (2016). Examining the use of talk and writing for students' development of scientific conceptual knowledge through constructing and critiquing arguments. *Cognition and Instruction, 34*(2), 100–147. https://doi.org/10.1080/07370008.2016.1145120

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

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). American Psychological Association. https://doi.org/10.1037/10096-006.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction, 49*, 13–20. https://doi.org/10.1016/j.learninstruc.2016.11.010

Fiorella, L., & Kuhlmann, S. (2020). Creating drawings enhances learning by teaching. *Journal of Educational Psychology, 112*(4), 811–822. https://doi.org/10.1037/edu0000392

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*(4), 281–288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology, 39*(2), 75–85. https://doi.org/10.1016/j.cedpsych.2014.01.001

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2020). Fostering generative learning from video lessons: Benefits of instructor-generated drawings and learner-generated explanations. *Journal of Educational Psychology, 112*(5), 895–906. https://doi.org/10.1037/edu0000408

Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology, 109*(5), 653–665. https://doi.org/10.1037/edu0000161

Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning, 8*(1), 1–18. https://doi.org/10.1007/s11409-012-9093-0

Golke, S., Hagen, R., & Wittwer, J. (2018). Lost in narrative? The effect of informative narratives on text comprehension and metacomprehension accuracy. *Learning and Instruction, 60*, 1–19. https://doi.org/10.1016/j.learninstruc.2018.11.003

Hertzog, C., Hines, J. C., & Touron, D. R. (2013). Judgments of learning are influenced by multiple cues in addition to memory for past test accuracy. *Archives of Scientific Psychology, 1*(1), 23. https://doi.org/10.1037/arc0000003

Hoogerheide, V., Deijkers, L., Loyens, S. M., Heijltjes, A., & van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology, 44*, 95–106. https://doi.org/10.1016/j.cedpsych.2016.02.005

Hoogerheide, V., Loyens, S. M., & van Gog, T. (2014). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction, 33*, 108–119. https://doi.org/10.1016/j.learninstruc.2014.04.005

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology, 111*, 45–56. https://doi.org/10.1037/edu0000272

Hoogerheide, V., Visee, J., Lachner, A., & van Gog, T. (2019). Generating an instructional video as homework activity is both effective and enjoyable. *Learning and Instruction, 64*, 101226. https://doi.org/10.1016/j.learninstruc.2019.101226

Hox, J. J. (2010). *Quantitative methodology series. Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY, US: Routledge/Taylor & Francis.

Jacob, L., Lachner, A., & Scheiter, K. (2020). Learning by explaining orally or in written form? Text difficulty matters. *Learning and Instruction, 68*, 101344. https://doi.org/10.1016/j.learninstruc.2020.101344

de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105–113. https://doi.org/10.1207/s15326985ep3102_2

Kant, J., Scheiter, K., & Oschatz, K. (2017). How to sequence video modeling examples and inquiry tasks to foster scientific reasoning. *Learning and Instruction, 52*, 46–58. https://doi.org/10.1016/j.learninstruc.2017.04.005

Klein, P., Boscolo, P., Kirkpatrick, L., & Gelati, C. (Eds.). (2014). *Writing as a learning activity*. Leiden: Brill.

Kobayashi, K. (2018). Learning by preparing-to-teach and teaching: A meta-analysis. *Japanese Psychological Research, 61*, 192–203. https://doi.org/10.1111/jpr.12221

Koh, A. W. L., Lee, S. C., & Lim, S. W. H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology, 32*(3), 401–410. https://doi.org/10.1002/acp.3410

Lachner, A., Backfisch, I., Hoogerheide, V., van Gog, T., & Renkl, A. (2020). Timing matters! Explaining between study phases enhances students' learning. *Journal of Educational Psychology, 112*(4), 841–853. https://doi.org/10.1037/edu0000396

Lachner, A., Ly, K.-T., & Nückles, M. (2018). Providing written or oral explanations? Differential effects of the modality of explaining on students' conceptual learning and transfer. *The Journal of Experimental Education, 86*(3), 344–361. https://doi.org/10.1080/00220973.2017.1363691

Lachner, A., & Neuburg, C. (2019). Learning by writing explanations: Computer-based feedback about the explanatory cohesion enhances students' transfer. *Instructional Science, 47*(1), 19–37. https://doi.org/10.1007/s11251-018-9470-4

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. https://doi.org/10.1037/xap0000038

Lachner, A., & Nückles, M. (2016). Tell me why! Content knowledge predicts process-orientation of math researchers' and math teachers' explanations. *Instructional Science, 44*(3), 221–242. https://doi.org/10.1007/s11251-015-9365-6

Lachner, A., Weinhuber, M., & Nückles, M. (2019). To teach or not to teach the conceptual structure of mathematics? Teachers undervalue the potential of principle-oriented explanations. *Contemporary Educational Psychology, 58*, 175–185. https://doi.org/10.1016/j.cedpsych.2019.03.008

Lakoff, R. T. (1982). Some of my favorite writers are literate: The mingling of oral and literate strategies in written communication. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 239–260). Norwood, NJ: Ablex.

Larsen, D. P., Butler, A. C., & Roediger, H. L., III (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education, 47*(7), 674–682. https://doi.org/10.1111/medu.12141

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.004.

Morehead, K., Dunlosky, J., & Rawson, K. A. (2019). How much mightier is the pen than the keyboard for note-taking? A replication and extension of mueller and oppenheimer (2014). *Educational Psychology Review*, 1–28. https://doi.org/10.1007/s10648-019-09468-2

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105. https://doi.org/10.1037/1082-989X.7.1.105

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*(6), 737–759. https://doi.org/10.1037/0033-2909.125.6.737

Niegemann, H. M., Hessel, S., Hochscheid-Mauel, D., Aslanski, K., Deimann, M., & Kreuzberger, G. (2013). *Kompendium E-learning*. Berlin, Germany: Springer.

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*(3), 259–271. https://doi.org/10.1016/j.learninstruc.2008.05.002

Okita, S. Y., & Schwartz, D. L. (2013). Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *The Journal of the Learning Sciences, 22*(3), 375–412. https://doi.org/10.1080/10508406.2013.807263

Ozuru, Y., Briner, S., Best, R., & McNamara, D. S. (2010). Contributions of self-explanation to comprehension of high- and low-cohesion texts. *Discourse Processes, 47*(8), 641–667. https://doi.org/10.1080/01638531003628809

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology, 84*, 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1

Park, S., & Catrambone, R. (2007). Social facilitation effects of virtual humans. *Human Factors, 49*(6), 1054–1060. https://doi.org/10.1518/001872007X249910

Pierce, B. H., & Smith, S. M. (2001). The postdiction superiority effect in metacomprehension of text. *Memory & Cognition, 29*(1), 62–67. https://doi.org/10.3758/BF03195741

Plötzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). *Learning by explaining to oneself and to others. Collaborative learning: Cognitive and computational approaches*. Oxford, UK: Elsevier.

Prinz, A., Golke, S., & Wittwer, J. (2018). The double curse of misconceptions: Misconceptions impair not only text comprehension but also metacomprehension in the domain of statistics. *Instructional Science, 46*, 723–765. https://doi.org/10.1007/s11251-018-9452-6

Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction, 23*, 98–114. https://doi.org/10.1016/j.learninstruc.2012.07.003

Rau, M. A., Aleven, V., & Rummel, N. (2015). Successful learning with multiple graphical representations and self-explanation prompts. *Journal of Educational Psychology, 107*(1), 30–46. https://doi.org/10.1037/a0037211

Renkl, A. (1995). Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. *Learning and Instruction, 5*(1), 21–36. https://doi.org/10.1016/0959-4752(94)00015-H

Renkl, A. (2013). Why practice recommendations are important in use-inspired basic research and why too much caution is dysfunctional. *Educational Psychology Review, 25*(3), 317–324. https://doi.org/10.1007/s10648-013-9236-0

Rittle-Johnson, B., & Loehr, A. M. (2017). Eliciting explanations: Constraints on when self-explanation aids learning. *Psychonomic Bulletin & Review, 24*(5), 1501–1510. https://doi.org/10.3758/s13423-016-1079-5

Rittle-Johnson, B., Saylor, M., & Swygert, K. E. (2008). Learning from explaining: Does it matter if mom is listening? *Journal of Experimental Child Psychology, 100*(3), 215–224. https://doi.org/10.1016/j.jecp.2007.10.002

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM, 49*(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z.

Roelle, J., & Berthold, K. (2017). Effects of incorporating retrieval into learning tasks: The complexity of the tasks matters. *Learning and Instruction, 49*, 142–156. https://doi.org/10.1016/j.learninstruc.2017.01.008

Roelle, J., & Nückles, M. (2019). Generative learning versus retrieval practice in learning from text: The cohesion and elaboration of the text matters. *Journal of Educational Psychology, 111*(8), 1341–1361. https://doi.org/10.1037/edu0000345

Roelle, J., & Renkl, A. (2019). Does an option to review instructional explanations enhance example-based learning? It depends on learners' academic self-concept. *Journal of Educational Psychology, 112*, 131–147. https://doi.org/10.1037/edu0000365

Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C. N. (2019). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology, 112*, 40–52. https://doi.org/10.1037/edu0000367

Roscoe, R. D. (2014). Self-monitoring and knowledge-building in learning by teaching. *Instructional Science, 42*(3), 327–351. https://doi.org/10.1007/s11251-013-9283-4

Roscoe, R. D., & Chi, M. T. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science, 36*(4), 321–350. https://doi.org/10.1007/s11251-007-9034-5

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. https://doi.org/10.1037/a0037559

Scheiter, K. (2015). Besser lernen mit dem tablet? Praktische und didaktische potenziale sowie anwendungsbedingungen von Tablets im unterricht. [Learning better with the tablet? Practical and didactical potentials of tablets for teaching]. In H. Buchen, L. Horster, & H.-G. Rolff (Eds.), *Schulleitung und Schulentwicklung* (3.eds, pp. 1–14). Stuttgart, Germany: Raabe-Verlag.

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *LGVT 6-12 Lesegeschwindigkeits- und Verständnistest für die Klassen 6-12*. Germany: Hogrefe: Göttingen.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123–164). Lawrence Erlbaum Associates Publishers.

Sindoni, M. G. (2014). *Spoken and written discourse in online interactions: A multimodal approach*. New York: Routledge.

Sperling, M. (1996). Revisiting the writing-speaking connection: Challenges for research on writing and writing instruction. *Review of Educational Research, 66*(1), 53–86. https://doi.org/10.3102/00346543066001053

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.004.

Waldeyer, J., Heitmann, S., Moning, J., & Roelle, J. (2020). Can generative learning tasks be optimized by incorporation of retrieval practice? *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1016/j.jarmac.2020.05.001. Advance online publication.

Waldeyer, J., Moning, J., Heitmann, S., Hoogerheide, V., & Roelle, J. (2020). *Does learning by teaching have double-edged effects?*. in preparation.

Wassenburg, S., de Koning, B., Koedinger, K., & Paas, F. (2020). *Limits of learning by teaching: Explaining to self versus explaining to others*. in preparation.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87*(3), 406–423. https://doi.org/10.1037/0022-0663.87.3.406

Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und ratingskalen* [interrater agreement und interrater reliability: Methods for calculating and improving the reliability of ratings by category systems and rating scales]. *Göttingen: Hogrefe*.

Wittrock, M. C. (2010). Learning as a generative process. *Educational Psychologist, 45*(1), 40–45. https://doi.org/10.1080/00461520903433554

Wittwer, J., & Ihme, N. (2014). Reading skill moderates the impact of semantic similarity and causal specificity on the coherence of explanations. *Discourse Processes, 51*, 143–166. https://doi.org/10.1080/0163853X.2013.855577

Wittwer, J., Nückles, M., & Renkl, A. (2010). Using a diagnosis-based approach to individualize instructional explanations in computer-mediated communication. *Educational Psychology Review, 22*, 9–23. https://doi.org/10.1007/s10648-010-9118-7

Wolf, L. K., Bazargani, N., Kilford, E. J., Dumontheil, I., & Blakemore, S. J. (2015). The audience effect in adolescence depends on who's looking over your shoulder. *Journal of Adolescence, 43*, 5–14. https://doi.org/10.1016/j.adolescence.2015.05.003