American Psychological Association
Journal of Educational Psychology

0022-0663/20/$12.00
2020, Vol. 112, No. 4, 841–853
http://dx.doi.org/10.1037/edu0000396

Timing Matters! Explaining Between Study Phases Enhances Students' Learning

Andreas Lachner
University of Tübingen and Leibniz-Institut für Wissensmedien, Tübingen, Germany

Iris Backfisch
Leibniz-Institut für Wissensmedien, Tübingen, Germany

Vincent Hoogerheide
Utrecht University and University of Freiburg

Tamara van Gog
Utrecht University and University of Tübingen

Alexander Renkl
University of Freiburg

Previous research has shown that explaining is an effective activity to enhance learning. In prior studies, students were instructed to explain the contents after completing an entire learning phase. Explaining at the end of a learning phase, however, may be less apt to support comprehension monitoring and subsequent regulation activities. In 2 experiments, we investigated whether explaining in earlier phases of studying (i.e., in-between explaining) would foster learning more than explaining after the entire study phase (i.e., afterstudy explaining). In Experiment 1, university students (N = 91) read a text about combustion engines and either explained the contents between the study phases or at the end of the entire study phase. A third group recalled the learning contents aloud at the end of the study phase to control for retrieval-processes that may also be involved in explaining. Results showed no overall effect of explaining in comparison to retrieval practice. However, in-between explaining enhanced students' conceptual knowledge as compared with afterstudy explaining. Verbal protocol analyses showed that this effect was due to students' increased monitoring. Experiment 2 (N = 126), had a 2 × 2-factorial design with between-subjects factors timing (in-between vs. afterstudy) and learning activity (explaining vs. written retrieval practice). We found a cascaded trend: In-between learning activities were more effective than afterstudy learning activities, whereas explaining was more effective than written retrieval practice. These findings suggest that the timing of learning activities is crucial to improve learning. Additionally, our findings reveal that explaining is not simply a result of retrieval practice.

Educational Impact and Implications Statement

Explaining is an effective learning activity which enhances students' learning. We examined whether explaining in earlier phases of studying would foster students' learning more than explaining after the entire study phase. Our findings demonstrate that preponing explaining between study phases enhances learning more than explaining at the end of the entire study phase. This benefit seems to be a result of students' higher awareness of knowledge gaps early on, which they can then attempt to remedy during further studying. This study demonstrates that the timing of explaining is crucial, and should be considered during designing learning environments.

Keywords: instructional timing, retrieval practice, learning by explaining, learning by teaching, generative learning

This article was published Online First July 18, 2019.
Andreas Lachner, Department of Psychology, University of Tübingen, and Leibniz-Institut für Wissensmedien, Tübingen, Germany; Iris Backfisch, Leibniz-Institut für Wissensmedien; Vincent Hoogerheide, Department of Education, Utrecht University, and Department of Psychology, University of Freiburg; Tamara van Gog, Department of Education, Utrecht University, and LEAD Graduate School and Research Network, University of Tübingen; Alexander Renkl, Department of Psychology; University of Freiburg.

We thank Eleonora Dolderer, Caroline Grau, 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 01JA1611.

Correspondence concerning this article should be addressed to Andreas Lachner, Leibniz-Institut für Wissensmedien, Schleichstraße 6, D-72076 Tübingen, Germany. E-mail: a.lachner@iwm-tuebingen.de

Students at all levels of education learn from expository texts, for example, during individual self-study or homework assignments. To obtain a deep understanding about the text's learning contents, students best employ different cognitive and metacognitive learning strategies (Kiewra, 2005; Mayer, 2014; Nückles, Hübner, & Renkl, 2009; Vidal-Abarca, Mañá, & Gil, 2010). Regarding cognitive strategies, students have to select the most relevant information from the text, reorganizing the selected information into a coherent mental representation, while continuously enriching and integrating the current text representation with their own prior knowledge and experiences to elaborate on the content of the text (Fiorella & Mayer, 2016; McNamara & Magliano, 2009; O'Reilly & McNamara, 2007). During these processes of selection, organization, and integration, students have to monitor their learning processes so that they are able to detect and resolve potential comprehension problems (Fiorella & Mayer, 2016; Kiewra, 2005; Nückles et al., 2009).

As students often do not employ these learning strategies spontaneously, the question arises which learning activities may help them engage in such strategies and at what point of the learning process these activities would ideally be used. Research has shown that explaining the content of learning materials to (fictitious) others is just such a beneficial activity that supports students' learning (e.g., Fiorella & Mayer, 2014; Palincsar & Brown, 1984; Plötzner, Dillenbourg, Preier, & Traum, 1999; Roscoe, 2014; Roscoe & Chi, 2008). Indeed, a meta-analysis by Kobayashi (2018) recently documented a medium effect for engaging in explanatory activities ($g = 0.56$, based on 28 comparisons).

It is an open question whether the effectiveness of generating explanations depends on the timing of the explanation activity, as the timing of explaining activities was not manipulated in previous studies and students were asked to explain almost exclusively at the end of the study phase. This preference for explaining being the final task of the learning phase might be a result of the implicit belief that explaining after studying the learning materials can help students elaborate on the learning contents which should lead to deeper learning (Bargh & Schul, 1980). However, we propose that generating explanations earlier in the study phase (e.g., explaining between study phases) should be even more conducive to learning. Explaining between study phases could trigger students' monitoring of their comprehension and help them become aware of potential knowledge gaps early on, which they can then (attempt to) remedy during further studying, thereby enhancing their understanding of the learning contents.

Against this background, we conducted two experiments to investigate whether explaining between study phases (i.e., in-between explaining) would be more effective for students' learning than explaining at the end of the study phase (i.e., afterstudy explaining). As explaining inherently involves retrieval and retrieval processes during explaining have been shown to improve learning (Koh, Lee, & Lim, 2018), we contrasted our explaining conditions to different retrieval practice conditions. This comparison allowed for investigating whether the expected effects on learning indeed came about through knowledge-building processes induced by explaining that go beyond retrieval processes (Koh et al., 2018).

Learning-By-Explaining to Fictitious Others

Explaining is a learning activity which has predominantly been applied in interactive settings in which students either provided explanations to fellow students who were physically present (e.g., Palincsar & Brown, 1984; Plötzner et al., 1999; Roscoe, 2014; Roscoe & Chi, 2008; Webb, Troper, & Fall, 1995) or to pedagogical agents in a computer-based learning environment (Chin et al., 2010; Okita & Schwartz, 2013). Hence, in most studies the peers or agents responded to the explanations (e.g., indicated whether or not they comprehended the explanations or asked further clarifying questions). This interaction can be assumed to benefit the explainers' learning by providing them with insight into the quality of their explanations (i.e., their own understanding) or prompting them to elaborate further. However, even without such interactions, explaining is still an effective instructional activity, as shown by recent research in which students provided explanations to a *fictitious* other student (cf. teaching on video; Fiorella & Mayer, 2013, 2014; Hoogerheide, Renkl, Fiorella, Paas, & van Gog, 2019).

For instance, Fiorella and Mayer (2014, Experiment 2) examined the effects of only preparing to explain learning contents (i.e., explaining-expectation only) versus preparing and actually explaining the contents on students' learning (i.e., explaining expectation + additional explaining) via a 2 × 2-factorial design. Students first read an expository text about the Doppler effect either with the intention to be tested or to provide an explanation about the learning contents. Next, students either explained the learning contents (recorded on video) or simply received additional study time. The authors demonstrated that explaining was more effective than additional studying for students' acquisition of conceptual knowledge. In addition, students who actually explained the learning contents outperformed those who only prepared to explain the learning materials. These findings were replicated in a similar study by Hoogerheide, Loyens, and van Gog (2014). The authors asked students to read a text on syllogistic reasoning either with the intention to complete a test or with the intention to explain the contents to a (fictitious) fellow student. A third group of students prepared and actually explained the learning contents (recorded on video). Again, students who provided a video explanation outperformed the students who only prepared to answer a test (or prepared to explain the contents to other students, Experiment 2). These findings show that it is the act of actually generating explanations that enhances students' comprehension.

But why is generating explanations (even without any interaction with others) a powerful learning strategy? There are three different views on how explaining might benefit individual learning. The *retrieval view* suggests that explaining simply functions as a special case of retrieval practice, as a considerable amount of explanation time is dedicated to actively retrieving the information about the previously studied material from memory (Koh et al., 2018). Actively recalling information may foster learning because it strengthens memory and retrieval cues (Rowland, 2014). In addition, explaining may help students build up new retrieval cues by so-called spreading activation (Carpenter, 2009; Endres, Carpenter, Martin, & Renkl, 2017; Endres & Renkl, 2015; Roediger & Butler, 2011; Rowland, 2014).

According to the *generative view*, explaining has benefits beyond retrieval because explaining triggers students' knowledge-building activities (Fiorella & Mayer, 2016; Rittle-Johnson, Saylor, & Swygert, 2008; Roscoe & Chi, 2008). For instance, explaining may help students in making inferences or elaborating on the contents, as they have to adapt their explanation to a particular audience (Hoogerheide, Deijkers, Loyens, Heijltjes, & van Gog, 2016; Lachner, Ly, & Nückles, 2018; Roscoe & Chi, 2008). As such, this generative view argues that the effectiveness of explaining is more than merely a retrieval practice effect (Fiorella & Mayer, 2016).

Lastly, the *metacognitive* view suggests that explaining improves learning because it helps students to monitor their understanding of the learning contents (Fukaya, 2013; Roscoe, 2014; Roscoe & Chi, 2008). This increase in comprehension monitoring should help students detect and repair potential knowledge gaps. Indeed, related evidence regarding students' metacognition indicates that monitoring activities contribute to students' learning (Loibl & Rummel, 2014; Nückles et al., 2009; Roelle, Schmidt, Buchau, & Berthold, 2017; Roscoe & Chi, 2008). This monitoring view proposes that explaining has benefits beyond retrieval practice or knowledge-building activities, provided that learners get the opportunity to repair the detected knowledge gaps.

Empirical evidence for the different accounts of explaining effects is mixed. For instance, in their laboratory study, Koh, Lee, and Lim (2018) compared the effects of explaining with retrieval practice. After studying a multimedia lesson, students provided a video explanation without teaching notes (explaining-with-retrieval), recalled the information in written form (retrieval-only), or simply performed a filler task which was not related to the learning materials (control group). Another group of students additionally received a worked-out teaching script and was required to provide an explanation by reading the teaching script verbatim (explaining-without-retrieval). The authors found that students in the explaining-with-retrieval condition and in the retrieval-only condition outperformed those who explained without retrieval and those in the control condition. However, there were no significant test performance differences between the explaining-with-retrieval and retrieval-only condition. These findings suggest that the benefits of explaining are primarily a result of retrieval practice. However, it must be noted that the explaining-without-retrieval condition simply required reading the teaching script aloud, which involved few if any generative learning processes. As such, the direct comparisons between the retrieval and the nonretrieval conditions have to be interpreted very cautiously.

Rittle-Johnson, Saylor, and Swygert (2008) investigated differential effects of explaining to another person as compared to retrieval practice. After learning how to solve multiple mathematical classification problems, children provided an explanation of the correct solution either to themselves (self-explanation) or to their moms, or were engaged in oral retrieval practice. Contrary to the findings by Koh et al. (2018), children who provided explanations outperformed children who simply rehearsed the contents on a transfer test and explaining to their moms yielded the highest knowledge gains, indicating that explaining additionally evoked distinct knowledge-building activities which contributed to their learning.

In a related study with university students, Roscoe and Chi (2008) compared different explanatory activities (explaining to a present peer, explaining to a fictitious peer, self-explaining), and analyzed the underlying cognitive and metacognitive processes during explaining. The students who explained to present or fictitious peers were instructed to provide an instructional explanation about the learning contents to another student after the study phase without the learning materials being available. The self-explainers were encouraged to continuously self-explain the content of the materials while studying. Contrary to the findings by Rittle-Johnson et al. (2008), the authors found that self-explaining was more beneficial than providing an instructional explanation. The beneficial effects of self-explaining occurred because students in the self-explaining condition were more engaged in knowledge-building processes and monitoring processes which contributed to their learning.

Based on the findings by Roscoe and Chi (2008), it could be argued that providing explanations (to fictitious others) may not always fully exploit the potential of generative activities, as it did not engage students in distinct knowledge-building and conducive monitoring processes. However, in contrast to Rittle-Johnson et al. (2008), the findings by Roscoe and Chi (2008) were potentially confounded by the timing of the explanations, as the self-explainers *continuously* self-explained the contents while studying the learning material, whereas the instructional explainers only had the opportunity to provide an explanation *after* the study phase. Thus, self-explainers may have had more opportunities to monitor their current understanding during learning and engage in regulation activities that would allow them to remedy gaps in their understanding, which would further enhance their learning. As such, the differences between the self-explainers and the instructional explainers may have emerged because of a timing effect of the explaining activities.

The Present Study: Preponing Explanatory Activities Between Study Phases

Against this background, one could expect that preponing explanation activities to early on in the study phase after only part of the text has been studied (i.e., in-between explaining) would be more beneficial than explaining after the study phase (i.e., afterstudy explaining). On the one hand, in-between explaining may lead to different and potentially more limited generative processes (e.g., elaborative or organizational processes) than afterstudy explaining because students have not yet studied all the material. On the other hand, in-between explaining does allow students to engage in comprehension monitoring which they can use to remedy potential knowledge gaps during subsequent study activities. This remedial option is not available in afterstudy explaining.

We conducted two experiments to investigate the hypothesis that in-between explaining is more beneficial to students' learning than afterstudy explaining. In Experiment 1, students either explained the contents of the learning materials on combustion engines at the end of the study phase (afterstudy explaining) or between the two study phases (in-between explaining). Additionally, we included a control group of students who engaged in oral free recall, which is a common form of retrieval practice but does not involve explaining (e.g., Carpenter, 2009; Endres et al., 2017). Additionally, we examined the cognitive and metacognitive processes underlying in-between versus afterstudy explaining as compared with retrieval practice. In Experiment 2, we used a full 2 × 2-factorial design with the between-subjects factors timing (in-between vs. afterstudy) and learning activity (explaining vs. retrieval practice) to additionally examine the effects of in-between retrieval practice.

Experiment 1

Research Questions and Hypotheses

Learning-outcome hypotheses. Based on previous findings indicating that explaining to fictitious others triggered distinct knowledge-building activities (e.g., Lachner et al., 2018; Roscoe & Chi, 2008), we hypothesized that explaining would be more conducive to students' learning than retrieval-based learning (Hoogerheide et al., 2014; Rittle-Johnson et al., 2008). The effect of explaining should be more pronounced when students were engaged in in-between explaining than afterstudy explaining, as in-between explaining allows students to not only engage in generative processes but also in monitoring their knowledge and acting upon that monitoring process by regulating their study activities during the remainder of the study phase.

Metacomprehension-accuracy hypotheses. Given that recent studies documented that explaining also has an effect on students' monitoring accuracy (e.g., Fukaya, 2013), we hypothesized that explaining generally should result in higher levels of metacomprehension accuracy than retrieval practice and that in-between explaining would lead to even more accurate metacomprehension judgments than afterstudy explaining.

Process hypotheses. Additionally, we analyzed students' explanations and verbal protocols of the retrieval practice to gain insight into the cognitive and metacognitive processes these learning activities evoked. We hypothesized that students who explained the learning contents should show more elaborations and organizational processes (i.e., bridging inferences across text passages) than students who simply recalled the learning contents (Roscoe & Chi, 2008).

Additionally, we hypothesized that students in the in-between explaining condition should show more monitoring activities than in the afterstudy explaining condition, as explaining between the study phases would likely raise students' metacognitive awareness toward potential knowledge deficits as compared with afterstudy explaining.

Mediation hypothesis. Following our previous hypotheses, we additionally tested whether the increase in monitoring processes during explaining not only affected students' judgments of their own understanding but also in turn contributed to students' learning gains. Therefore, we hypothesized that the main effect of in-between versus afterstudy explaining on students' learning outcomes should be serially mediated by students' monitoring activities and their metacomprehension accuracy (see also Wiley et al., 2016, for related findings).

Method

Participants. Ninety-nine students from nontechnical study programs of a German university participated in this study. We had to exclude eight students, as they either did not adhere to the experimental instructions (n = 3) or were not German native-speakers (n = 5). The mean age of the final sample (N = 91) was 24.76 (SD = 7.74). Twenty students were male. The students were in their sixth semester on average (SD = 3.79). In our sample, 69 students were enrolled in humanities or social study programs and the remaining 22 students were enrolled in science programs without an engineering focus. Students received 12 euros for participating.

Design. The experiment had a between-subjects design. Students were randomly assigned to one of three experimental conditions: 1) in-between explaining (n = 32) in which students explained between studying the two text units, 2) afterstudy explaining (n = 30) in which students explained after studying the entire text units, or 3) retrieval practice (n = 29) in which students were asked to recall the previously learned information aloud also after studying the two text units.

Test power with the acquired sample size was good (the required sample size was N = 90, when setting α-error to .05, power to .80, and the assumed effects to partial η^2 = .08, as recent studies showed medium to large effects of explaining on students' learning, e.g., Fiorella & Mayer, 2014; Hoogerheide et al., 2014).

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

Study text. The study text was an adapted German Wikipedia article on four-stroke engines, a popular type of internal combustion engines (previously used in a study by Lachner et al., 2018). The text consisted of two larger text units (total word count: 2,048 words). The first text unit (Study Phase 1) dealt with the general functions and processes of four-stroke engines and comprised 884 words and two static pictures visualizing the general construction of four-stroke engines. The second text unit (Study Phase 2) exemplified the general functions and constructions by two concrete four-stroke engines: the gasoline engine and the diesel engine (word count: 1,164, one static picture).

Conceptual knowledge pretest and posttest. Based on the information provided in the text, we constructed a test to measure students' conceptual knowledge about internal combustion engines. This test was used as both a pretest and a posttest. It included eight open-answer questions (e.g., "What is the basic functioning of internal combustion engines?"; "Which factors affect the efficiency of internal combustion engines?"; "How is an internal combustion engine generally constructed and what are the central functions of the individual components?"). Students received two points for each correctly solved question, yielding a possible maximum score of 16. To avoid memory effects, the order of the posttest questions was randomized per participant. Ten percent of the conceptual knowledge tests were scored independently by two trained raters who were blind to the experimental conditions. Interrater reliability was very good, ICC = .96 (Wirtz & Caspar, 2002). Thus, only one rater coded the rest of the explanations. The reliability of the conceptual knowledge test was good, α = .80 (Cronbach's alpha).

Transfer posttest. We used the transfer test by Lachner, Ly, and Nückles (2018), which consisted of two open questions that required students to transfer the general functions and processes of internal combustion engines to other internal combustion engines (i.e., gas turbine and two-stroke engines; for more details, see Lachner et al., 2018). The combustion engines used in the transfer tasks were comparable regarding the general principles of internal combustion engines (e.g., transformation of chemical energy into mechanical energy), but differed in the construction of the particular engines and the underlying sequence of the four-stroke processes. Again, 10% of the transfer tasks were scored independently by two trained raters who were blind to the experimental conditions. Interrater reliability was very good, $ICC = .93$ (Wirtz & Caspar, 2002). Thus, only one rater coded the rest of the explanations. Cronbach's alpha was relatively low ($\alpha = .45$), which was to be expected given that the number of items was low and that the two tasks required students to transfer their knowledge to two completely different combustion engines. As Cronbach's alpha is an indicator of the interrelatedness of test items, values of Cronbach's alpha are commonly high for unidimensional constructs, such as for our conceptual knowledge test, but not necessarily for knowledge tests which require the application of knowledge from broad knowledge bases such as transfer tests (Tavakol & Dennick, 2011).

Mental effort. Students were asked to self-report how much mental effort they had invested in studying the first text unit (i.e., the general functions and components of combustion engines), in studying the second text unit (i.e., examples of combustion engines), and in retrieval practice or explaining (depending on assigned condition) immediately upon completing each activity. They rated their invested mental effort on a 9-point rating scale from 1 (*very low effort*) to 9 (*very high effort*; see Paas, 1992).

Metacomprehension accuracy. To investigate students' metacomprehension accuracy, students were asked to make prospective judgments of learning (after studying the first text unit, the second text unit, and after the learning activity) about their expected performance on the conceptual knowledge test. We used the final judgment of learning at the end of the entire learning phase for determining students' metacomprehension accuracy. Students estimated how many points they would achieve on the conceptual posttest (eight questions, two points), resulting in a scale from 0 to 16. 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 and absolute accuracy (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., $X_{Judgment} - X_{Performance}$). 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. Note that despite its benefits, one caveat of the bias measure is that it is relatively unreliable to investigate the average accuracy of the sample mean for experimental comparisons (Prinz et al., 2018). For instance, when the bias is computed, and there are underconfident and overconfident students within one or multiple condition(s), over- and underestimations may cancel each other out and result in less robust metacomprehension estimates of the average mean per condition.

Therefore, we additionally used the absolute accuracy measure (Baars et al., 2017; Prinz et al., 2018). Absolute accuracy is operationalized as the difference between judgment and performance regardless of the direction of the difference (i.e., $|X_{Judgment} - X_{Performance}|$). Positive and negative differences are both counted as inaccuracies and—in contrast to the bias measure—do not cancel each other out while calculating mean differences within conditions (Prinz et al., 2018). Higher values indicate larger inaccuracies and a value of zero reflects an accurate judgment. Given that the bias measure has a trade-off regarding its reliability for statistical comparisons between conditions, we used the absolute accuracy measure for investigating potential statistical differences among experimental conditions. Nevertheless, we report the descriptive bias measure to indicate the direction of the metacomprehension accuracies within the conditions (see Prinz et al., 2018).

Procedure. Students were randomly assigned to the particular session via our online-participant environment. The students were seated in sound-proof cubicles (maximum: $n = 6$) with a tablet computer (Apple iPad 4) and a wireless keyboard. The doors were open during the study so that the students could listen to the experimenters' verbal instruction. The doors were only closed during the learning activities so that the students were not distracted by each other's talking. During the experimental sessions (duration: 1.5 hr), the students were not allowed to proceed before being signaled by the experimenter (exact-time-on-task). At the beginning of the study, the students were informed that they would take part in a study on learning about combustion engines. They were instructed that they would engage in different learning activities in different phases of the experiment. After providing written consent, students in all conditions completed the pretest (15 min). Then, they studied the first text unit (10 min). After studying the first text unit, participants were required to indicate their invested mental effort and to provide a judgment of learning. Then, students in the retrieval and afterstudy explaining condition engaged in the same procedure for the second text unit (10 min), while students in the in-between explaining condition first provided an explanation about the first text unit (15 min), estimated how much effort they invested in explaining, and provided their judgment of learning before moving to the second text unit. After completing the second text unit, students in the afterstudy explaining and the retrieval practice condition engaged in explaining or retrieval followed by effort and judgment of learning ratings. Finally, all students completed the conceptual knowledge (15 min) and transfer tests (10 min). See Table 1 for an overview of the study procedure.

For the explanation activity (either between the two text units or after the second text unit), students in the explaining conditions were instructed to create a video message for a potential novice student by using the built-in camera function of the tablet:

> Unfortunately, one of your fellow students could not participate in this study but is very interested in combustion engines. Therefore, the student asks you to record an explanation of the contents of the text on

Table 1
Conditions and Materials Used in the Study

In-between explaining	After-study explaining	Retrieval practice
Pretest	Pretest	Pretest
Text unit 1	Text unit 1	Text unit 1
Mental effort/JoL	Mental effort/JoL	Mental effort/JoL
Explaining	Text unit 2	Text unit 2
Mental effort/JoL	Mental effort/JoL	Mental effort/JoL
Text unit 2	**Explaining**	**Free recall (aloud)**
Mental effort/JoL	Mental effort/JoL	Mental effort/JoL
Posttest	Posttest	Posttest

Note. Bold items varied across experimental conditions.

> video. Please make sure that your explanation is comprehensible for your fellow student and can be understood without any additional learning material.

For the retrieval practice activity, students in the retrieval practice condition were instructed to freely recall the contents of the text while thinking aloud and to utter everything that came to their mind:

> Please recall all the information of the text. As we are interested in what you think while you recall the information, we ask you to think aloud. By "thinking aloud" we mean that you mention everything that comes to your mind during the recall phase. We therefore ask you to speak all the time, from the time the recall starts until the end of the recall phase.

The free recall was recorded by the built-in recording function of the tablet.

Analysis and coding. For the analyses of the students' learning processes during the learning activities, the recordings of the explanation and the free recall activities were transcribed and segmented into idea units. We used a categorization scheme by Lachner and Nückles (2015) to categorize the idea units (see Table 2). A total of 20 transcriptions (22.72%) were corated by a second rater. Interagreement between the two raters was very good for all categories ($.78 < \kappa > .98$).

Results

We used an alpha level of .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 .06 and .14 as a medium effect, and values $>$ .14 as a large effect (see Cohen, 1988).

Preliminary analyses. ANOVAs and a χ^2 test showed no significant differences between the experimental conditions concerning age, $F(2, 88) = 0.21$, $p = .807$, $\eta_p^2 = .005$; prior knowledge, $F(2, 88) = 1.37$, $p = .260$, $\eta_p^2 = .030$; or gender, $\chi^2(2) = 2.26$, $p = .323$, indicating that there were no significant a priori differences among conditions. Additionally, a MANOVA showed that there were no significant differences among conditions in reported mental effort investment in the different activities (Reading Text Unit 1, Reading Text Unit 2, learning activity), $F(6, 174) = 1.09$, $p = .373$, $\eta_p^2 = .036$. Apparently, students were mentally engaged in a comparable manner across the different activities during the experiment. Thus, potential effects of in-between explaining did not result due to different levels of engagement in the activities. For the descriptive statistics of our dependent measures, see Table 3.

Learning-outcome hypotheses. As we had directed hypotheses, we conducted contrast analyses to test our learning-outcome-hypotheses (Furr & Rosenthal, 2003). To test our first hypothesis that explaining is more beneficial for learning outcomes than retrieval practice, we contrasted the explaining conditions against the retrieval practice condition with the following contrast weights: in-between explaining: 1; afterstudy explaining: 1; retrieval practice: −2. The second contrast (in the following: "in-between-contrast") was used to test the prediction that in-between explaining would be superior to afterstudy explaining (in-between explaining: 1; afterstudy explaining: −1; retrieval practice: 0). In the contrast analyses, the contrasts represented a fixed factor and the students' learning outcomes (i.e., conceptual knowledge, transfer) were the dependent variables. Additionally, we included prior knowledge as covariate to control for potential differences with regard to students' knowledge prerequisites.

Regarding students' performance on the conceptual knowledge test, in contrast to our hypothesis, the explaining-contrast was not significant, $F(1, 87) = 0.47$, $p = .496$, $\eta_p^2 = .005$, suggesting that explaining was as effective as retrieval practice. In line with our hypotheses, however, the in-between-contrast was significant, $F(1, 87) = 5.88$, $p = .017$, $\eta_p^2 = .063$, indicating that in-between explaining was more beneficial than afterstudy explaining (see Table 3, for the descriptive statistics).

Regarding students' performance on the transfer test, contrary to our expectations, none of the contrasts were significant ($F < 1$), indicating that students showed similar (and very low) transfer performance across all experimental conditions (see Table 3).

Metacomprehension-accuracy hypotheses. The descriptive findings suggested that students generally underestimated their comprehension (see Table 3) and that the underestimation was most pronounced when students were engaged in in-between explaining. However, the contrast analyses on the absolute

Table 2
Categories Used to Categorize the Transcripts of the Explanations and the Think-Aloud Protocols in Experiment 1

Category	Description	Examples
Paraphrase	Student simply restated or paraphrased a text segment from the online text without linking the information to other text segments or to her or his prior knowledge.	"The online text says that the general principle of combustion engines is the transformation of chemical to mechanic energy."
Bridging inferences	Student related different text passages across the text to better understand relations between sentences to establish coherence.	"A working cycle of the engine refers to the different strokes of internal combustion engines."
Elaboration	Student connected new information out of the online text to her/his prior knowledge. Indicators are the generation of examples, report of own experience, or making predictions.	"Motor scooters also have internal-combustion engines." "I have also experienced that diesel engines emit more exhaust gases during combustion."
Monitoring	Student expressed her or his (non)-understanding.	"I don't understand how the four-stroke engine works." "I am not sure, what this means." "Strange, I thought that . . ."

Table 3
Means and Standard Deviations for the Dependent Measures of Experiment 1

Dependent variable	Retrieval practice	In-between explaining	After-study explaining
Prior knowledge[a]	.07 (.07)	.06 (.07)	.09 (.10)
Learning outcome			
Conceptual knowledge[b]	.52 (.22)	.61 (.19)	.51 (.28)
Transfer[b]	.29 (.15)	.27 (.16)	.29 (.16)
Metacomprehension			
Bias	−.11 (.16)	−.17 (.24)	−.01 (.26)
Absolute accuracy[c]	.16 (.11)	.22 (.18)	.21 (.14)
Mental effort			
Reading unit 1	5.86 (1.81)	6.50 (1.19)	6.37 (1.16)
Reading unit 2	5.93 (1.73)	6.66 (1.45)	6.67 (1.30)
Learning activity	6.38 (1.50)	6.47 (1.68)	6.63 (1.43)
Learning activities			
Paraphrase	.66 (.23)	.86 (.80)	.77 (.16)
Inferences	.12 (.20)	.11 (.15)	.09 (.11)
Elaboration	.06 (.09)	.03 (.06)	.07 (.12)
Monitoring	.16 (.15)	.16 (.14)	.07 (.12)

[a] Students' performance on the knowledge tests were transformed to percentage scores. [b] The measures for students' learning activities were transformed to proportions. [c] The accuracy measure could range from 0 = *completely accurate* to 1 = *completely inaccurate*.

accuracy of students' final metacomprehension judgments (after the entire study phase), controlling for students' metacomprehension judgments during the reading phase by using that as covariate (see also Hertzog, Hines, & Touron, 2013, for similar approaches), showed that neither the explaining-contrast, nor the in-between-contrast was statistically significant ($F < 1$), indicating that students' mean accuracy did not statistically differ across conditions (see Table 3).

Process hypotheses. In the next step, we examined students' underlying cognitive and metacognitive processes during the different learning activities. Unfortunately, two recordings (one explanation of the in-between explanation condition, one think-aloud protocol of the retrieval practice condition) were not stored on the internal data storage of the tablets. Thus, these analyses are based on a sample of 89 students. We hypothesized that explaining activities should trigger more knowledge-building processes (i.e., inference and organization) than retrieval practice. To test this hypothesis, we contrasted the explaining conditions with the retrieval condition (early explaining: 1; late explaining: 1; retrieval practice: −2). However, the test of the contrast was not significant, neither for students' inferential processes, $F(1, 86) = 0.47$, $p = .494$, $\eta_p^2 = .005$, nor for the elaborative processes, $F(1, 86) = 0.29$, $p = .590$, $\eta_p^2 = .003$. Additionally, the proportion of paraphrases was comparable across conditions, $F(1, 86) = 1.89$, $p = .173$, $\eta_p^2 = .021$.

Regarding students' monitoring activities, we expected that in-between explaining should trigger students' metacognitive monitoring during explaining more than afterstudy explaining. To test this assumption, we contrasted the in-between explaining condition to the afterstudy explaining condition (in-between explaining: 1; afterstudy explaining: −1; retrieval practice: 0). In line with our hypothesis, the test of the contrast was significant, $F(1, 86) = 6.57$, $p = .012$, $\eta_p^2 = .071$, indicating that in-between explaining triggered students to engage in more monitoring of potential knowledge gaps in early stages of their study phases than afterstudy explaining.

Mediation analyses. Finally, we tested our mediation hypothesis: We tested whether in-between explaining enhanced students' conceptual learning more than afterstudy explaining, because in-between explaining triggered more monitoring activities than afterstudy explaining. The higher levels of monitoring activities should result in higher levels of metacomprehension accuracy and contribute to students' conceptual learning. To test this hypothesis, we conducted a mediation analysis with two serial mediators (proportion of monitoring, absolute metacomprehension accuracy), the type of learning activity was a contrast-coded predictor (0 = retrieval practice, −1 = afterstudy explaining, 1 = in-between explaining), and students' conceptual knowledge was the dependent variable (see Figure 1). As with previous analyses, we controlled for students' prior knowledge as well as students' metacomprehension accuracy in the reading phase. We used OLS regression-based path analyses and applied the bootstrapping methodology via the PROCESS macro for SPSS (Hayes, 2013). We ran 10,000 bootstrap samples to derive a 95%-bias-corrected confidence interval for the indirect effect.

We found a significant indirect effect of the two-serial mediator model (via monitoring activities → absolute metacomprehension accuracy), $a \times d \times b = .07$, 95% CI [.01, .27], as zero was not included in the confidence interval (see Figure 1 for the full mediation model). The significant serial mediation effect indicates that the effect of in-between explaining on students' conceptual knowledge emerged at least partially because of a higher proportion of monitoring activities and subsequent differences in their metacomprehension accuracy. The sign of the regression weights (between monitoring accuracy and learning outcome), however, was positive—indicating that inaccurate rather than accurate judgments improved students' conceptual learning. Given that the lower metacomprehension accuracy resulted primarily from underestimations, the fact that it resulted in higher learning gains suggests that learners may

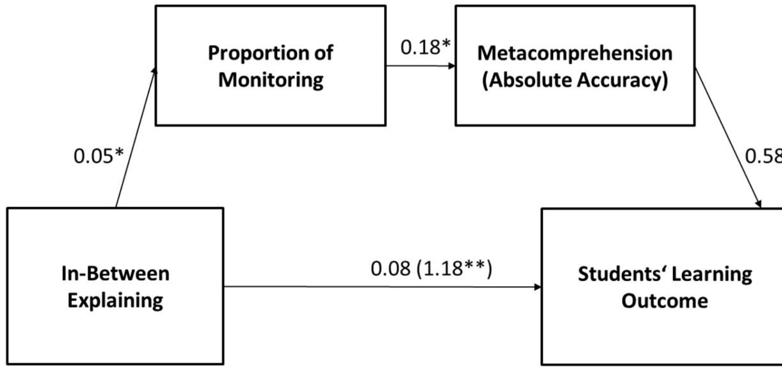

Figure 1. Main findings of the serial mediation analysis. Numbers represent unstandardized path coefficients for the direct and total effects (in parentheses). * $p < .05$. ** $p < .01$.

have been attempting to remedy their understanding during further studying.

Discussion

Our findings provide evidence that in-between explaining supported students' acquisition of conceptual knowledge more than afterstudy explaining particularly because in-between explaining increased students' monitoring activities. As such, our findings extend previous studies on explaining to fictitious others (e.g., Fiorella & Mayer, 2014; Hoogerheide et al., 2014) by showing that not only the *act* of explaining contributed to students' knowledge acquisition but also the *instructional timing*, that is, at what point in the study phase an explanation was provided.

A surprising finding was that explaining was not more beneficial than retrieval practice, as comparable amounts of cognitive processes were triggered in the explaining conditions and in the retrieval practice condition. This finding may have resulted because we audio-recorded students' verbal recall activities. Thus, the students in the retrieval practice condition might have believed that someone else was listening along or would do so later (i.e., social presence; Hamilton & Lind, 2016; Hoogerheide et al., 2016, 2018). The awareness of the potential audience during free recall could have evoked all sorts of advantageous meta(-cognitive) processes, such as higher levels of cognitive engagement during retrieval practice (see Sundqvist, Mäntylä, & Jönsson, 2017, for empirical evidence on written vs. oral retrieval practice). This would also explain why explaining afterstudy resulted in similar levels of cognitive processes as retrieval practice. If this explanation holds true, then written recall should be less beneficial than explaining, as written retrieval should trigger lower amounts of knowledge-building activities (see also Hoogerheide et al., 2016; Lachner et al., 2018, for related empirical evidence on explanations).

Additionally, the three-group design of Experiment 1 did not allow us to disentangle the effects of instructional timing (i.e., in-between vs. afterstudy learning activities) and the type of learning activities (i.e., retrieval practice vs. explaining). Therefore, it is an open question whether in-between retrieval practice would also contribute to students' learning as compared with afterstudy retrieval.

Experiment 2

Research Questions and Hypotheses

To address these open questions, in Experiment 2 we used a factorial design with two between-subjects factors: instructional timing (in-between activity vs. afterstudy activity) and the type of learning activity (written retrieval practice vs. explaining). Additionally, we asked the students to engage in written retrieval practice—that is, without being (audio) recorded. This procedure allowed us to replicate the findings of Experiment 1 in written settings of retrieval practice (see Carpenter, 2009; Endres et al., 2017; Koh et al., 2018, for similar approaches), and potentially lower amounts of social presence during retrieval, which allowed us to disentangle the effects of active retrieval practice and social presence on students' learning.

Based on the findings of Experiment 1, we hypothesized that in-between learning activities would be more beneficial for students' conceptual understanding than afterstudy learning activities. More importantly, we assumed that the effect of learning activities should be more pronounced when students actually explained the learning contents as compared with being engaged in written retrieval practice. This was expected because written recall likely triggers lower feelings of social presence (as compared with the oral retrieval practice in Experiment 1; see also Hoogerheide et al., 2016; Sundqvist et al., 2017). Thus, we assumed a cascaded trend (in-between explaining > in-between retrieval practice > afterstudy explaining > afterstudy retrieval practice). As in Experiment 1, we explored potential effects of our interventions on students' invested mental effort and metacomprehension accuracy.

Method

Participants. Participants were 132 students from nontechnical study programs from a German university. We excluded six students, as they were not German native speakers. The remaining students ($N = 126$) were comparable with the sample of Experiment 1 regarding study programs (66.46% humanity programs), age ($M = 23.24$; $SD = 3.17$), and gender (79.37% female). Test power with the acquired sample size was good (the required sample size was $N = 126$, while setting α-error to .05, power to .80, and the assumed effects to partial $\eta^2 = .06$, based on Experiment 1). As in Experiment 1, the students received 12 euros for participating.

Design and materials. We used the same materials as in Experiment 1. As students' reading skills could impact their learning from text (Golke & Wittwer, 2017; Ozuru, Dempsey, & McNamara, 2009), we additionally measured their reading skills by a parallel version of the German reading and speed comprehension test (LGVT 6–12; see Schneider, Schlagmüller, & Ennemoser, 2007) to see whether the experimental conditions were comparable regarding their reading skills. The test was a speed test and comprised a reading task with 25 gaps for which students had to decide which of three provided words had to be filled in. Students' reading skill was determined by summing the number of correctly filled-in gaps. Cronbach's alpha was very good: $\alpha = .84$.

Procedure. The procedure was identical to Experiment 1. The only exception was that we included an "in-between retrieval practice condition." Thus, students were either assigned to in-between explaining ($n = 34$), in-between retrieval practice ($n = 30$), to afterstudy explaining ($n = 29$), or to afterstudy retrieval practice ($n = 33$). In contrast to Experiment 1, the students in the retrieval practice conditions recalled the learning material in written form to reduce potential effects of social presence: "Recall all the contents of the text on combustion engines. Try to remember as much as possible and write them down on the separate sheet (see also Sundqvist et al., 2017, for similar approaches)."

Results

Table 4 provides the descriptives of this study.

Preliminary analyses. There were no significant differences among experimental conditions regarding age, $F(3, 122) = 0.90$, $p = .443$, $\eta_p^2 = .022$; prior knowledge, $F(3, 122) = 1.39$, $p = .250$, $\eta_p^2 = .033$; average reading skills, $F(3, 122) = 0.66$, $p = .576$, $\eta_p^2 = .016$, and gender, $\chi^2(6) = 4.80$, $p = .569$. As with Experiment 1, we performed a MANOVA with timing and learning activity as independent factors and students' invested mental effort for the different activities (Reading Text Unit 1, Reading Text Unit 2, learning activity) as dependent variables. There was no main effect of learning activity, $F(3, 120) = 0.86$, $p = .463$, $\eta_p^2 = .021$, but a significant main effect of timing, $F(3, 120) = 3.75$, $p = .013$, $\eta_p^2 = .021$. The interaction between learning activity and timing was not significant, $F(3, 120) = 2.65$, $p = .052$, $\eta_p^2 = .062$. Separate ANOVAs showed that the effect of timing on students' mental effort was only significant for reported effort investment in reading the second text unit, $F(1, 122) = 8.36$, $p = .005$, $\eta_p^2 = .064$, as students with an in-between learning activity invested more mental effort in reading the second text unit than students with an afterstudy learning activity (see Table 4 for the descriptives). None of the other effects were significant ($F < 1$). Apparently, preponing learning activities between study phases contributed to students' deep-processing when reading the second text unit.

Learning-outcome hypotheses. As we had directed hypotheses for students' learning outcomes, as in Experiment 1, we conducted a priori contrast analyses for full-factorial designs (Wiens & Nilsson, 2017). This approach allowed us to flexibly test for specific test patterns (e.g., cascaded trends) and at the same time reduce the risk of alpha-inflation, as we only had to conduct one test (Furr & Rosenthal, 2003). Furthermore, it is suggested to prefer contrast analyses, especially when the test of a particular hypothesis cannot be captured by conventional ANOVAs (main effects and interactions), such as cascaded or synergistic trends (Wiens & Nilsson, 2017). As in Experiment 1, we controlled for students' prior knowledge in our analyses. Based on our learning-outcome hypotheses, we assumed a cascaded trend of in-between learning activities: In-between learning activities should be more effective than afterstudy learning activities, whereas providing explanations should be more effective than written retrieval practice (in-between explaining > in-between retrieval practice > afterstudy explaining > afterstudy retrieval). We used the following contrast weights for this analysis: in-between explaining: 2; in-between retrieval practice: 1; afterstudy explaining: −1; afterstudy retrieval practice: −2 (see also Wiens & Nilsson, 2017). The

Table 4
Means and Standard Deviations for the Dependent Measures of Experiment 2

	In-between		After-study	
Dependent variable	Explaining	Retrieval	Explaining	Retrieval
Prior knowledge[a]	.09 (.10)	.04 (.04)	.06 (.10)	.07 (.11)
Reading skills[b]	24.24 (7.25)	23.47 (10.71)	22.07 (8.92)	21.58 (7.47)
Learning outcomes[a]				
Conceptual knowledge	.62 (.21)	.56 (.23)	.52 (.22)	.49 (.25)
Transfer	.28 (.14)	.26 (.17)	.26 (.14)	.23 (.15)
Mental effort				
Reading unit 1	6.21 (1.09)	6.53 (1.17)	6.45 (.95)	6.00 (1.56)
Reading unit 2	6.76 (1.04)	7.07 (1.39)	6.59 (1.59)	5.76 (1.68)
Learning activity	6.38 (1.10)	6.67 (1.54)	6.97 (1.05)	6.03 (1.31)
Metacomprehension				
Bias	−.08 (.22)	−.01 (.25)	−.01 (.24)	−.01 (.24)
Absolute accuracy[c]	.18 (.14)	.21 (.11)	.18 (.15)	.21 (.13)

[a] Students' performance on the knowledge tests were transformed to percentage scores. [b] Fifty points were possible on the reading skill test. Given that the LGVT is a speed test with very strong time constraints, the obtained values show high reading skills, see Schneider, Schlagmüller, and Ennemoser (2007) for the norm values. [c] The accuracy measure could range from 0 = *completely accurate* to 1 = *completely inaccurate*.

contrast was indeed significant, $F(1, 121) = 6.13$, $p = .015$, $\eta_p^2 = .048$, indicating that students' acquisition of conceptual knowledge was best supported when they were engaged in in-between explaining (see Table 4). Regarding students' transfer, as in Experiment 1 the contrast was not significant, $F(1, 121) = 1.43$, $p = .234$, $\eta_p^2 = .012$.

Metacomprehension-accuracy hypothesis. As in Experiment 1, the descriptive findings again indicated that students generally underestimated their comprehension (see Table 4), and that the underestimation was most pronounced when students were engaged in in-between explaining. However, again, a contrast analysis with students' final metacomprehension judgments (i.e., the absolute accuracy measure at the last measuring point) as a dependent variable and students' metacomprehension judgments during the reading phase as covariate (Hertzog et al., 2013), was not significant, $F(1, 121) = 1.02$, $p = .315$, $\eta_p^2 = .008$.

Discussion

In Experiment 2, we successfully confirmed our hypotheses: We found a cascaded trend showing that in-between explaining significantly enhanced students' conceptual learning compared to in-between retrieval practice and afterstudy explaining. Afterstudy retrieval practice showed the lowest test performance (see Table 4). These findings indicate that explaining was more conducive to students' learning as compared to written retrieval practice both in the in-between and the afterstudy conditions. Presumably, written retrieval practice triggered lower amounts of social presence, which likely reduced the effectiveness of the retrieval practice activity. Comparing the descriptives of Experiment 2 and Experiment 1 also indicated that the means of students' conceptual knowledge for the afterstudy written recall conditions in Experiment 2 ($M = .49$) were somewhat lower than the oral retrieval practice condition in Experiment 1 ($M = .52$). This finding may indicate that the verbalization during learning activities additionally affected students' conceptual learning.

General Discussion

Despite the fact that a lot of research has shown that explaining is a powerful strategy for learning (Fiorella & Mayer, 2014; Hoogerheide et al., 2019; Palincsar & Brown, 1984; Plötzner et al., 1999; Roscoe, 2014), research has not taken the timing of explaining into account. Most research investigated effects of explaining after studying a text entirely. Because explaining between study phases (i.e., after having studied part of the text) might provide opportunities for identifying and repairing knowledge gaps, the central goal of the present study was to investigate the effects of in-between explaining activities on students' learning. Our findings indicated that in-between explaining improved learning compared with afterstudy explaining, because in-between explaining triggers distinct monitoring activities which help students instantiate adequate regulation processes in subsequent study phases (Experiment 1). Furthermore, we showed that the effects of in-between activities were dependent on the type of learning activity, as students who were engaged in in-between explaining outperformed those who engaged in in-between retrieval practice (Experiment 2). We attribute the significant differences between explaining and retrieval regarding students' conceptual learning (Experiment 2) to the fact that we asked students to realize the free recall in written form. Comparisons of the findings across Experiments 1 and 2 suggest that particularly the verbalization may have accounted for the differences across our experiments: In Experiment 1, explaining and aloud retrieval practice were comparably effective. The analyses of students' explanations and students' verbal protocols of Experiment 1 showed a similar pattern: Students who provided an explanation expressed a comparable amount of cognitive processes (paraphrases, knowledge inferences, elaboration), as students who were engaged in verbalizing their recall-activities aloud. In contrast, in Experiment 2 the contrast analyses showed significant differences between the explaining and written retrieval practice conditions. The only difference between the two studies was that we administered a written free-recall task (Experiment 2) instead of an oral recall task (Experiment 1). We attribute these differences in test performance to differences of social presence across experiments, as written retrieval practice may have induced lower levels of feelings of social presence as compared with oral retrieval practice. Therefore, our findings suggest that besides distinct learning activities, the verbalization in combination with being aware of a potential audience particularly plays a key role during explaining. However, the social presence hypothesis is only one potential explanation for the differences between written and oral learning activities. An alternative explanation might lie in the fact that writing and speaking involve different neurocognitive mechanisms (Cleland & Pickering, 2006; Sperling, 1996). Speaking is a primary and automated communicative act, whereas writing is a secondary cultural practice that has to be explicitly instructed and deliberately practiced (Cleland & Pickering, 2006; Paas & Sweller, 2012). Further research is needed to shed light on the different (combinations of) neuro-cognitive and sociocognitive mechanisms that account for the effectiveness of explaining.

An unexpected finding of our experiments was that we did not find a direct effect of in-between explaining on students' metacomprehension accuracy, suggesting that in-between explaining did not further contribute to students' ability to judge their understanding accurately. However, it has to be noted that overall the level of metacomprehension accuracy was rather high among our students (see Table 3 and Table 4). The high levels of metacomprehension accuracies were also reflected by their average reading skills, as the students of Experiment 2 (and likely the students of Experiment 1) showed relatively high performance on the reading skill test (see Table 4), which can be attributed to the fact we tested a sample of university students who were regularly confronted with complex reading assignments. Therefore, it would be interesting to replicate this study with less skilled learners to investigate whether in-between explaining would affect those students' metacomprehension accuracy, as well as their learning.

A further unexpected finding of Experiment 1 was that students' metacomprehension *in*accuracy was positively related to students' learning gains, indicating that less accurate judgments yielded more learning. This finding is in contrast to current research which presumes that accurate monitoring is critical for the adequate regulation of cognitive processes (e.g., Dunlosky & Rawson, 2012; Gutierrez, Schraw, Kuch, & Richmond, 2016; Schleinschok, Eitel, & Scheiter, 2017; Wiley et al., 2016). However, it has to be noted that previous research mostly referred to students' overconfidence as being problematic (e.g., Baars et al., 2017; Dunlosky & Rawson, 2012; Roelle et al., 2017). For instance, in their early studies, Begg, Martin, and Needham (1992) showed that test performance was even greater for students who less accurately judged their performance than students who accurately monitored their learning. Similarly, Roelle, Schmidt, Buchau, and Berthold (2017) experimentally reduced metacomprehension accuracy by providing relevance instruction about the dangers of overconfident judgments. In three studies, they found that the availability of this information reduced students' confidence judgments. Reduced confidence judgments subsequently contributed to students' learning. Thus, in light of these findings, one may conclude that at least slight underestimations of one's knowledge in early phases of studying may enhance learning as compared to overestimations, as students likely invest more effort in subsequent study phases. Which cognitive and metacognitive processes were triggered in the subsequent second study phase, however, is an open question. The use of eye tracking in future research could perhaps offer particular insights into how students regulated their learning processes in the subsequent study phase.

Additionally, we have to note that students' proportion of knowledge-building activities and, by implication, their knowledge gains were rather low in our studies. Apparently, explaining to fictitious others did not evoke a sufficiently grounded communicative situation in which students were required to adapt and transform their knowledge to a potential student's needs so that the provided information is tangible to the recipient (Kobayashi, 2018; Lachner et al., 2018; Scardamalia & Bereiter, 1987). Given that video explanations are a parsimonious instructional tool which can be easily implemented, particularly in online educational settings, such as flipped classrooms or distance education, it would seem fruitful to further investigate whether and how explaining to fictitious others can be made more effective. A simple and easy-to-use strategy would be to provide students with distinct information about the knowledge of a potential audience to trigger knowledge-building processes. For instance, Wittwer, Nückles, Landmann, and Renkl (2010) provided tutorial explainers with an assessment tool in which they received information about a recipients' individual knowledge level. The authors found that explainers who had such information available also produced more adapted explanations that were customized to the potential audience (see also Nückles, Wittwer, & Renkl, 2005, for related findings). Future research could explore whether explicit information about a fictitious recipient's knowledge not only has an effect on the explanatory product but also on the explainers' knowledge gains.

One caveat of our study is that we cannot be absolutely sure that only the timing of explaining accounted for students' learning, as the students in the in-between condition were required to provide an explanation after they received abstract information about the general principles of combustion engines (Study Phase 1), whereas afterstudy explainers provided an explanation after they additionally received concrete information about examples of combustion engines (Study Phase 2). We chose such a design, as we wanted to use authentic learning material that is in line with the common rhetorical structure of expository texts (e.g., broad general introduction, progressively narrowing concrete information, see Lachner & Neuburg, 2019; Linderholm et al., 2001) to heighten the ecological validity of our studies. Nevertheless, it is not clear whether the obtained effects of in-between explaining resulted because students were engaged in in-between explaining activities or because they provided more generalized and abstract explanations, as they had no concrete information about examples available as compared with the afterstudy explaining condition. Abstract explaining may have helped students more easily relate the abstract principles and functions to the concrete examples in the second study phase and thereby advance their conceptual understanding (Hinds, Patterson, & Pfeffer, 2001; Kalyuga, Renkl, & Paas, 2010). Our findings of the different qualities of the explanations, however, showed that the explanations in the in-between and afterstudy explaining condition were comparable in their level of concreteness/abstractness, as the students included a comparable proportion of concrete elaborations in their explanations (see Table 3). Therefore, we are still tempted to attribute our findings to the specific timing of the explanations and not to the different levels of abstractness of students' explanations. Nevertheless, future studies should use comparable instructional material across study phases regarding the level of abstractness to finally disentangle the effects of abstractness of information and timing of the explanatory activities.

A final caveat refers to the ecological validity of our study, as we conducted the study in a laboratory setting, which likely reduced the students' motivation to engage in the different learning activities. Therefore, future studies should replicate our findings in more authentic and applied learning settings, in which actual novice students of the particular domain are engaged in in-between versus afterstudy explaining.

All in all, our findings are a promising starting point for further research on the effects of in-between learning activities on students' acquisition of conceptual knowledge, as our study is one of the first which documented that the sequencing of explaining activities is crucial for the effectiveness of explaining activities. Our findings demonstrate that learning by explaining can be more effective when it is instantiated between study phases rather than after studying. These findings also have implications for educational practice, as they demonstrate the relevance of the timing of distinct learning activities. Providing in-between video explanations can, therefore, represent an alternative instructional strategy and increase teachers' repertoire to engage students in meaningful knowledge-building and monitoring processes.

References

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,* 810–834. http://dx.doi.org/10.1080/01443410.2016.1150419

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72,* 593–604. http://dx.doi.org/10.1037/0022-0663.72.5.593

Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *The European Journal of Cognitive Psychology, 4,* 195–218. http://dx.doi.org/10.1080/09541449208406182

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,* 1563–1569. http://dx.doi.org/10.1037/a0017021

Chin, D. B., Dohmen, I. M., Cheng, B. H., Oppezzo, M. A., Chase, C. C., & Schwartz, D. L. (2010). Preparing students for future learning with teachable agents. *Educational Technology Research and Development, 58,* 649–669. http://dx.doi.org/10.1007/s11423-010-9154-5

Cleland, A. A., & Pickering, M. J. (2006). Do writing and speaking employ the same syntactic representations? *Journal of Memory and Language, 54,* 185–198. http://dx.doi.org/10.1016/j.jml.2005.10.003

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

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

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. http://dx.doi.org/10.1016/j.learninstruc.2016.11.010

Endres, T., & Renkl, A. (2015). Mechanisms behind the testing effect: An empirical investigation of retrieval practice in meaningful learning. *Frontiers in Psychology, 6,* 1054. http://dx.doi.org/10.3389/fpsyg.2015.01054

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38,* 281–288. http://dx.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,* 75–85. http://dx.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,* 717–741. http://dx.doi.org/10.1007/s10648-015-9348-9

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

Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics, 2,* 33–67. http://dx.doi.org/10.1207/S15328031US0201_03

Golke, S., & Wittwer, J. (2017). High-performing readers underestimate their text comprehension: Artifact or psychological reality? In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2108–2113). London, UK: Cognitive Science Society.

Gutierrez, A. P., Schraw, G., Kuch, F., & Richmond, A. S. (2016). A two-process model of metacognitive monitoring: Evidence for general accuracy and error factors. *Learning and Instruction, 44,* 1–10. http://dx.doi.org/10.1016/j.learninstruc.2016.02.006

Hamilton, A. F. C., & Lind, F. (2016). Audience effects: What can they tell us about social neuroscience, theory of mind and autism? *Culture and Brain, 4,* 159–177. http://dx.doi.org/10.1007/s40167-016-0044-5

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

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,* 23–32. http://dx.doi.org/10.1037/arc0000003

Hinds, P. J., Patterson, M., & Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology, 86,* 1232–1243. http://dx.doi.org/10.1037/0021-9010.86.6.1232

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. http://dx.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. http://dx.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. http://dx.doi.org/10.1037/edu0000272

Kalyuga, S., Renkl, A., & Paas, F. (2010). Facilitating flexible problem solving: A cognitive load perspective. *Educational Psychology Review, 22,* 175–186. http://dx.doi.org/10.1007/s10648-010-9132-9

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. http://dx.doi.org/10.1016/j.learninstruc.2017.04.005

Kiewra, K. A. (2005). *Learn how to study and SOAR to success*. Upper Saddle River, NJ: Pearson Prentice Hall.

Kobayashi, K. (2018). Learning by preparing-to-teach and teaching: A meta-analysis. *The Japanese Psychological Research*. Advance online publication. http://dx.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,* 401–410. http://dx.doi.org/10.1002/acp.3410

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. *Journal of Experimental Education, 86,* 344–361. http://dx.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,* 19–37. http://dx.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,* 101–115. http://dx.doi.org/10.1037/xap0000038

Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2001). Effects of causal text revision on more or less–skilled readers' comprehension of easy and difficult texts. *Cognition & Instruction, 18,* 525–556. http://dx.doi.org/10.1207/S1532690XCI1804_4

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction, 34,* 74–85. http://dx.doi.org/10.1016/j.learninstruc.2014.08.004

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.005

McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, pp. 297–284). New York, NY: Elsevier Science. http://dx.doi.org/10.1016/S0079-7421(09)51009-2

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

Nückles, M., Wittwer, J., & Renkl, A. (2005). Information about a layperson's knowledge supports experts in giving effective and efficient online advice to laypersons. *Journal of Experimental Psychology: Applied, 11,* 219–236. http://dx.doi.org/10.1037/1076-898X.11.4.219

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

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43,* 121–152. http://dx.doi.org/10.1080/01638530709336895

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19,* 228–242. http://dx.doi.org/10.1016/j.learninstruc.2008.04.003

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. http://dx.doi.org/10.1037/0022-0663.84.4.429

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24,* 27–45. http://dx.doi.org/10.1007/s10648-011-9179-2

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

Plötzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). Learning by explaining to oneself and to others. In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 103–121). 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. http://dx.doi.org/10.1007/s11251-018-9452-6

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,* 215–224. http://dx.doi.org/10.1016/j.jecp.2007.10.002

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. http://dx.doi.org/10.1016/j.tics.2010.09.003

Roelle, J., Schmidt, E. M., Buchau, A., & Berthold, K. (2017). Effects of informing learners about the dangers of making overconfident judgments of learning. *Journal of Educational Psychology, 109,* 99–117. http://dx.doi.org/10.1037/edu0000132

Roscoe, R. D. (2014). Self-monitoring and knowledge-building in learning by teaching. *Instructional Science, 42,* 327–351. http://dx.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,* 321–350. http://dx.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,* 1432–1463. http://dx.doi.org/10.1037/a0037559

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. *Advances in Applied Psycholinguistics, 2,* 142–175.

Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control during learning from text? *Learning and Instruction, 51,* 10–25. http://dx.doi.org/10.1016/j.learninstruc.2017.02.002

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *LGVT 6–12 Lesegeschwindigkeits- und Verständnistest für die Klassen 6–12* [LGVT 6-12 reading speed- and comprehension test for grade 6-12]. Göttingen, Germany: Hogrefe.

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

Sundqvist, M. L., Mäntylä, T., & Jönsson, F. U. (2017). Assessing boundary conditions of the testing effect: On the relative efficacy of covert vs. overt retrieval. *Frontiers in Psychology, 8,* 1018. http://dx.doi.org/10.3389/fpsyg.2017.01018

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2,* 53–55. http://dx.doi.org/10.5116/ijme.4dfb.8dfd

Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology, 102,* 817–826. http://dx.doi.org/10.1037/a0020062

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

Wiens, S., & Nilsson, M. E. (2017). Performing contrast analysis in factorial designs: From NHST to confidence intervals and beyond. *Educational and Psychological Measurement, 77,* 690–715. http://dx.doi.org/10.1177/0013164416668950

Wiley, J., Griffin, T. D., Jaeger, A. J., Jarosz, A. F., Cushen, P. J., & Thiede, K. W. (2016). Improving metacomprehension accuracy in an undergraduate course context. *Journal of Experimental Psychology: Applied, 22,* 393–405. http://dx.doi.org/10.1037/xap0000096

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, Germany: Hogrefe.

Wittwer, J., Nückles, M., Landmann, N., & Renkl, A. (2010). Can tutors be supported in giving effective explanations? *Journal of Educational Psychology, 102,* 74–89. http://dx.doi.org/10.1037/a0016727

Received December 7, 2018
Revision received April 30, 2019
Accepted June 12, 2019 ■