Active ingredients and factors for deep processing during an example-based training intervention

Markus H. Hefter · Kirsten Berthold

Received: 7 February 2020 / Accepted: 24 December 2020 / Published online: 23 January 2021
© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract
In two experiments with German secondary school students ($N_1 = 43; N_2 = 41$), we aimed to analyze and optimize an effective learning environment. We sought further active ingredients and crucial factors for deep processing during an example-based training intervention, which consisted mainly of video examples and self-explanation prompts. In Experiment 1, we analyzed whether presenting learning goals influenced learning processes (i.e. mental effort and self-explanation quality) and learning outcomes (i.e. declarative knowledge about argumentation). In Experiment 2, we examined the role of the ratio between the number of self-explanation and practice tasks. Both experiments revealed that learners highly improved their declarative knowledge on argumentation. These effects remained stable after 2 and 3 weeks—regardless of whether learning goals had been presented or the ratio between self-explanation and practice tasks. Furthermore, mental effort and the Need for Cognition (NFC) were identified and discussed as crucial factors for deep processing the given examples and for the knowledge gain from pretest to delayed posttests. Presenting learning goals reduced mental effort, albeit at the expense of beneficial cognitive processing of the given examples. Finally, NFC contributed to learners’ mental engagement not only during but also after the training intervention.

Keywords Example-based learning · Mental effort · Need for cognition · Self-explanation · Training intervention

Introduction

Example-based training interventions have emerged as promising learning environments that are short-term yet effective for initial knowledge and skill acquisition in various domains. For instance, Berthold and Renkl’s (2010) training intervention fostered the focused processing of explanations. Furthermore, Hoogveld et al. (2005) analyzed a training approach for teachers on educational design. Focusing on video-based examples, Hefter et al. (2014, 2015, 2018) designed training interventions on argumentative knowledge.
Broadly speaking, we consider an example-based training intervention as a learning environment built around examples. Those examples model the to-be-learned principles and are usually accompanied by various instructional measures to ensure the examples’ deep processing. In short, the rationale behind example-based training interventions is to have learners deeply process and thereby learn from given examples. Learning from examples can be very effective, especially for learners with low prior knowledge, because it helps to reduce cognitive overload (see Sweller et al. 2011 for an overview on cognitive load). Examples do not require that learners seek a problem’s solution. Hence, learners could now dedicate their limited cognitive resources to deeply processing the examples’ underlying principles (Paas and van Gog 2006; Renkl 2014; Wittwer and Renkl 2010). Indeed, learners could deeply process those principles, but how do we make them actually engage in doing so?

**Deep processing during an example-based training intervention**

A crucial and well-studied approach for fostering deep processing is to have learners self-explain the underlying principles that are modeled in given examples (Renkl 2014). A self-explanation is “explicating the rationale of the solution for oneself, especially under reference to the underlying domain principles” (Renkl 2011, p. 278). Thus, in terms of an example-based learning environment, self-explaining means generating an explanation for oneself by deeply processing the given example-based material. Generally speaking, self-explaining can be a powerful and deeply constructive learning activity (Roy and Chi 2005)—a learning strategy, so to speak. However, as learners do not automatically generate self-explanations, they often need to be nudged to do so. So-called self-explanation prompts have proven to be effective measures to encourage learners to perform self-explanation activities. As mentioned earlier, such self-explanation activities mean nothing other than to “think deeply and to cognitively engage with the learning materials […]” (Wylie and Chi 2014, p. 423). Those self-explanation prompts are often questions that request learners to generate an explanation for themselves and thereby deeply process the examples’ principles (Berthold and Renkl 2009; Hefter et al. 2014, 2015, 2018; Roelle et al. 2015; Schworm and Renkl 2007; Wittwer and Renkl 2010). For instance, in Hefter et al.’s (2014, 2015) training interventions on argumentative thinking, the self-explanation quality mediated the interventions’ effect on learning outcomes 1 week after the interventions (Hefter et al. 2014, 2015). In other words, the higher the quality of how learners explained the principles to themselves, the stronger the effect on their declarative knowledge about these principles in the delayed posttest. All in all, self-explanation quality indicates the deep processing of the example’s principles, and is beneficial for acquiring knowledge that learners retain for at least a week.

Considering that the deep processing of given examples is such a beneficial learning activity, it makes sense to seek further measures to improve it. Following research recommendations by Renkl (2015), it appears less worthwhile to compare different instructional methods in horse-race studies (Salomon 2002). An example for such a horse race would be a direct comparison between two learning environments, one based on example-based learning and the other problem-based learning. One learning environment would likely win. However, this approach is highly unlikely to provide any insights about which parts of the learning environment—the so-called active ingredients (Clark 2009)—affected the learning process. At the end of the day, because too many factors would probably have varied, generalizability would be limited (Renkl 2015). Finally, it is the instructional method’s
implementation quality that would have played a more important role in winning the learning environments’ race, than whether the winner was actually example- or problem-based.

Hence, we did not aim to compare different learning environments in this study. Rather, we focused on a learning environment that previous studies had already considered effective. We deconstructed it and experimentally adjusted its screws to learn about its mechanisms. Our goal was to analyze active ingredients (Clark 2009) as well as crucial factors for deep processing during the example-based training intervention on argumentative knowledge by Hefter et al. (2014). Various studies (Schworm and Renkl 2007; Hefter et al. 2014, 2015, 2018) have already shown the effectiveness of this training intervention, as well as its forerunner and derivative models. Our present goal was therefore to refrain from simple treatment/control comparisons to avoid triviality. Learners who have undergone a training intervention on argumentative principles surely would be able to recall more of those principles than learners who have undergone no intervention on these principles in the first place. Hence, we hope to shed more light on the active ingredients and crucial factors that actually contribute to the well-known effectiveness of this learning environment to deduce theoretical and practical implications.

Potential active ingredients

Presentation of learning goals

As mentioned before, self-explanation prompts are well established and have undergone intensive analysis as a learning environment’s active ingredients for deep processing when learning from examples. Looking for active ingredients other than prompts, we first suggest a presentation of learning goals before establishing and explaining the basic principles modeled in given examples. Presenting learning goals can be considered an important aspect of effective strategy instruction (e.g. Harris et al. 2008). Studies of reading comprehension show that providing learners with learning goals improved understanding of relevance and thereby learning (e.g. McCrudden and Schraw 2007). Furthermore, from the Perspective of Focused Processing (Renkl 2015), a presentation of learning goals might help learners to focus their limited cognitive resources on the relevant principles to be learned. This seems to be particularly important when a learning environment addresses other topics to exemplify its relevant principles. For instance, the learning goals of Berthold and Renkl’s (2010) training intervention referred to principles about how to process given explanations. During their intervention, they used the topic of heredity to exemplify these principles. The learning goals of Hefter et al.’s (2014) training intervention referred to principles of argumentation. During their intervention, they used the topic of ecology to exemplify these principles. In both learning environments, the relevant principles were just not about biology or ecology, but about explanations or argumentation. Briefly, presenting learning goals should show learners what is relevant, prevent them from focusing too much on the exemplifying topics, and help them to center their attention on the underlying principles. Put another way, it might help to reduce cognitive overload (Sweller et al. 2011) and leave learners with enough cognitive resources for relevant learning processes. These processes comprise the previously-discussed self-explaining.

Against this background, effective example-based training interventions in various studies relied on a presentation of learning goals, such as Berthold and Renkl’s (2010) training intervention on focused processing of explanations, Renkl et al.’s (2013) training intervention on self-explaining, or Hefter et al.’s (2014, 2015, 2018) training interventions on skill
and will of argumentative thinking. These training interventions revealed positive effects on learning processes (such as self-explanation quality) as well as learning outcomes (such as declarative knowledge). However, the authors did not empirically analyze how presenting learning goals contributed to these effects. Rather, the training interventions in question were primarily developed and analyzed as effective interventions as a whole, thus leaving the search for their active ingredients a research desideratum. We therefore analyzed the effects of presenting learning goals on learning processes and outcomes in Experiment 1.

**Ratio of self-explanation and practice tasks**

As another active ingredient for deep processing during an example-based training intervention, we suggest the ratio of self-explanation tasks and practice tasks. This ratio relates to the point of transition from the initial phase of self-explaining the principles of given examples to the later phase of applying those principles by solving a problem. As Renkl (2014) discussed in detail, self-explaining principles of given examples is an effective endeavor at the earlier stages of knowledge and skill acquisition. This can be explained from a cognitive perspective: the more knowledge learners have already acquired, the less complexity—and thus intrinsic load—a given problem-solving task imposes on them (Renkl and Atkinson 2003). Hence, more of learners’ limited and precious cognitive resources become available for problem-solving tasks. Furthermore, according to the Expertise Reversal Effect (Kalyuga et al. 2003), self-explaining can lose its effectiveness for learners with more knowledge. Problem-solving tasks are thus recommended for more-experienced learners. In other words, instructors should provide self-explaining tasks before problem-solving tasks. Against this rationale, Hefter et al. (2014, 2015, 2018) designed the structure of their effective training interventions on argumentative thinking. After a short theoretical introduction on argumentative principles, three learning phases took place. The first two learning phases consisted of self-explanation tasks. Each thereof provided a video-based example and respective self-explanation prompts on the argumentative principles. The third and final learning phase was a practice task to provide the opportunity to solve a problem by applying the argumentative principles without any support. Put simply, the authors implemented a 2:1-ratio of self-explanation and practice tasks. This approach proved effective in that Hefter et al.’s (2014) training intervention fostered learning processes (i.e. self-explanation quality and argument quality) as well as learning outcomes (i.e. declarative and procedural knowledge about argumentation). In particular, the authors detected a positive effect on procedural knowledge—but only immediately after the training intervention. A week later, the effect had vanished. Thus, the question arises if their approach of implementing a 2:1-ratio of self-explanation and practice tasks was already the optimum with respect to effectiveness.

Differentiating between declarative and procedural knowledge, we consider the first to be primarily factual knowledge. In the studies of Schworm and Renkl (2007) or Hefter et al. (2014), declarative knowledge referred simply to listing names and functions of argumentative elements. In contrast, we consider procedural knowledge to be about knowing how to perform complex actions, such as generating actual argumentative elements. It thus seems feasible that, for lasting effects on procedural knowledge, a higher number of practice tasks should be required to produce more enduring effects. While keeping the number of learning tasks constant (i.e. three), effects of implementing either a 2:1- or 1:2-ratio of self-explanation and practice tasks on learning processes and outcomes (both declarative and procedural knowledge) were analyzed in Experiment 2.
Further potential factors

Mental effort

The training interventions described above aimed to have learners engage in deeply processing given examples. To do so, they should focus their limited cognitive resources on the respective learning task. Because of learners’ limited cognitive resources, instructional designers should be aware of the risk of overtaxing them. The load on these resources—that is, cognitive load (see Sweller et al. 2011 for an overview)—is a crucial factor in instructional effectiveness. It can be differentiated into three types: intrinsic, extraneous, and germane load. The complexity of the learning material itself already imposes demands on learners’ cognitive resources (intrinsic cognitive load), especially for those with low prior knowledge. So do the manner in which the learning material is presented (extraneous cognitive load) and the actual learning processes (germane cognitive load) (Sweller et al. 1998). All these demands on learners’ cognitive resources contribute to “the total amount of cognitive processing” (Paas and Van Merriënboer 1993, p. 738), which depends on the mental effort a learner invests in a given task. Thus, mental effort is allocated to the three categories (i.e. intrinsic, extraneous, and germane) of cognitive load (Sweller 2005), but it does not specify to which category it relates.

Obviously, mental effort is needed to deeply process the given examples, which would relate to germane cognitive load. The more germane is cognitive load invested, the better the quality of the learning process. Hence, there should be a positive correlation between mental effort and self-explanation quality. However, high mental effort could also be a sign that a learner is overstrained because the cognitive resources are strongly affected by the learning task’s complexity (intrinsic cognitive load) or by inapt instructional design (extraneous cognitive load). Finally, mental effort is also needed to hold information in memory, which could be the introductory information about the to-be-learned principles. In a nutshell, mental effort alone seems to be an ambiguous measurement, because it does not specify to which type of cognitive load or combinations of types it refers. It should be combined with a performance measure to analyze the mental costs of a training intervention (Paas and Van Merriënboer 1993). To shed some light on these issues, we analyzed the influence of mental effort as a factor on learning processes and outcomes in Experiments 1 and 2.

Need for cognition

Whereas mental effort can at least be partly influenced by instructional design, we finally suggest a learners’ attribute as another factor that influences learning processes and outcomes: Need for Cognition (NFC). NFC is a personality trait that refers to an individual’s enjoyment “engaging in effortful and complex thinking” (Nussbaum and Bendixen 2003, p. 576). Learners with a high NFC share a positive attitude toward engaging in cognitive tasks (Cacioppo and Petty 1982). They “naturally tend to seek, acquire, think about, and reflect back on information” (Cacioppo et al. 1996, p. 243). Thus, individuals with a high NFC would be more likely to engage in cognitive activities such as self-explaining and thereby deeply processing given examples. As mentioned above, the deep processing of given examples goes hand-in-hand with high self-explanation quality and the retention (for at least 1 week) of declarative knowledge about the examples’ principles. Put another way, NFC is likely to influence these learning benefits positively. We analyzed such assumptions in Experiment 2 to deduce potential practical implications for an effective example-based learning environment.
Experiment 1

Hypotheses

In Experiment 1, our goal was to examine the effect of presenting learning goals as a potential active ingredient in an example-based training intervention on argumentative knowledge. We aimed to investigate whether presenting learning goals influenced a given training intervention’s effect on learning processes (i.e. self-explanation quality) and learning outcomes (i.e. declarative knowledge). Against the background of the aforementioned theoretical considerations, presenting learning goals is likely to provide beneficial support for learners. We thus assumed that the presentation of learning goals...

H1 … lowers cognitive load.
H2 … has a positive effect on self-explanation quality.
H3 … has a positive effect on declarative knowledge.

Method

Sample and design

We recruited two classes containing a total of 45 German students in their second to last year (year 11) of a secondary school in the German state of North Rhine-Westphalia. For all participants, we received parental consent and provided laptops and headsets. They were under our supervision in the schools’ classrooms. Because of technical problems with their computers, we had to exclude two participants, resulting in a final sample of 43 ($N=43$, 25 female, 18 male; $M_{age}=16.65$; $SD_{age}=0.78$). We randomly assigned the sample to one of two experimental conditions, each of which featured one of two versions of the computer-based learning environment: (a) training intervention with learning goals (goal condition, $n=23$) and (b) training intervention without learning goals (non-goal condition, $n=20$). Five participants from the goal condition and three participants from the non-goal condition were unable to take part in the delayed posttests after 3 weeks.

Computer-based learning environment

Participants in the goal condition received the computer-based training intervention on argumentative knowledge by Hefter et al. (2014), which consisted of the following components: It began with a presentation of learning goals and a theoretical introduction on argumentative principles. Participants could navigate through this introduction via button clicks. The presentation of learning goals was only shown once at the beginning of the introduction. After the participants had clicked through the introduction, two self-explanations tasks followed. Each self-explanations task featured a video-based example and respective self-explanation prompts on the argumentative principles. The self-explanation prompts should make the learners self-explain the principles that were modelled in the video-based examples. These prompts were questions such as “Which elements of argumentation do you recognize in the sequence from Manuel? What are their functions?” See Fig. 1 for a screenshot of such a prompt with a textbox. By typing in their self-explanations (i.e. their answers to the self-explanation prompts), learners need to deeply process the examples’ principles. Finally, a practice task asked learners to compose their own position
in a text box. This task thus gave them the opportunity to apply the argumentative principles without any support. The topics of the self-explanation and practice tasks concerned ecological positions, namely, ‘the ecological consequences of resettling the lynx, the consequences of global warming on forest dieback, and the ecological consequences of cultivating genetically-engineered plants’.

Participants in the non-goal condition received the same computer-based training intervention except for one detail: the learning goals were not mentioned. Thus, the non-goal condition started immediately with the theoretical introduction on argumentative principles.

**Instruments**

**Declarative knowledge**

We assessed *declarative knowledge* about argumentative principles as pretest, posttest, and delayed posttest. As we used Hefter et al.’s (2014) learning environment on argumentative knowledge, we also used their one open-format question to assess declarative knowledge about argumentation: “What are the elements of good argumentation, and what function does each element serve?” We applied a 6-point scale from 1 (*very low quality*) to 6 (*very high quality*) to rate the participants’ answers and gave the maximum rating of 6 to an answer that named or described all six argumentative elements (i.e. theory, genuine evidence, alternative theory, counterargument, rebuttal, and synthesis).

**Declarative knowledge gain**

We calculated the (delayed) posttest’s values minus the pretest’s values to arrive at the *declarative knowledge gain*. 
Self-explanation quality

To rate the participants’ answers to each of the 8 self-explanation prompts during the training intervention, we used a 6-point scale ranging from 1 (very low quality) to 6 (very high quality). We gave the maximum rating of 6 to an answer that contained correct and exhaustive self-explanations (Hefter et al. 2014).

For each of our ratings of declarative knowledge and self-explanation quality, a student research assistant and the first author rated a randomly-selected set of ~23% of our sample’s (i.e. 10 participants’) data. Because of high interrater reliabilities (intraclass coefficient for a two-way mixed model with measures of absolute agreement, ICC_{declarative knowledge} = .96, ICC_{self-explanation quality} = .88), we had our student research assistant rate the remaining data.

Cognitive load

To assess cognitive load, we applied the one-item 9-point rating scale by Paas (1992) on the subjective invested mental effort (“How much effort did you invest in the last task?”). This scale is used frequently in instructional research (Le et al. 2018; Tabbers et al. 2004; van Gog et al. 2012). We applied it three times during training intervention: after the participants answered the last self-explanation prompt referring to the first video-example; after they answered the last self-explanation prompt referring to the second video-example; and after they finished the practice task.

Learning time

Learning time was calculated as the difference between the logged time stamps when the participants started and finished working with the training intervention.

Procedure

Based on previous studies reporting similar interventions and our own pretesting, we assumed a mean learning time of about 50 min. To ensure an unstressed procedure including tests, including greetings and saying thank you, we arranged a 90-min timeslot in class (two consecutive lessons). First, participants filled out a demographic questionnaire (about sex, age, and school grades) and a pretest on declarative knowledge. They then learned in their respective computer-based learning environment, during which we assessed self-explanation quality, cognitive load, and learning time. Following the intervention, participants received posttests on declarative knowledge. Three weeks later, we administered delayed posttests on declarative knowledge. In the 3 weeks between posttest and delayed posttest, participants had their regular school education with no feedback, comments or any instruction referring to our intervention or tests.

Results for Experiment 1

To test our hypotheses, we used $t$ tests with $d$ as the effect size measure—qualifying values around 0.20 as small, values around 0.50 as medium, and values of 0.80 or more as large effects (Cohen 1988). For correlations, we used Pearson’s correlation coefficient $r$—qualifying values around .10 as small, values around .30 as moderate, and values of 50 or more as
large correlations (Cohen 1988). Referring to our repeated measure of declarative knowledge, we performed a one-way ANOVA and used partial η² as the effect size measure—qualifying values < 0.06 as small effect, values between .06 and .13 as medium effect, and values > .13 as large effect (Cohen 1988). All our statistical analyses refer to an alpha level of .05. See Table 1 for all measures of Experiment 1.

### Learning prerequisites

We observed no statistically significant differences between the experimental groups in the participants’ learning prerequisites, such as school grades, learning time, prior ecological knowledge, or prior declarative knowledge. As students in the second to last year (year 11) of a secondary school, our participants would probably have had some prior experience with argumentation, such as writing dialectical essays. However, as the pretest revealed, prior declarative knowledge about our intervention’s main principles (i.e. the six argumentative elements) was rather sparse.

### Learning processes

We assumed that the presentation of learning goals would support the participants’ learning processes. As anticipated in our first hypothesis (H1), the non-goal group reported having invested greater mean mental effort than the goal group, t(41) = 1.93, p = .030, d = 0.59 (one-sided t test, medium effect).

We found no positive effect of the presentation of learning goals on self-explanation quality (H2). In fact, the pattern that appeared was different. Learners in the non-goal condition actually achieved higher self-explanation quality than learners in the goal condition, t(41) = 2.10, p = .042, d = 0.64 (two-sided, medium effect).

Furthermore, the mean mental effort and self-explanation quality correlated positively, r(41) = .40, p = .004 (one-sided, moderate correlation). When comparing these correlations between the two conditions, the small sample size might have compromised

---

**Table 1** Means (with standard deviations in parentheses) for all measures in Experiment 1

| Measures                              | Goal condition | Non-goal condition |
|---------------------------------------|----------------|-------------------|
| Declarative knowledge<sup>a</sup>     |                |                   |
| Pretest                               | 2.00 (1.24)    | 2.68 (1.53)       |
| Posttest                              | 2.52 (1.75)    | 3.00 (1.28)       |
| Delayed posttest                      | 2.89 (0.90)    | 3.41 (1.33)       |
| Declarative knowledge gain<sup>b</sup> | 0.62 (1.83)    | 0.47 (2.15)       |
| Immediate                             |                |                   |
| After 2 weeks                         | 0.89 (1.45)    | 0.69 (1.78)       |
| Self-explanation quality<sup>c</sup>  | 3.38 (0.74)    | 3.86 (0.76)       |
| Cognitive load<sup>d</sup>            | 4.72 (1.42)    | 5.70 (1.88)       |
| Learning time<sup>d</sup>             | 55.08 (10.73)  | 56.51 (15.43)     |

<sup>a</sup>6-point rating scale from 1 (very low quality) to 6 (very high quality)

<sup>b</sup>Difference between values from (delayed) posttest and pretest

<sup>c</sup>Scale from 1 to 9, with higher values reflecting higher invested mental effort

<sup>d</sup>Time in minutes
further analyses, but there were hardly any differences, $r_{goal\ condition}(21) = .37, p = .041$ and $r_{non\-goal\ condition}(18) = .32, p = .086$.

Finally, we analyzed the relationship between mental effort and knowledge gains. Mental effort and immediate knowledge gain correlated positively, $r(36) = .35, p = .017$ (one-sided, moderate correlation) as also did mental effort and delayed knowledge gain, $r(32) = .48, p = .002$ (one-sided, moderate correlation).

**Learning outcomes**

We had assumed that the presentation of learning goals would influence participants’ learning outcomes (H3). We identified no statistically significant differences between the experimental groups in the participants’ declarative knowledge immediately after the training intervention, $t(37) = 0.96, p = .173$, or after 3 weeks, $t(33) = 1.37, p = .090$ (all one-sided t tests).

Furthermore, to analyze effectiveness on declarative knowledge (Fig. 2), we conducted a one-way repeated-measures ANOVA with condition (i.e. goal vs. non-goal) as a between-subjects factor and measurement time (pretest vs. delayed posttest) as a within-subjects factor. Those results revealed a statistically significant effect of measurement time, $F(1, 32) = 8.10, p = .008$, partial $\eta^2 = .20$ (large effect). However, there was no statistically significant effect of condition, $F(1, 32) = 4.00, p = .054$, and no statistically significant interaction effect, $F(1, 32) = 0.13, p = .719$.

With respect to a connection between learning processes and learning outcomes, self-explanation quality correlated positively with declarative knowledge after 3 weeks, $r(33) = .37, p = .015$ (one-sided, moderate correlation) but not with declarative knowledge immediately after the training, $r(37) = .01, p = .470$ (one-sided).

![Fig. 2 Declarative knowledge rating in Experiment 1](image-url)
Discussion for Experiment 1

In Experiment 1, we examined the effect of presenting learning goals as a learning environment’s potential ‘active ingredient’ (Clark 2009) for deeply processing the presented video-examples. As expected, a presentation of learning goals resulted in lower cognitive load or, more precisely, a lower mean of reported mental effort (H1). This presentation of learning goals, however, actually seemed to be detrimental to the self-explanation quality (H2). Learners not given learning goals generated self-explanations of higher quality than those who had received learning goals. This unexpected result raises the question of whether and how the learning goals were formulated might have actually handicapped self-explaining. After all, the learning goals clearly focused on what learners should grasp (i.e. principles of argumentation) and not explicitly on what learners should do (i.e. self-explain) in the learning environment. For future investigations, including specific goals involving self-explanation might prove fruitful.

Furthermore, self-explanation quality and reported mental effort correlated positively. It thus seems feasible to assume that in this experiment, mental effort referred to engaging in cognitive processes that benefit the deep processing of presented video-examples. After all, mental effort correlates positively with knowledge gains also. However—as we consider this more thoroughly in the General Discussion—a mental effort assessment does not distinguish between the three categories of cognitive load (de Jong 2010). All in all, providing a presentation of learning goals seems to have reduced the total amount of cognitive processing. However, this reduction occurred at the expense of beneficial cognitive processing of the given examples.

With regard to learning outcomes, we observed a large effect of measurement time on declarative knowledge. Learners in both groups demonstrated more declarative knowledge after the intervention and even 3 weeks later than right before the intervention. The reason for this improvement can only be the training intervention itself. Indeed, learners had no other access to declarative information about argumentative principles. The training was the sole source of information that they could have retrieved to score on the declarative knowledge test. Additionally, the immediate posttest on declarative knowledge might also have contributed to further consolidating that knowledge and thus a better result in the delayed posttest (see retrieval practice effect, Roediger and Butler 2011).

The declarative knowledge that was retained for 3 weeks after the intervention correlated positively with self-explanation quality. In other words, the better learners explained the examples’ principles to themselves, the more they knew about those principles 3 weeks later. This finding underscores the long-term benefits of the deep processing of given examples, as already shown in previous studies such as that by Hefter et al. (2014), who found that self-explanation quality was a predictor for learning outcome in delayed posttests.

In contrast, declarative knowledge assessed immediately after the intervention did not correlate with self-explanation quality. This pattern of results is plausible, because to score a high declarative-knowledge rating at the posttest, a learner simply has to remember up to six argumentative elements seen only minutes ago. Such an achievement does not require deep processing (i.e. self-explaining the video examples). As the recent results of Hefter and Berthold (2020) indicate, it might not even require processing video examples at all—the short introduction before the videos would already suffice.

However, we detected no statistically significant influence of the condition on learning outcomes (H3). While this null finding naturally delivers no evidence for the null hypothesis, at least it indicates that learning goals were not as influential on our training intervention’s learning outcomes as expected. However, it also seems reasonable to surmise that the two
experimental conditions were not that different in terms of their cognitive representation of learning goals. First, the learning-goals presentation was only one additional table right at the beginning of the intervention. This bears the risk that learners might have inadequately processed the learning goals, or even overlooked them. Second, it is also possible that, even without the explicit presentation of learning goals, learners in both groups knew what was being expected from them. The training intervention’s components, such as the theoretical introduction and self-explanation prompts, focused closely on argumentative elements and their function. Thus, it might have been obvious to the learners that they should concentrate on those principles, regardless of an additional presentation of learning goals. Hence, this presentation of learning goals might have been redundant. Overall, the learners’ actual awareness of learning goals should therefore be addressed in future studies. Furthermore, future studies should gather objective data, such as eye tracking and time logs, to analyze whether and how learners process given learning goals. Finally, these studies might also assess the learners’ epistemological understanding. This reflects how reasonable learners consider argumentation (e.g. Mason and Scirica 2006) and might thus be an important moderator of the intervention’s effect.

**Experiment 2**

**Hypotheses**

The first goal of Experiment 2 was to compare the effect of different ratios of self-explanation tasks and practice tasks of an example-based training intervention on argumentative knowledge. We aimed to investigate whether implementing either a 2:1- or 1:2-ratio of self-explanation and practice tasks would influence a given training intervention’s effect on learning processes (i.e. self-explanation quality) and learning outcomes (i.e. declarative knowledge and additionally procedural knowledge).

Against the aforementioned background, we assumed the following hypotheses:

H1 A self-explanation task imposes less cognitive load than a practice task.
H2 The ratio of self-explanation and practice tasks affects learning outcomes (i.e. declarative and procedural knowledge).

The second goal of Experiment 2 was to examine NFC (Need for Cognition) as a crucial learning prerequisite. We assumed that NFC is a beneficial factor for deep processing the principles of given examples. After all, NFC represents a positive attitude toward cognitively-demanding endeavors, which include self-explaining. Furthermore, we were interested in how NFC contributes to the knowledge gain about these principles in the delayed posttest (i.e. the difference between delayed posttest and pretest ratings):

H3 NFC positively affects self-explanation quality.
H4 NFC positively affects a sustained gain in declarative knowledge.

**Method**

**Sample and design**

We recruited 42 German students in 2 classes in their second to last year (year 11) at a secondary school in the German state of North Rhine-Westphalia. This was a different school...
from in Experiment 1. We obtained parental consent and provided laptops with headsets for all participants. Because of computer problems, we had to exclude one participant, resulting in a final sample of 41 (N=41, 26 female, 15 male; M_{age}=16.80; SD_{age}=0.46). We randomly assigned them to one of two experimental conditions, featuring one of two versions of the computer-based learning environment: (a) training intervention with two self-explanation tasks and one practice task (SE-condition, n=20) and (b) training intervention with one self-explanation task and two practice tasks (P-condition, n=21). One participant from the P-condition was unable to take part in the delayed posttests after 2 weeks.

Computer-based learning environment

Participants in both experimental conditions received very similar versions based on the training intervention on argumentative knowledge by Hefter et al. (2014)—analogous to Experiment 1. The only difference between the versions occurred after the first video. As Table 2 shows, the SE-condition was Hefter et al.’s (2014) original training intervention with a 2:1-ratio of self-explanation and practice tasks. However, in the P-condition, we replaced the second self-explanation task with a practice task, resulting in a 1:2-ratio of self-explanation and practice tasks.

Instruments

We used the same instruments as in Experiment 1 for declarative knowledge, cognitive load, self-explanation quality, and learning time. Moreover, we calculated the declarative knowledge gain as the difference between values from the (delayed) posttest and pretest. We also assessed procedural knowledge about argumentative principles as posttest and delayed posttest. As previously mentioned, unlike declarative knowledge that is essentially factual knowledge in list form, procedural knowledge is about performing actions. Hence, whereas the assessment of declarative knowledge referred to knowing the names and functions of argumentative elements, the assessment of procedural knowledge referred to knowing how to actually generate argumentative elements. We assessed such procedural knowledge via Hefter et al.’s (2014) six open-format questions (e.g. “Assume somebody disagrees with your position. What might this person say to show you are wrong?”). We used a 10-point scale of generative knowledge from 0 (very low quality) to 9 (very high quality) to rate participants’ answers, giving the maximum rating of 9 for an answer that consisted of all six argumentative elements (i.e. theory, genuine evidence, alternative theory, counterargument, rebuttal, and synthesis). As we achieved high interrater reliability between a student research assistant and the first author on ~24% of our sample’s (i.e. 10 participants’) data on all our

Table 2  Versions of the computer-based learning environment in Experiment 2

| Phase | SE-condition | P-condition |
|-------|--------------|-------------|
| 1     | Self-explanation task | Practice task |
| 2     | Self-explanation task |            |
| 3     | Practice task |              |
open format instruments (i.e. ICC_{declarative knowledge} = .92, ICC_{procedural knowledge} = .94, ICC_{self-explanation quality} = .98), our student research assistant rated the remaining data. Furthermore, we used the 16 items of the German NFC short scale on Need for Cognition (Bless et al. 1994), whose authors reported high internal consistency (Cronbach’s α = .83). We achieved a similar value (Cronbach’s α = .81). Moreover, with respect to validity, NFC correlated with our participants’ grades in biology (r = .41, p = .005, one-sided) and mathematics (r = .46, p = .002, one-sided). We also asked participants to rate the extent of their interim engagement (“I have engaged with the training’s subject matter…”) using a Likert-type response format ranging from −3 (very little) to 3 (very much).

Procedure

Participants filled out a demographic questionnaire (about sex, age, and school grades) and a paper-and-pencil pretests on declarative knowledge (just to check for possible a priori group differences). Then, the learning environment started, during which we assessed self-explanation quality, cognitive load, and learning time. Finally, participants received posttests on declarative knowledge and procedural knowledge. Two weeks later, we assessed NFC and re-administered the posttests. In between, the students received their regular school education during which there was no reference to our study.

Results for Experiment 2

As in Experiment 1, we performed t tests (d as effect size) and a one-way repeated-measures ANOVA (partial η² as effect size) with an alpha-level of .05. See Table 3 for all measures of Experiment 2.

Learning prerequisites

We identified no statistically significant differences between the experimental groups in the participants’ learning prerequisites such as grades, learning time, prior declarative knowledge, or need for cognition.

Learning processes

We assumed that a practice task imposes more cognitive load than a self-explanation task. Thus, we compared the mental effort that learners reported after their second learning phase. For learners in the SE-condition, this was a self-explanation task. For learners in the P-condition, this was a practice task. As expected, the process of self-explaining principles caused less cognitive load than applying those principles, t(37) = −2.26, p = .015, d = 0.72 (one-sided t test, medium effect).
Learning outcomes

We assumed that implementing either a 2:1- or 1:2-ratio of self-explanation and practice tasks would affect the interventions’ effectiveness on learning outcomes. However, we found no statistically significant differences between the experimental groups in the participants’ declarative knowledge, t(34) = −0.44, p = .661, declarative knowledge after 2 weeks, t(37) = −0.65, p = .519, procedural knowledge, t(33) = 0.35, p = .729, or procedural knowledge after 2 weeks, t(38) = −0.26, p = .796 (all two-sided t-tests).

To analyze the effectiveness on declarative knowledge (Fig. 3), we conducted a one-way repeated-measures ANOVA with condition (i.e. SE- vs. P-condition) as a between-subjects factor and measurement time (pretest vs. delayed posttest) as a within-subjects factor. We noted a statistically significant effect of measurement time, F(1, 35) = 23.05, p < .001, partial η² = .40 (large effect), but no statistically significant effect of condition, F(1, 35) = 0.13, p = .724 and no statistically significant interaction effect, F(1, 35) = 0.71, p = .406.

Table 3  Means (with standard deviations in parentheses) for all measures in Experiment 2

| Measures                                      | SE-condition | P-condition |
|-----------------------------------------------|--------------|-------------|
| Declarative knowledgea                        |              |             |
| Pretest                                       | 2.42 (1.12)  | 2.35 (1.04) |
| Posttest                                      | 3.53 (1.73)  | 3.76 (1.38) |
| Delayed posttest                              | 3.45 (1.91)  | 3.79 (1.27) |
| Declarative knowledge gainb                   |              |             |
| Immediate                                     | 1.47 (1.89)  | 1.50 (1.50) |
| After 2 weeks                                  | 1.05 (1.87)  | 1.50 (1.30) |
| Procedural knowledgec                         |              |             |
| Posttest                                      | 2.20 (2.01)  | 2.00 (1.38) |
| Delayed posttest                              | 1.95 (1.85)  | 2.10 (1.80) |
| Self-explanation quality (phase 1)a           | 3.50 (0.78)  | 3.08 (0.78) |
| Cognitive loadd                              |              |             |
| Phase 1                                       | 5.65 (1.31)  | 5.48 (1.54) |
| Phase 2                                       | 4.56 (1.58)  | 5.71 (1.62) |
| Phase 3                                       | 4.76 (2.08)  | 5.48 (1.72) |
| Learning timee                               | 53.88 (13.72)| 49.88 (13.27)|
| Need for cognitionf                          | 0.58 (0.81)  | 0.25 (0.77) |
| Interim engagementg                          | 1.40 (0.60)  | 1.40 (0.60) |

*a*6-point rating scale from 1 (very low quality) to 6 (very high quality)

*b*Differences between values from (delayed) posttest and pretest

*c*10-point rating scale from 0 (very low quality) to 9 (very high quality)

*d*Scale from 1 to 9, higher values reflecting higher invested mental effort

*e*Time in minutes

*f*Scale from −3 to +3, higher values reflecting higher need for cognition

*g*Scale from 1 to 5, higher values reflecting more interim engagement
NFC (need for cognition)

We assumed that NFC positively affects self-explanation quality. Indeed, NFC correlated with self-explanation quality, $r(38) = .48$, $p < .001$ (moderate correlation, one-sided). Moreover, we further analyzed the sustained declarative knowledge gain that our previous repeated-measures ANOVA had revealed. We calculated a multiple linear regression to predict learners’ gain in declarative knowledge after 2 weeks based on two variables: The first—and rather obvious—predictor was the immediate knowledge gain after the training. The second predictor was NFC, because it reflects a penchant for mental engagement and thus for deep processing the learning material—ideally, deep enough to remember it after 2 weeks. We detected a statistically significant regression, $F(2, 30) = 15.36$, $p < .001$, $R^2 = .51$. Immediate knowledge gain was a significant predictor, $\beta = 0.68$, $t(32) = 5.27$, $p < .001$ (one-sided t test), as was NFC, $\beta = 0.26$, $t(32) = 2.05$, $p = .024$ (one-sided t test). Multicollinearity was no concern for any predictor (Immediate knowledge gain: Tolerance = .996, VIF = 1.004; NFC: Tolerance = .996, VIF = 1.004). All in all, the participants’ knowledge gain after 2 weeks amounted to $0.04 + 0.70$ (immediate knowledge gain) $+ 0.58$ (NFC) (Fig. 4).

Discussion for Experiment 2

In Experiment 2, we compared the effect of implementing either a 2:1- or 1:2-ratio of self-explanation and practice tasks during an example-based training intervention on argumentative knowledge. As expected with respect to learning processes, learners assigned to a self-explanation task reported less mental effort than learners given a practice task (H1). This is in line with previous research comparing learning from examples and learning from solving problems. In earlier stages of knowledge and skill acquisition, learning from examples is more effective than learning from solving problems. This is because learning from examples does not overstrain the learners’ limited cognitive resources, as
problem-solving tasks would do. However, there is one possible limitation to consider: The mental effort during the task of either self-explaining or practicing could be influenced by prior knowledge about that task. For learners in the P-condition, this was their first practice task whereas, for learners in the SE-condition, it was their second self-explanation task. Thus, familiarity of the task might have lessened the mental effort that learners invested in the SE-condition.

With respect to learning outcomes (H2), learners demonstrated more declarative knowledge on argumentative principles even 2 weeks after the intervention than they had right before the training—regardless of the condition. Whether one combines two self-explanation tasks with one practice task or one self-explanation task with two practice tasks, both training versions improved learners’ declarative knowledge statistically significantly. Perhaps the intervention’s components that both conditions featured—namely, the presentation of learning goals, theoretical introduction, and the first self-explanation task—already contributed exhaustively to this improved knowledge. Following these highly effective components and preceding a final practice task, a second self-explanation task might be just as effective as a practice task. This effectiveness might also be attributable to learners’ relatively low prior knowledge. For learners with more prior knowledge, a possible expertise reversal effect (Kalyuga et al. 2003) should be considered. Finally, and analogous to Experiment 1, the immediate posttest on declarative knowledge might have helped learners to further consolidate declarative knowledge.

We observed no statistically significant difference between the experimental groups with respect to procedural knowledge. We thus assume that neither version of the intervention
was effective enough to foster procedural knowledge that was retained for 2 weeks after the training. After all, in Hefter et al.’s (2014) training intervention with two self-explanation tasks followed by one practice task, the effect on procedural knowledge vanished after 1 week. Furthermore, in Hefter et al.’s (2018) training intervention, one self-explanation task followed by one practice task was insufficient to enable even an immediate effect on procedural knowledge. Consequently, fostering procedural knowledge that is retained for 2 weeks requires more than just three learning phases or even additional training interventions. Future studies might analyze more than three learning phases and thus ratios other than only a 2:1- and 1:2-ratio of self-explanation and practice tasks.

Furthermore, the ratings of procedural knowledge for both conditions and both measurement times are surprisingly lower than those that Hefter et al. (2014) achieved. Apparently, our sample performed less well than Hefter et al.’s (2014). Admittedly, our sample received no cash incentives (as opposed to 32€ per person) and was slightly younger ($M_{age} = 16.80$, $SD_{age} = 0.46$ in contrast to $M_{age} = 17.76$; $SD_{age} = 0.93$).

Nevertheless, procedural knowledge is knowledge about performing complex actions. According to Anderson’s (1983, 1993) ACT-R theory, it is built on available declarative knowledge. In our case, knowing argumentative elements and their functions (i.e. declarative knowledge) might be the foundation for acquiring knowledge of how to actually generate these elements (i.e. procedural knowledge). Thus, the training intervention that we tested (either version) can serve as a promising springboard for establishing a fruitful base of declarative knowledge for further training interventions on procedural knowledge.

With respect to Experiment 2’s second goal, we focused on NFC as a crucial learning prerequisite. NFC statistically significantly correlated positively with self-explanation quality (H3). This result suggests that the extent to which learners were inclined towards effortful cognitive activities (i.e. high NFC) was associated with the extent to which they deeply processed the argumentative principles of given examples during the training intervention.

Self-explanation quality was a crucial mediator of the effect of Hefter et al.’s (2014) training intervention on declarative knowledge about these principles 1 week later. We did not directly manipulate learners’ deep processing of the argumentative principles in this study. Rather, we analyzed whether NFC as a personality trait might contribute to the amount of enduring knowledge that learners acquire about the argumentative principles. Indeed, NFC was a predictor of the gain in declarative knowledge from pretest to delayed posttest after 2 weeks (H4). As a possible reason for this finding, learners with stronger NFC might have benefited more from learning with the training intervention because they were more engaged in deeper processing of the argumentative principles. Furthermore, learners with stronger NFC might merely have been more likely to engage mentally with argumentative principles after the training intervention. A positive correlation between NFC and interim engagement hints at this assumption, $r(38) = .41, p = .004$ (moderate correlation, one-sided).

**General discussion**

**Main conclusions**

In this article, we focused on analyzing an example-based learning environment’s active ingredients, as well as crucial factors for deep processing by concentrating on an effective
example-based training intervention on argumentative knowledge. In light of our two experiments’ findings, we come to the following three main conclusions:

First, a short-term, example-based training intervention on argumentative knowledge that features a theoretical introduction, self-explanation tasks, and practice tasks helps learners to improve their declarative knowledge: Even 2 and 3 weeks after the intervention, our learners demonstrated more declarative knowledge than before the intervention. This effectiveness was apparent regardless of ingredients such as a presentation of learning goals (see Experiment 1) and of implementing either a 2:1- or 1:2-ratio of self-explanation and practice tasks (see Experiment 2). All versions of the training intervention vastly improved the learners’ declarative knowledge. This result partly replicates those of these interventions’ forerunners (Hefter et al. 2014, 2018) and thus contributes to the now-more-than-ever important research goal of replication (e.g. Yong 2012). Our findings underscore that an example-based training intervention provides an effective learning environment.

Second, mental effort was a crucial factor for deep processing these examples. Although self-reported mental effort does not enable accurate discrimination among the three categories of cognitive load, it correlated positively with self-explanation quality and knowledge gains (see Experiment 1). Hence, we assume that during the intervention, mental effort was mostly required for germane cognitive processes that foster deep processing of the presented video-examples. However, extraneous load (i.e. evoked by the learning environment but unrelated to learning) might have also contributed to mental effort. This would have affected both experimental conditions, though. Apart from that, our learning environment was designed to be low-extraneous-load-inducing by following instructional design recommendations such as avoiding seductive details, considering learner control and segments for the videos etc. (e.g. Mayer and Moreno 2003).

In Experiment 1, presenting learning goals apparently reduced mental effort. It also reduced self-explanation quality and thereby the deep processing of the given examples. At first sight, this result is unexpected, because it might seem contrary to the focused processing stance and the usual advantages of presenting learning goals, namely focusing the learners. However, this result might also be interpreted as a desirable difficulty (Bjork 1994; Bjork et al. 2013). Making the learning process more difficult and challenging for learners can have beneficial effects on retention and transfer. One could argue that, in this specific learning environment, a lack of learning goals provided learners with a greater challenge and encouraged them to invest more mental effort processing the given examples.

Third, NFC (Need for Cognition) is a personality trait that can be identified as a crucial factor for deep processing given examples (see Experiment 2). It correlated positively with self-explanation quality. It stands to reason that it was also found to be a predictor of the declarative knowledge gain 2 weeks after the intervention. Furthermore, it correlated also with the extent to which learners reconsider the training intervention’s principles after the training intervention. Thus, we propose that NFC contributed to learners’ mental engagement both during and after the training intervention.

Limitations

One of this study’s limitations is its focus on one specific training intervention on argumentative knowledge (Hefter et al. 2014). It is necessary to analyze active ingredients and crucial factors on other types of knowledge in other domains, such as mathematics...
Furthermore, the training intervention was designed against the rationale and recommendations (e.g. Mayer and Moreno 2003) of being a short-term yet effective intervention requiring as little extraneous cognitive load as possible. Therefore, all the intervention’s components clearly focused on argumentative principles, lacked seductive details, displayed well-marked coherence, and featured sensible segmentation. These design decisions seem to be a plausible reason for why learners could allocate their mental effort for cognitive processes that foster the deep processing of the presented video-examples (i.e. germane cognitive load) and not for cognitive processes imposed by inapt instructional design (i.e. extraneous cognitive load). In different and probably more complex learning environments taking other instructional designs into consideration, learners’ allocations of mental effort might vary—meaning a presentation of learning goals might actually support them. Likewise, the self-explanation/practice phase ratio might play a bigger role in different learning environments, such as those for more experienced learners.

Another limitation is how we measured cognitive load. We used a widely-applied and established method, namely, the subjective rating of invested mental effort by (Paas 1992). Nevertheless, there are critical aspects such as subjectivity. Furthermore, mental effort does not differentiate the three categories of cognitive load (de Jong 2010) and is more closely related to the total amount of cognitive load (Paas et al. 1994).

As a final limitation, our sample sizes were quite small. This limits the interpretation of those results that did not reach statistical significance, such as the test for a potential effect of learning goals on learning outcomes. Future studies with larger samples might therefore allow more complex analyses, such as multivariate models of the relationship between the learning environments’ active ingredients and learners’ crucial factors. To test our hypotheses, we performed four t tests in Experiment 1 and five t tests in Experiment 2, which might be problematic considering the family-wise error rate. Overall, however, these issues hardly affect our main finding (referring to the repeated-measures ANOVAs). Being quite unaffected by manipulations (i.e. presenting learning goals or the ratio of self-explanation and practice phases), our training intervention’s effect on declarative knowledge persisted for 2 or 3 weeks.

Practical implications

From a more practical perspective, it is reassuring that our example-based training intervention enabled an effective learning environment in which to acquire declarative knowledge. Its effectiveness remained robust over several weeks and was rather unaffected by a presentation of learning goals or by different self-explanation/practice phase ratios. However, our learners had relatively low prior declarative knowledge. More experienced learners might not have benefitted likewise, considering the expertise reversal effect (Kalyuga et al. 2003). Regarding learners’ experience, future studies might analyze how instructors can optimize the learning environment to ensure deep processing in their lessons. Depending on learners’ prior knowledge, instructors—manually or via an algorithm—might adapt the number of self-explanation prompts or the self-explanation/practice phase ratios in the learning environment.

With respect to procedural knowledge though, our learning environment can only serve as a promising and fruitful base to build upon. After all, it seems feasible that constructing procedural knowledge (such as how to generate actual argumentative elements) takes much
more time and effort than constructing declarative knowledge (such as remembering names and functions of argumentative elements). To foster procedural knowledge effectively, we propose more long-term learning environments in the framework of cognitive apprenticeship (Collins et al. 1988). Such an environment might start with rather direct instruction such as modelling and prompts (as in our current intervention). Extending over several sessions, these scaffolds would then gradually fade by providing opportunities for more self-regulated practice tasks. For instance, Stark’s (2004) effective learning environment in business administration started with worked examples, followed by incomplete examples, and finished with problem-solving tasks. Hence, for future studies, we might expand and modify our interventions into several sessions implementing this fading approach over a series of self-explanation and practice phases.

Finally, for instructors, having information on their learners’ NFC might help them to adapt their classroom’s learning environment accordingly. For instance, learners with low NFC might need extra incentives to engage mentally, such as feedback or interactivity. Future studies might analyze further connections between NFC and learning processes by analyzing interactions between NFC and different versions of training interventions with varying incentives such as feedback or interactivity. Such studies could contribute to insights into how instructors or even an algorithm might adapt the learning environment to different learners with varying NFC.

To summarize, a sensibly designed example-based training intervention could provide a highly-effective learning environment for acquiring declarative knowledge. In such a specific learning environment as the one in this article, deciding whether learners are presented with learning goals or if they engage in one or two self-explanation or practice tasks seems rather insignificant for learning outcomes. What does have a significant influence, however, is the extent to which learners deeply process the given examples. This crucial endeavor is manifested in high self-explanation quality, feeds upon learners need for cognition, and benefits the acquisition of declarative knowledge that can be retained for at least 3 weeks after the intervention.

Acknowledgements We would like to thank Isabella Dirks, Alina Mertgen, Cornelia Schulz and Julia Werner for their assistance in conducting the experiments and in coding the qualitative data, and Carole Cuerten for proofreading. Finally, we would like to thank all students who took part in our experiment as well as all teachers and the school personnel for their support.

References

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.
Anderson, J. R. (1993). Rules of the mind. Hillsdale, NJ: Erlbaum.
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.
Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. Educational Psychology Review, 22, 25–40. https://doi.org/10.1007/s10648-010-9124-9.
Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185–205). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/4561.003.0011.
Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. Annual Review of Psychology, 64, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823.
Bless, H., Wänke, M., Bohnr, G., Fellhauer, R. F., & Schwarz, N. (1994). Need for cognition: eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben [Need for cognition: a scale measuring engagement and happiness in cognitive tasks]. Zeitschrift für Sozialpsychologie, 25, 147–154.
Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., Blair, W., & Jarvis, G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197.

Clark, R. E. (2009). Translating research into new instructional technologies for higher education: The active ingredient process. *Journal of Computing in Higher Education, 21*(1), 4–18. https://doi.org/10.1007/s12528-009-9013-8.

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

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children, 8*(1), 2–10. https://doi.org/10.5840/thinking19888129.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105–134. https://doi.org/10.1007/s11215-009-9110-0.

Harris, K. R., Alexander, P., & Graham, S. (2008). Michael Pressley’s contributions to the history and future of strategies research. *Educational Psychology, 43*(2), 86–96. https://doi.org/10.1080/00461520801942300.

Hefter, M. H., & Berthold, K. (2020). Preparing learners to self-explain video examples: Text or video introduction? *Computers in Human Behavior, 110*, 1–9. https://doi.org/10.1016/j.chb.2020.106404.

Hefter, M. H., Berthold, K., Renkl, A., Riess, W., Schmid, S., & Fries, S. (2014). Effects of a training intervention to foster argumentation skills while processing conflicting scientific positions. *Instructional Science, 42*(6), 929–947. https://doi.org/10.1007/s11251-014-9320-y.

Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. *Learning and Instruction, 39*, 11–22. https://doi.org/10.1016/j.learninstruc.2015.05.002.

Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2018). Training interventions to foster skill and will of argumentative thinking. *The Journal of Experimental Education, 86*(3), 325–343. https://doi.org/10.1080/002200973.2017.1363689.

Hoogveld, A. W., Paas, F., & Jochems, W. M. (2005). Training higher education teachers for instructional design of competency-based education: Product-oriented versus process-oriented worked examples. *Teaching and Teacher Education, 21*(3), 287–297. https://doi.org/10.1016/j.tate.2005.01.002.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4.

Le, Y., Liu, J., Deng, C., & Dai, D. Y. (2018). Heart rate variability reflects the effects of emotional design principle on mental effort in multimedia learning. *Computers in Human Behavior, 89*, 40–47. https://doi.org/10.1016/j.chb.2018.07.037.

Mason, L., & Scirica, F. (2006). Prediction of students’ argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction, 16*(5), 492–509. https://doi.org/10.1016/j.learninstruc.2006.09.007.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/S15326985ep3801_6.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19*(2), 113–139. https://doi.org/10.1007/s10648-006-9010-7.

Nussbauma, E. M., & Bendixen, L. D. (2003). Approaching and avoiding arguments: The role of epistemological beliefs, need for cognition, and extraverted personality traits. *Contemporary Educational Psychology, 28*(4), 573–595. https://doi.org/10.1016/S0361-476X(02)00620-0.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics—A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429.

Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*(2), 87–91. https://doi.org/10.1016/j.learninstruc.2006.02.004.

Paas, F., & Van Merriënoer, J. J. G. (1993). The efficiency of instructional conditions—An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737–743. https://doi.org/10.1177/001872089303500412.

Paas, F., Van Merriënoer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1), 419–430. https://doi.org/10.2466/pms.1994.79.1.419.

Renkl, A. (2011). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 272–295). New York: Routledge.
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*(1), 1–37. https://doi.org/10.1111/cogs.12086.

Renkl, A. (2015). Different roads lead to Rome: The case of principle-based cognitive skills. *Learning: Research and Practice, 1*(1), 79–90. https://doi.org/10.1080/23735082.2015.994255.

Renkl, A. (2017). Learning from worked-examples in mathematics: Students relate procedures to principles. *ZDM Mathematics Education, 49*(4), 571–584. https://doi.org/10.1007/s11888-017-0859-3.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3.

Renkl, A., Solymosi, J., Erdmann, M., & Aleven, V. (2013). Training principle-based self-explanations: Transfer to new learning contents. *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*, 1205–1210.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003.

Roelle, J., Lehmkühl, N., Beyer, M.-U., & Berthold, K. (2015). The role of specificity, targeted learning activities, and prior knowledge for the effects of relevance instructions. *Journal of Educational Psychology, 107*(3), 705–732. https://doi.org/10.1037/edu0000010.

Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271–286). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511681918.018.

Salomon, G. (2002). Technology and pedagogy: Why don’t we see the promised revolution? *Educational Technology, 42*(2), 71–75.

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*(2), 285–296. https://doi.org/10.1037/0022-0663.99.2.285.

Stark, R. (2004). Implementing example-based learning and teaching in the context of vocational school education in business administration. *Learning Environments Research, 7*(2), 143–163. https://doi.org/10.1023/B:LERI.0000037197.78134.cd.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511681689.003.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer. https://doi.org/10.1007/978-1-4419-8126-4.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205.

Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*, 71–81. https://doi.org/10.1348/000709904322848824.

van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favour of repeated measures. *Applied Cognitive Psychology, 26*, 833–839. https://doi.org/10.1002/acp.2883.

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review, 22*(4), 393–409. https://doi.org/10.1007/S10648-010-9136-5.

Wylie, R., & Chi, M. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 413–432). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.021.

Yong, E. (2012). Replication studies: Bad copy. *Nature, 485*(7398), 298–300. https://doi.org/10.1038/485298a.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.