

Article

# Effects and Prerequisites of Self-Generation in Inquiry-Based Learning

Irina Streich \* and Jürgen Mayer

Institute for Biology, Biology Education, University of Kassel, 34119 Kassel, Germany; jmayer@uni-kassel.de

\* Correspondence: i.streich@uni-kassel.de

Received: 3 August 2020; Accepted: 26 September 2020; Published: 10 October 2020



**Abstract:** The goal of this study is to investigate the effect of self-generation in inquiry-based learning and to identify the role of feedback. While open-ended inquiry-based learning with a high degree of self-generation requirements has long been considered optimal for facilitating effective learning, its long-run effects have been critically challenged. This study employed a 3 (learning condition)  $\times$  2 (retention interval) mixed factorial design ( $N = 98$ ). An inquiry activity involving the self-generation of content knowledge with or without subsequent feedback was compared to an inquiry task in which students simply read hypotheses and data interpretations. Self-generation without feedback was subject to rereading and self-generation with feedback. However, no differences were found under the two latter conditions. An additional analysis of individual learners' abilities revealed that different abilities (e.g., cognitive load, self-generation success) served as predictors of performance in the disparate treatments.

**Keywords:** inquiry-based learning; generation effect; self-generation; retention; self-generation success; cognitive load

## 1. Introduction

Inquiry-based learning is an activity-oriented, student-centered, and collaborative learning approach in which students generate new knowledge by employing an idealized hypothetico-deductive method [1]. It represents an important approach to teaching and learning in science classes. While direct instruction advocates argue that discovery methods like inquiry restrict long-term retention and overstrain human working memory by increasing cognitive load [2,3], their opponents state that open-ended instructional methods promote deeper learning and understanding through their own exploration [1,4]. The component of (self-) generation of knowledge represents a key factor in this debate. The effect of generation (*generation effect*) [5,6] is a powerful encoding phenomenon that asserts a memory advantage for self-generated stimuli in comparison to material that was simply read. The effect was found to sustain retention intervals of up to a week [7,8]. However, previous studies were mostly limited to rather simple familiar material (e.g., synonyms and rhymes) primarily in controlled laboratory studies with relatively homogeneous groups of (adult) subjects who worked individually. Thus, whereas many laboratory studies have shown that simple actively generated information is retrieved more successfully than passively learned information, effects on complex content knowledge acquired in a natural learning environment like inquiry remain unclear [9].

### 1.1. Self-Generation within Inquiry-Based Learning

Self-generation plays a unique role in inquiry-based learning. During inquiry, students investigate authentic scientific problems related to a specific scientific phenomenon by forming hypotheses, planning and carrying out experiments, and analyzing the resulting data. This allows them to acquire both content knowledge (declarative knowledge) and scientific understanding of experimental

procedures (procedural knowledge), also known as scientific reasoning skills or inquiry skills [10,11]. While content knowledge tends to play an important role during the phases of *generating hypotheses* and *analyzing data*, scientific reasoning skills are critical for *planning experiments* and *discussing results*. Moreover, the degree of open-endedness of both the methodological and content phases can vary. It is linked to students' autonomy within different inquiry levels (see Table 1). At the lowest level (*verification inquiry*), the teacher provides students with a lot of information (question, methods, interpretation) and instructional support, whereas in the highest inquiry level (*open inquiry*) students develop and organize their learning process themselves like real scientists [12]. Thus, on the instructional level, the degree of open-endedness is determined by the teacher's inputs and the extent of instructional guidance. Regarding the cognitive process, a high degree of open-endedness is characterized by high self-generation requirements. Self-generating requires students' self-construction of knowledge—containing ideas that go beyond the presented information—and is achieved when students infer new knowledge and integrate new information with existing knowledge [13].

**Table 1.** Levels of Inquiry [14] adapted from Schwab [15] and Colburn [16], modified by Kaiser and Mayer.

Phases	Levels of Inquiry			
	Verification	Structured	Guided	Open
Source of the question	Given	Given	Given	Open
Data collection methods	Given	Given	Open	Open
Interpretation of results	Given	Open	Open	Open
<b>Self-generation requirement</b>				
	low			high
<b>Instructional support</b>				
	high			low
<b>Cognitive load</b>				
	low			high

Note. Given = Given by teacher, Open = Open to student.

Supporters of maximum open-endedness (*open inquiry*) consider strategies with high self-generation requirements and withholding information (e.g., [17]) to be optimal for promoting effective learning; moreover, some recent findings claim that only high self-generation requirements have a long-term impact on learning outcomes [18]. Conversely, other findings reveal that high self-generation requirements impede learning [19], as greater open-endedness is often accompanied by a higher element of interactivity and thus higher *cognitive load* [19,20]. Learners who must process a large number of elements in working memory while also generating new ideas from instructions may experience increased *intrinsic cognitive load* if they do not receive guidance (e.g., [19,21]). Thus, generating new ideas and giving them a sense of coherence in realistic learning situations may impede learning, with the result that less knowledge can be acquired and transferred to long-term memory (e.g., [2,20]). Kirschner and colleagues [2] point out that learners' cognitive load can only be reduced, and strong long-term learning outcomes arise by providing appropriate assistance (through *guided/structured/verification inquiry*).

Research on the single crucial component of self-generation in the context of inquiry-based learning could contribute to clarification in this debate. In a recent study by Kaiser, Mayer, and Malai [22] on the self-generation of scientific reasoning skills (procedural knowledge), reading had a clear advantage over self-generation in facilitating students' acquisition of scientific reasoning skills immediately after

the inquiry task. However, after a one-week delay, students with high self-generation success exhibited a clear generation effect. Similar research on the self-generation of declarative knowledge in the context of inquiry-based learning, as well as the impact of feedback on self-generation success and retention, is still outstanding.

### 1.2. Prerequisites of Self-Generation within an Authentic Learning Environment

Despite all the evidence supporting the robust positive effects of self-generation, the effectiveness of self-generation is built on essential prerequisites that can be inferred from previous studies concerning (1) the *learning content*, (2) the *learning environment*, and (3) *learners' characteristics*.

**Learning content:** The generation effect only arises for information based on preexisting knowledge [23,24]. Self-generation has no or a greatly reduced effect on the recognition of unfamiliar material like non-words or new material like unfamiliar sentences from textbooks [25,26]. As preexisting representations are necessary for active self-generation to benefit item memory, self-generation may increase semantic processing [27]. In fact, only a few studies have examined and found the generation effect using educationally-relevant science material (e.g., [22,28,29]). Foos et al. [28] discuss how effects have rarely been found in applied settings, because examining total test performance rather than just successfully generated items masked the effect, which can only be expected among successfully generated items, not for non-generated items. The research of Kaiser et al. [22] on self-generation of procedural knowledge using an inquiry-based learning environment confirms this claim. The findings reveal that students with a lower error rate during learning in the generation condition, and thus higher self-generation success, exhibit better long-term learning outcomes than students in the reading condition with comparable grades (in German, biology and mathematics) within the same school type (for details, see Appendix B). Furthermore, studies in cognitive psychology have demonstrated a generation effect not only when it comes to strategies and procedures such as conducting inquiry or multiplying and adding numbers, but also for declarative knowledge. However, the effects for procedural knowledge are much larger than for declarative knowledge [9].

**Low cognitive load and feedback:** The construction of knowledge is always accompanied by certain difficulties, especially in authentic learning environments. Realistic learning situations, particularly in science education, involve a high element of interactivity—referred to as *cognitive load* [20]. Dynamic and effortful active learning techniques, such as generating, i.e., connecting, distinguishing, organizing, and structuring ideas, require a considerable investment of cognitive effort and time [30]. Providing greater guidance can help to reduce mental exertion [31]. Further, actively generating complex knowledge in an authentic learning environment instead of just reading information inevitably results in students making mistakes. According to results by Metcalfe and Kornell [32], generating errors does not result in interference as long as the errors are corrected. Previous studies with simple learning material have never found a detrimental effect of generating errors as long as feedback is given. In fact, feedback improved performance and learning outcomes [32,33]: Clarity, purposefulness, meaningfulness, and compatibility with learners' prior knowledge are important features of effective feedback. It should include logical links to exciting knowledge and instructions for active information processing. Furthermore, it must be on a low complexity level and relate to concrete targets [34], since simple feedback tends to be even more effective than complex [35]. Diverse meta-analyses have revealed that even the most common type of feedback—called *corrective feedback* or *knowledge of the correct response*—can be very powerful (e.g.,  $d = 1.13$  [36]). It has proven to be most effective when it refers to interpretation and leads to more effective and efficient strategies for processing and understanding the learning content [37]. Kulik and Kulik [38] reported that errors occurring while engaging in processing classroom activities need to be corrected immediately ( $d = 0.28$ ). Especially when performance levels are relatively low and feedback is provided during an intervening short answer task, performance improves [32,33]. In fact, immediate error correction during task acquisition can result in faster acquisition rates and provide a greater boost to final retention one week later [39]. Finally, feedback not only helps compensate for generation failures and repair faulty knowledge, but it

also contributes to reinforcing the corrected response and increases retention [39]. However, providing too much feedback at the task level can decrease performance, because it may encourage students to focus only on short-term goals. To avoid this, further instruction instead of additional feedback information should be given [34].

High Need for Cognition (NFC): Regardless of the effectiveness of self-generation, in the end, whether or not students apply a learning strategy depends on their cognitive abilities and their intrinsic cognitive–motivational disposition—the tendency to engage in and enjoy effortful cognitive endeavors (or activities with a high cognitive load) called *need for cognition* [40]. NFC plays a distinctive role in describing and predicting information processing in addition to cognitive abilities (e.g., [40]). Recent studies show that higher NFC correlates with deeper learning strategies [41] and positively relates to attitudes and use of learning strategies like generating of materials [42].

All factors have to be considered when a constructive learning strategy like self-generation is applied in an authentic learning environment in order to achieve long-term retention. Experimental research on the influence of new and complex content-related (1) information generated in a realistic teaching situation like inquiry-based learning causing high intrinsic and extraneous cognitive load and a high error rate (2) within a heterogeneous group of students (3) on self-generation success and long-term retention and also the impact of feedback is still lacking.

### 1.3. Research Questions

Our study was conducted to analyze the impact of the generation effect in inquiry-based learning and discover learning conditions that enhance long-term retention of conceptual scientific issues by examining five main questions:

- Q1 Does self-generating content knowledge during inquiry affect long-term retention among 6th and 7th graders?
- Q2 Does additional feedback improve the long-term retention of generated information?
- Q3 Does successful self-generation enhance the effect?
- Q4 Does additional feedback (including a prompt to revise the generated answer) enhance successful self-generation during learning?
- Q5 Do individual learning abilities and characteristics (e.g., need for cognition, cognitive load, reading competency, self-generation success) influence the learning outcome?

## 2. Method

### 2.1. Participants

We conducted an a priori power analysis using G \* Power [43] with a significance level of  $\alpha = 0.05$ , a medium effect size of  $f = 0.3$  (which is a conservative estimate; cf. *Sentence completion* ( $d = 0.6$ ); [9]), and a desired power of 0.8; the results indicated a recommended sample size of  $N = 111$ . Consequently, a total of 136 6th and 7th graders from four different secondary schools with three different school types (Gymnasium, Realschule, Hauptschule) were originally recruited (see Appendix B for details of the German school system). In the end, a sample of 98 students (power of 0.75), with a mean age of 13 years ( $SD = 0.71$ ), 49% boys, participated in our experiment. A dropout of 34 students can be explained by the presence of different participants in at least one of the three total sessions or a lack to consent to data usage. Additionally, four participants in the rereading condition had to be excluded from further analysis due to extreme values.

### 2.2. Research Design

A  $3 \times 2$  mixed factorial design was used with an encoding task for biological content knowledge (self-generation (G) vs. self-generation with feedback (GF) vs. rereading (RR)) as a between-subjects factor and retention interval (10 min vs. 1 week) as a within-subjects factor. Three experimental groups

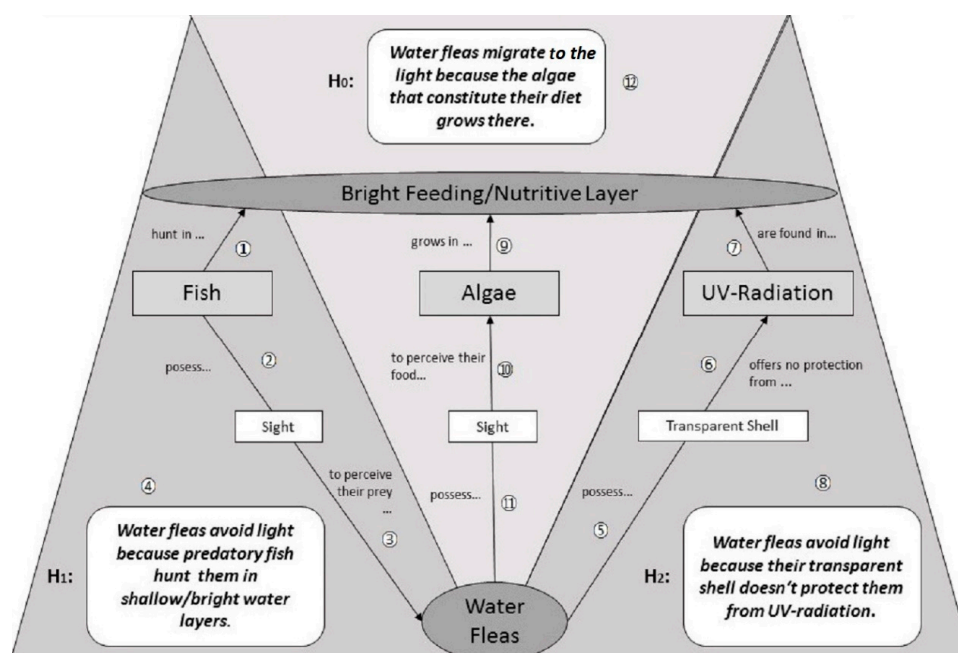
were formed on the basis of these different types of encoding tasks to which the students were randomly assigned. The randomization controls revealed no significant differences with respect to biology grade, German grade, sex, and need for cognition, although there was a difference in the students' age. Students in the generation group were coincidentally significantly younger ( $M = 12.6$  yrs.,  $SD = 0.56$ ) than the students in both other groups ( $M = 13.03$  yrs.,  $SD = 0.77$ ). However, this difference had no statistically significant influence (T1:  $F(3, 101) = 0.29$ ,  $p = 0.835$ ; T2:  $F(3, 101) = 0.61$ ,  $p = 0.609$ ).

As the second independent variable was manipulated within subjects, all students were tested 10 min after manipulation (T1) and one week later (T2) with the same content knowledge questionnaires.

### 2.3. Learning Content

The learning content focused on the concept of biological adaptation, which is a core disciplinary idea in German and international science curricula. The periodic daily vertical migration of water fleas (*Daphnia magna*) served as an example of biological adaptation in an inquiry task. Water fleas move to the surface of a pond when the sun rises. As the intensity of light increases, the animals descend to a characteristic maximum depth. They then rise to the surface again as the light decreases and finally show a typical midnight sinking [44]. It can be reasonably assumed that all students possessed the same low prior knowledge of the learning content at the beginning of the study, as it had not been covered previously by any of the participating classes. Thus, element interactivity or *intrinsic cognitive load* was at a high level for all students (cf. [19]).

To avoid overexertion, certain isolated ideas/elements (①, ②, ③, ⑤, ⑥, ⑦, ⑨, ⑩, ⑪) presented in a concept map given below (see Figure 1) were provided via an eLearning program. However, they remained disconnected for the students until the manipulation. Providing information about the anatomical structure of water fleas, their habitat, their prey, and predators allowed students to understand and/or generate three different hypotheses ( $H_0$ ,  $H_1$ ,  $H_2$ ) about the daily vertical migration of these animals in the pond during inquiry. All three hypotheses and interpretations of the phenomena (④, ⑧, ⑫ in Figure 1) could be formulated on the basis of the presented isolated ideas ①, ②, ③, ⑤, ⑥, ⑦, ⑨, ⑩, and ⑪ from the introductory session. Each hypothesis was assigned to four single items, so that three items could be combined to generate a sense of coherence in a fourth item (e.g., ①, ②, ③, results in ④).



**Figure 1.** Concept map of the learning content. Presenting the complex coherences among the three hypotheses for the vertical migration of water fleas (*Daphnia magna*).

### 3. Materials

#### 3.1. Introductory Session

In the first session, all students were introduced to the concept of biological adaptation of animals of the pond as well as the habitat and anatomical structure of an exemplary animal, the water flea, via a short, standardized computer-based program. Thus, the students were familiarized with the basic information (Figure 1) they needed for the subsequent inquiry task with the help of videos, pictures, and short text passages (see Supplementary Materials).

#### 3.2. Inquiry Task

The second session took place in an inquiry-based learning environment conducted in a student lab (the Experimental Biology Lab at the University) one week later. The students' learning process was supported by a research workbook that led all students through the steps of a scientific study (generating hypotheses, carrying out an experiment, and drawing conclusions). Students of the generation groups were asked to document their generated hypotheses and interpretations of their results in there.

Using several different instructional formats—individual, pair, and group work—created a natural setting for us to implement the repetition of the generation process.

The workbook format for both self-generation conditions comprised 5 short self-generation prompts (short answer tasks) concerning the students' hypotheses and interpretations of their data (see Supplementary Materials) and a cloze concerning information about all three hypotheses and interpretations at the end of the research workbook (see Supplementary Materials). The research workbooks for the RR-condition contained reading texts instead of generation prompts and a cloze. They were graded with regard to their readability and showed a Flesch Reading Ease [45] of 61 (simple), which is appropriate for students in Grades 6 and 7. In order to ensure that information in all encoding formats was similar and processed to the same extent, feedback material for the self-generating treatment (GF) were developed to reflect the text material used in the reading condition.

### 4. Procedure

#### 4.1. Introductory Session

In the first session, all students received the same standardized introduction to the basic content they needed for the second learning session. In order to consistently transmit the basic content knowledge and minimize teaching style effects, all students worked independently for 30 min on the eLearning program after receiving a short briefing by a trained supervisor. To ensure that all material was read carefully without skipping any slides, the program contained control mechanisms (e.g., time controls).

#### 4.2. Inquiry Task

In the inquiry session, students within a class were randomly assigned to three conditions (self-generation (G) vs. self-generation with feedback (GF) vs. rereading (RR)) and then divided into small groups of up to five students. All groups were guided by trained supervisors during a three-hour inquiry task. The supervisors moderated the self-controlled learning process by providing instructional support: They were specially trained in mentoring all three conditions and received comprehensive scripts with detailed information on each phase in each condition (see Supplementary Materials). Both self-generation groups were tasked with actively generating their own hypotheses and appropriately interpreting the data they collected in individual work, partner work, and team-based group work using the information acquired in the introductory unit. By contrast, students in the RR-condition were provided with the same information that is hypotheses and interpretations that simply had to be read.

While the G-condition received no feedback, the GF-condition group were provided with immediate corrective feedback after both individual work phases during the learning process (see Supplementary Materials). We decided to restrict feedback to just two time points and prompt the students to revise their answers, as too much feedback has proven to detract from performance. The feedback comprised reading texts from the reading condition. Thus, the students received the same information as students in the RR-condition. In the end, the GF-condition group were provided with knowledge of the correct responses for all three hypotheses and were instructed to supplement and/or revise own hypotheses and interpretations. In this way, the students were given the opportunity to reject faulty hypotheses/interpretations and use the provided cues as tips for generating correct answers. Apart from this, manipulation in the phases required declarative knowledge, *generating hypotheses*, and *analyzing/interpreting data*; the study procedure was identical in all three conditions.

Immediately following the inquiry-based learning session, students in all treatments completed a questionnaire regarding cognitive load and applied content knowledge. Additionally, all students were tested one week later using the same questionnaire on content knowledge.

## 5. Instruments

Three assessment time points were integrated into the experimental design: the first, after the introduction to the “Animals of the Pond” via an eLearning program, the second after the inquiry task, and a final test after a retention interval of one week. In addition, the students’ success in self-generation and perceived cognitive load [46] was measured during the experimental unit. Moreover, data were collected on the students’ demographics, their grades in biology and German, and their need for cognition [47]. All four questionnaires were completed within three weeks. Total data acquisition was completed within four months.

### 5.1. Learning Outcome

Learning condition was chosen as the between-subjects factor, while retention interval served as the within-subject factor. Therefore, a questionnaire was designed to assess the acquisition and retention of content knowledge. All students were tested immediately after the inquiry task (T1) and one week later (T2) using the same questionnaire. It consisted of 12 single-choice items offering four possible answer options. The 12 items corresponded to those described in the concept map (see Figure 1) in order to examine whether the disparate ideas could be retrieved and whether a sense of coherence among them was developed after the inquiry task.

For this paper-and-pencil test, item difficulty, internal consistency, and discrimination parameters were calculated using classical test theory. Item difficulty was satisfactory ( $p = 0.64$ ), and the test was sufficiently reliable ( $\alpha = 0.66$ ) for comparing groups [48]. Furthermore, the discrimination parameters were all above  $r_{it} 0.30$ .

### 5.2. Learners’ Self-Generation Success

To assess successful self-generation, a qualitative analysis of the hypotheses and interpretations generated by the students was conducted in order to determine how much information (hypotheses, interpretations) was generated how often. In total, a maximum of 29 credit points plus 9 extra points for an additional/third hypothesis (for details Appendix C) could be earned. In addition, we determined whether feedback was incorporated into subsequent generation processes. For the *self-generation of hypotheses* criteria like correctness, completeness, and quality were evaluated in all three working phases (individual, pair, and group work). It was analyzed whether at least two of three possible hypotheses were correctly generated and justified with at least one or two arguments the students learned in the eLearning program (max. 18). The same criteria were used for the assessment of the *interpretation of the data* in an individual working phase (max. 8 credit points): The raters checked whether the interpretation was suitable to the results and for correctness. The conclusion of the experimental results could involve 1–6 reasons from the eLearning program that were put

into context. The subsequent cloze received a score on a scale of 0 to 12 credits, 4 gaps for each hypothesis/interpretation (see Supplementary Materials for further information on self-generation prompts and supervisor instructions).

Consistency among two independent raters was determined by means of an interrater reliability analysis using the Kappa statistic. Interrater reliability was found to be Cohen's  $\kappa = 0.82$  ( $p < 0.001$ ), which reflects almost perfect agreement [49].

### 5.3. Learners' Abilities

Learners' need for cognition [47] and cognitive load [46] were measured via standardized questionnaires. The instrument for need for cognition comprised 19 items measured on a five-point Likert scale ( $\alpha = 0.85$ ). The questionnaire for cognitive load consisted of 8 items to which responses were recorded on a seven-point Likert scale ( $\alpha = 0.76$ ).

### 5.4. Data Analysis

To identify differences between experimental conditions and among learners of different ability levels, statistical analyses rooted in classical test theory were conducted using the SPSS software. All results were significant at the 0.05 level unless otherwise stated. Pairwise comparisons were Bonferroni-corrected to the 0.05 level. Partial eta squared ( $\eta_p^2$ ) are reported for all ANOVAs, Cohen's  $d$  for all t-tests, and  $r$  for Mann–Whitney U-tests as measures of effect size.

## 6. Results

The descriptive results of the learners' performance during the inquiry session and in all test sessions are shown in Table 2.

**Table 2.** Means and standard deviations (in parentheses) of performance assessed during the inquiry session and in posttest 1 and 2.

	Generation with Feedback	Generation	Rereading
Posttest 1 (%)	73.33 (20.33)	62.99 (17.75)	81.92 (8.75)
Posttest 2 (%)	70.00 (20.42)	56.37 (21.33)	77.42 (17.28)
Self-generation success	12.97 (4.37)	8.62 (3.86)	-
<i>N</i>	34	40 (35 <sup>a</sup> )	24

<sup>a</sup> Self-generation success could only be analyzed in 35 of 40 research workbooks.

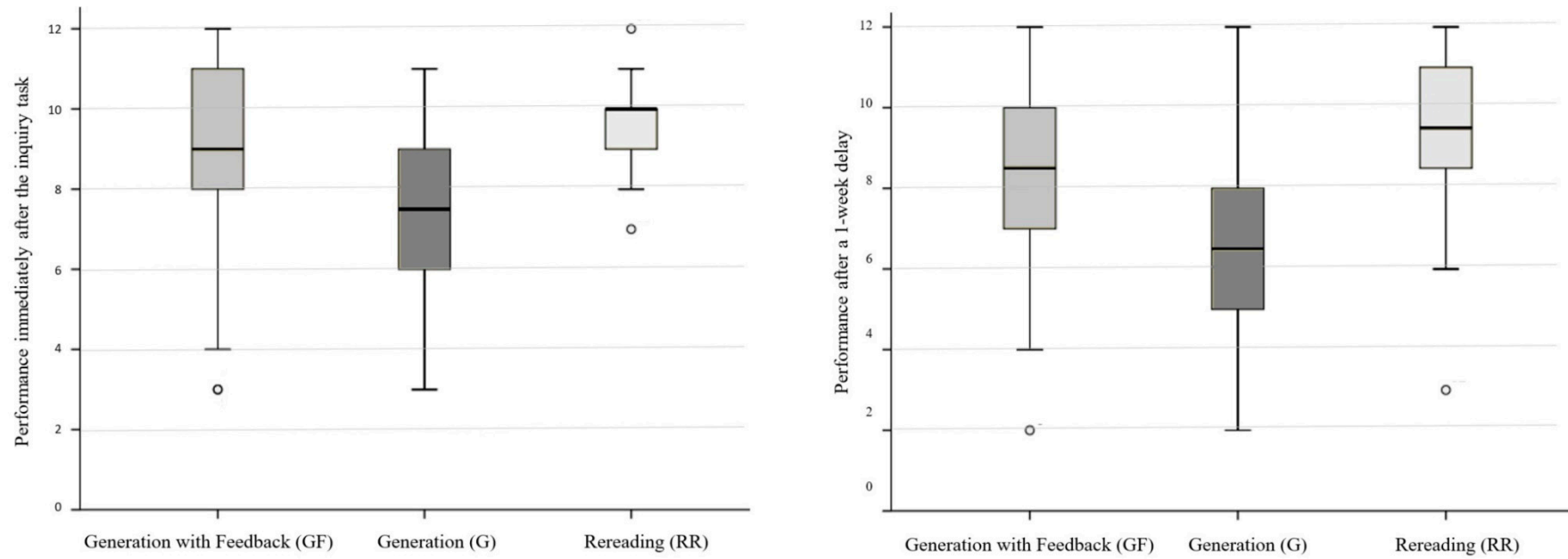
### 6.1. Retention by Treatment and Time (Q1, Q2)

A multi-factorial analysis of variance with repeated measures was used to compare the different treatments (G vs. GF vs. RR) at both measurement points.

Findings revealed a main effect for the condition  $F(2, 95) = 11.62$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.197$  and time  $F(1,95) = 7.17$ ,  $p = 0.009$ ,  $\eta_p^2 = 0.070$ .

No significant interaction between retention interval and condition was found. Performance on the short-term (T1) and long-term (T2) retention tests was significantly higher in the RR-condition and the GF-condition than in the pure G-condition (see Figure 2). Pairwise comparisons (Bonferroni-corrected) demonstrated that receiving feedback led to significantly higher short-term and long-term retention among students who experienced self-generation requirements (T1: GF vs. G:  $MD = 1.24$ ,  $SE = 0.48$ , 95% CI [0.63, 2.42],  $p = 0.035$ ; T2:  $MD = 1.64$ ,  $SE = 0.55$ , 95% CI [0.33, 2.94],  $p = 0.009$ ). Rereading outperformed self-generation without feedback at both measurement points (T1: L vs. G:  $MD = 2.27$ ,  $SE = 0.55$ , 95% CI [0.93, 3.62],  $p < 0.001$ ; T2: L vs. G:  $MD = 2.53$ ,  $SE = 0.61$ , 95% CI [1.04, 4.02],  $p < 0.001$ ). In contrast to G-condition, RR-condition ( $n = 24$ ) and GF-condition ( $n = 40$ ) were not normally distributed at both measurement points, as assessed by the Shapiro–Wilk test (T1:  $p = 0.014$ , T2:  $p = 0.015$  for RR; T1:  $p = 0.002$ , T2:  $p = 0.027$  for GF). Mann–Whitney U-Tests revealed no significant differences between the RR- and GF-encoding formats at both measurement points (T1:  $U(40, 24) = 1,49$ ,  $p = 0.136$ ; T2:  $U(40, 24) = 1,43$ ,  $p = 0.153$ ).





**Figure 2.** Correct responses by treatment (self-generation with feedback vs. self-generation vs. rereading) and time (short-term and long-term).

A significant decline in retention between T1 and T2 was only found in the pure G-condition without feedback, G:  $t(33) = 2.08, p = 0.045, d = 0.4$ .

### 6.2. Success in Self-Generation (Performance Success) (Q3, Q4)

To confirm the treatment effects and examine the influence of self-generation success on short-term and long-term retention, we further collected qualitative data in the form of 69 student responses to the self-generation prompts in the inquiry-based environment, including their hypotheses, interpretations, and the final cloze in their research workbooks (see *Learners' Self-Generation Success*). The analyses revealed that at the end of the learning session students in the GF-condition achieved significantly higher success in self-generation than students in the pure G-condition,  $t(67) = 4.35, p < 0.001, d = 1,1$  (see Figure 2).

We have indicated how many students were able to solve fewer than 50% and more than 50% of the generation tasks at the end of the inquiry unit. It turned out that 27 students (19 from the GF-condition, 8 from the G-condition) were able to generate more than 50% of the items correctly at the end of the learning unit; 32 students (16 from the GF-condition, 16 from the G-condition) generated fewer than 50% correctly. One GF-condition of 5 students lacked the corresponding documentation. Performance improved in both groups in Posttest 1, but fell again in Posttest 2 for the G-condition.

### 6.3. Learners' Abilities and Prerequisites (Q5)

We conducted a manifest path model (see Appendix A for details on the data fit values) to analyze the complex interactions between variables referring to the learning process, the immediate and delayed tests, and learners' characteristics (see Table 3) [50]. Multiple variables were considered in the path model, which specified several dependent variables as well as dependent and independent variables at the same time [50]. We also controlled for indirect correlations between variables (i.e., mediation effects) in the path model [51].

Success in self-generation was measured in credit points (0–29), cognitive load on a scale of 1 (very low) to 7 (very high) and need for cognition on a scale of 1 (very low) to 5 (very high). Reading competency was measured by the students' German grade.

Path analysis was used to test if individual learning abilities and prerequisites significantly predicted learners' self-generation success and retention. The results presented in Table 3 revealed that in the GF-condition, short-term retention was influenced by learners' perceived cognitive load, success in self-generation, and need for cognition. Moreover, need for cognition had the highest influence on self-generation success.

**Table 3.** Results of the Path Analyses by Treatment (self-generation with feedback and rereading) and Time (short-term (T1) and long-term (T2)).

Variable	Short-Term (T1)		Long-Term (T2)		Self-Generation Success (GS)
	GF	RR	GF	RR	GF
GS	0.52 ***	–	–	–	–
ind. (T1)			0.34 ***		
NFC					0.46 ***
ind. (GS)	0.24 *		0.19 *		
CL	–0.37 ***				
ind. (T1)			–0.27 **		
RC		0.45 *		0.44 *	
T1	–		0.54 ***		–
R <sup>2</sup>	0.47 ***	0.20 *	0.43 ***	0.20 *	–

*Note.* T1 = immediate test; T2 = 1-week-delayed test; GS = self-generation success; NFC = need for cognition; CL = cognitive load; RC = reading competency; ind. = indirect path; GF: the model for the treatment condition generation with feedback; RR: the model for the treatment condition rereading. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

All three variables were still found to have an indirect effect on long-term retention. However, the highest correlation was found between test scores on the two posttests.

In the RR-condition, reading competency predicted performance for both posttests. 20% of the variance in short-term and long-term retention was explained solely by this factor. Unexpectedly, performance on the immediate test turned out to be no predictor for the second posttest in RR.

## 7. Discussion

The present study sought to analyze the unique role of self-generating content knowledge in inquiry-based learning as well as the impact of feedback and other components on the effectiveness of this encoding format. We compared an inquiry activity involving the self-generation of content knowledge with and without subsequent feedback to an inquiry task in which students simply reread all possible hypotheses and appropriate interpretations of their experimental results.

### *7.1. Self-Generating with Feedback and Rereading Both Promote Long-Term Retention in Inquiry-Based Learning (Q1)*

We assumed that generating the answer rather than simply reading it would be an advantage in terms of memory and a disadvantage regarding immediate performance [6]. Contrary to expectations and recent findings [22], we found no negative generation effect between self-generation with feedback and simple rereading immediately after the inquiry task. There was also no generation advantage after a one week delay. Thus, the findings of McDaniel et al. [25], Slamecka and Graf [6] and others could not be confirmed. In line with expectations, pure self-generation without feedback led to the worst performance and learning outcomes among all three conditions. Moreover, retention significantly declined after pure self-generation, while it remained stable over one week when a feedback was given or hypotheses and interpretations were simply read [39].

### *7.2. Success in Self-Generating is Critical for Long-Term Retention (Q3)*

In contrast to the findings by Richland et al. [29], success during learning reliably predicted learning outcomes. Retention depended on the amount of successfully self-generated information and the recurrence of the self-generation process. These results support the recent findings of the argumentation of Kaiser et al. [22] and Foos et al. [28] that there is only a generation effect for successfully generated items. The fact that self-generation success was moderated by learners' high NFC, underlines Cazan and Indreica's finding that NFC correlates with deeper learning strategies like self-generation [41]. At the same time, feedback during task acquisition was a fundamental requirement, as it increased the probability of successful learning through self-generation and improved performance. This was particularly the case because self-generation referred to erroneous interpretations [34].

### *7.3. Feedback Is a Key Prerequisite for Self-Generating Complex Content Knowledge (Q2, Q4)*

Corrective feedback did not only improve success in self-generation (i.e., performance). Furthermore, the fact crystallized that additional feedback on generated answers had a crucial influence on memory. Feedback increased learning outcomes at both measurement points and promoted long-term retention by reducing forgetting. This was mainly due to reduced cognitive load. While pure self-generation required the construction of relations between all of the individual elements (= items), in the feedback condition, connections that were generated incorrectly or not at all could be supplemented by feedback [19]. Students with disconnected ideas on hypotheses and interpretations tended to separate new ideas and were more likely to forget the information they had learned; conversely, students with integrated understandings of the content were able to connect new information through the knowledge integration process. Enhancing knowledge integration is a crucial step towards providing a solid basis for students' lifetime learning [30]. Indeed, while it could be argued that feedback alone led to this result, feedback can only build on self-constructed basic knowledge; there is no feedback effect when initial active learning, i.e., self-generation, is missing [34].

#### 7.4. Generating and Rereading Place Different Demands on Learners (Q5)

While the encoding formats of self-generation with feedback and rereading were similarly effective, they placed different demands on learners. An analysis of individual learners' abilities revealed that reading competency had the highest influence on short-term and long-term retention in the RR-condition. In the GF-condition, short- and long-term retention were moderated by learners' success in self-generation, cognitive load, and NFC. Students with low cognitive load, high self-generation success, and NFC profit the most from self-generation of content knowledge. It was particularly interesting that short-term performance had a substantial impact on long-term learning outcomes when the learning content was generated but no significant influence when the information was simply read. Thus, in contrast to Richland et al. [29], performance turned out to be an unreliable predictor of long-term learning for structured inquiry, but a robust predictor for guided inquiry.

#### 7.5. Limitations and Directions for Future Research

The lack of a positive generation effect in the long run—that we expected to find between GF and RR according to previous laboratory results—can be explained by the unique features of the inquiry-based learning environment, the complexity of the learning content, and learners' characteristics (see *Introduction*). By using inquiry-based learning, which is a collaborative form of learning, it could not be guaranteed that all students would successfully generate all information completely independently and unaffected. Communication among students with different learning characteristics (e.g., cognitive abilities, prior knowledge) through collaborative learning, which is common in everyday instructional practice, led to interaction and the exchange of information among students. However, in the end, own performance and therefore a high degree of activity is precisely a distinctive prerequisite for the generation effect. As long as information could not be generated due to missing preexisting schemata, it could not be remembered [27]. Additionally, even though we tried to implement this requirement by using individual work phases, feedback, and a computer-based introduction into the material, there were still students who failed to generate their own hypotheses and interpretations.

Further decisions regarding the learning environment may have also overshadowed differences in learning between GF and RR. Based on the assumption that the generation effect may only occur for stimuli with preexisting representations [23,24], a short, standardized introduction to the basic content knowledge was conducted in order to create preexisting representations and reduce learners' cognitive load. However, since increased expertise reduces element interactivity and promotes semantic processing [2,19,23,24,27], students of the rereading condition might also have felt motivated to generate relations between the single information they learned in that eLearning program. Thus, our methodological decisions may have been enough for students of the rereading condition to generate hypotheses or interpretations even when they were not explicitly prompted to do so. Therefore, the cognitive process could not adequately be controlled. Even a seemingly passive task on the instructional level can induce the retrieval of prior knowledge and active generation of information on the cognitive level, and this is particularly the case when the students already find themselves in an activity-based and problem-based learning environment [13]. Renkl [52] argues that a perfectly pure form of inquiry-based learning in the sense of a minimum or maximum generation requirement does not exist. Even reading, which is a receptive form of learning, demands that connections be generated on the basis of the text content and the learner's prior knowledge for understanding to be achieved [52]. The coherent reading and generating material in the present study inevitably led to conceptual processing in both treatments, which brings about deeper understanding and long-term retention to both conditions according to Jacoby [53], whereas typical laboratory studies induce primarily perceptual processing with reading material and conceptual processing with generating material [54]. Nevertheless, we decided to provide all students with the same introductory information in order to ensure comparability of generated and read information across all treatments. We also made a deliberate choice in favor of a hands-on activity in order to be able to examine an authentic inquiry environment that is commonly used in science classes [1]. However, the open learning

environment in the present study risked creating obstacles not only for the investigations but also for the analyses of the results. The analysis of only successfully generated items—as conducted by Foos and colleagues—could not be implemented [28], because these items could only have been isolated for inspection by restricting the *hypothesis space* [10] and concretely specifying all hypotheses that had to be generated. However, according to Hmelo-Silver et al. [1], such practices are contrary to the definition of inquiry-based learning. Unfortunately, our relatively small sample, thus a low statistical power, has prevented an alternative analysis: the comparison of highly successful students in the GF-condition with a group of similar students in the RR-condition (on the basis of grades). Thus, further replications with greater power are required. Instead, we calculated the influence of self-generation success on retention using path analysis and compared self-generation success between the pure G- and the GF-treatment. Our investigations suggest that self-generation with feedback might bring a similar benefit for content knowledge (declarative knowledge) as for inquiry skills (procedural knowledge) when only students who successfully generated or only material that was actually generated are considered [22,28].

## 8. Implications

Regarding theoretical implications, this study can expand the empirical research base on the generation effect and the open pedagogical question of whether discovery-oriented or direct instructional methods result in better learning outcomes and retention [4,55]. It could be demonstrated that the long-term effectiveness of self-generation in an authentic setting like inquiry learning depends on certain learners' characteristics like NFC and interrelated contextual factors like students' self-generation success, cognitive load, and feedback. Feedback is the determining factor and initiator when it comes to effectively integrating self-generation as an encoding format in an authentic educational setting like inquiry. Self-generation within inquiry-based learning results in better long-term learning outcomes in guided inquiry when intrinsic and extraneous load is reduced by effective feedback. Promoting the process of self-generation through feedback, in turn, results in higher self-generation success, and thus retention. Ultimately, it is self-generation success that is critical for better long-term retention. Only when self-generation is successful can new information properly be linked with previous knowledge and stored in long-term memory.

In the end, this work suggests that when it comes to teaching content knowledge within inquiry-based learning, open inquiry proves to be ineffective for students aged 11–14. However, even guided inquiry does not automatically promote deep learning and retention. Self-generation success must be ensured. This can (partly) be achieved by feedback. Immediately correcting errors during task completion results in higher performance and improves final retention. However, for a higher rate of self-generation success, further scaffolds (e.g., incremental scaffolds) and a sufficient basis of preexisting schemata (e.g., video modeling examples [56]) might be required.

## 9. Conclusions

Giving students the opportunity to connect isolated notions about scientific phenomena and allowing them to make sense of complex phenomena are important parts of scientific education. However, guided and verification/structured inquiry can both provide such an opportunity. A high self-generation requirement did not turn out to be more effective and efficient than simply rereading content knowledge in an inquiry-based learning environment. Rather, the results indicate that the self-generation of hypotheses and appropriate interpretations delivers equivalent learning outcomes as rereading, when feedback is given. Open inquiry, thus self-generation without feedback, did not prove to be effective. Only through feedback can cognitive load be reduced, self-generation success be increased, and retention be improved. Thus, self-generation in the context of inquiry-based learning should always include feedback.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2227-7102/10/10/277/s1>.

**Author Contributions:** Conceptualization, I.S. and J.M.; Funding acquisition, J.M.; Investigation, I.S.; Methodology, I.S.; Project administration, J.M.; Writing—original draft, I.S.; Writing—review & editing, J.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project was funded by the LOEWE Excellence Programme: Desirable Difficulties in Learning from the Hessian Ministry for Science and the Arts.

**Conflicts of Interest:** The authors declare no conflict of interest

## Appendix A. Fit Values of the Path Models

The data fit values of the model of the GF-group are  $RMSEA_{GF} < 0.01$ ;  $p(RMSEA_{GF}) = 0.5$ ;  $CFI_{GF} = 1.000$ ;  $SRMR_{GF} = 0.04$ , which are acceptable in respect to the model fit and other parameters. In the case of the RR-group, the tests of model fit show a degree of freedom of “0”, which reveal a “saturation” or barely identification of the model, making it impossible to test the structure of covariance (Geiser, 2011) ( $RMSEA_{RR} < 0.001$ ;  $p(RMSEA_{RR}) < 0.001$ ;  $CFI_{RR} = 1.000$ ;  $SRMR_{RR} < 0.001$ ). For each dependent variable, the amount of explained variance ( $R^2$ ) is reported.

## Appendix B. German School System

The German school system is a three-tiered system in which students are divided into three different tracks depending on their cognitive abilities and educational goals: (1) *Gymnasium*, (2) *Realschule*, and (3) *Hauptschule*. Participants in the study were both students from schools with separate “Haupt-”, *Realschule*”, and “*Gymnasium*” tracks (cooperative comprehensive schools) and 6th grade students enrolled in the “*Förderstufe*” who had not yet been tracked into a school type. Teachers of the 6th grade “*Förderstufe*” were asked to provide a recommendation/assessment of the school type they thought each student should attend in 7th grade in order to improve the comparability of the school data.

## Appendix C. The Term “Hypothesis”

The focus of the study was on the acquisition of content-related knowledge. Therefore, in the context of this study, the students were explicitly prepared in advance only with respect to content-related information and not methodological knowledge. The technical term “hypothesis”, which was quite difficult for the group of young learners to remember, was avoided and described using the term “educated guess” (in the original: “*begründete Vermutung*”). Nevertheless, the students were informed by the supervisors and an additional accompanying booklet what a hypothesis consists of—a presumption and a content-related justification. Therefore, they were encouraged to formulate hypotheses, even if they were not explicitly designated as such.

## References

1. Hmelo-Silver, C.E.; Duncan, R.G.; Chinn, C.A. Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educ. Psychol.* **2007**, *42*, 99–107. [[CrossRef](#)]
2. Kirschner, P.A.; Sweller, J.; Clark, R.E. Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educ. Psychol.* **2006**, *41*, 75–86. [[CrossRef](#)]
3. Paas, F.; Renkl, A.; Sweller, J. Cognitive load theory and instructional design: Recent developments. *Educ. Psychol.* **2003**, *38*, 1–4. [[CrossRef](#)]
4. Dean, D., Jr.; Kuhn, D. Direct instruction vs. discovery: The long view. *Sci. Ed.* **2007**, *91*, 384–397. [[CrossRef](#)]
5. Jacoby, L.L. On interpreting the effects of repetition: Solving a problem versus remembering a solution. *J. Verbal Learn. Verbal Behav.* **1978**, *17*, 649–667. [[CrossRef](#)]
6. Slamecka, N.J.; Graf, P. The generation effect: Delineation of a phenomenon. *J. Exp. Psychol. Hum. Learn. Mem.* **1978**, *4*, 592–604. [[CrossRef](#)]
7. Crutcher, R.J.; Healy, A.F. Cognitive operations and the generation effect. *J. Exp. Psychol. Learn. Mem. Cogn.* **1989**, *15*, 669–675. [[CrossRef](#)]

8. McNamara, D.S.; Healy, A.F. A Procedural Explanation of the Generation Effect: The Use of an Operand Retrieval Strategy for Multiplication and Addition Problems. *J. Mem. Lang.* **1995**, *34*, 399–416. [CrossRef]
9. Bertsch, S.; Pesta, B.J.; Wiscott, R.; McDaniel, M.A. The generation effect: A meta-analytic review. *Mem. Cogn.* **2007**, *35*, 201–210. [CrossRef]
10. Klahr, D.; Dunbar, K. Dual space search during scientific reasoning. *Cogn. Sci.* **1988**, *12*, 1–48. [CrossRef]
11. Klahr, D. *Exploring Science: The Cognition and Development of Discovery Processes*; MIT Press: Cambridge, MA, USA, 2000. Available online: <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=32641> (accessed on 28 September 2020).
12. Bell, R.L.; Smetana, L.; Binns, I. Simplifying inquiry instruction: Assessing the inquiry level of classroom activities. *Sci. Teach.* **2005**, *72*, 30–33.
13. Chi, M.T.H. Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Top. Cogn. Sci.* **2009**, *1*, 73–105. [CrossRef] [PubMed]
14. Abrams, E.; Silva, P.C.; Southerland, S.A. (Eds.) Contemporary research in education. In *Inquiry in the Classroom: Realities and Opportunities*; IAP: Charlotte, NC, USA, 2008.
15. Schwab, J. The teaching of science as enquiry. In *The Teaching of Science*; Schwab, J., Brandwein, P.F., Eds.; Harvard University Press: Cambridge, MA, USA, 1962; pp. 1–103.
16. Colburn, A. An inquiry primer. *Sci. Scope* **2000**, *23*, 42–44.
17. Zion, M.; Mendelovici, R. Moving from structured to open inquiry: Challenges and limits. *Sci. Educ. Int.* **2012**, *23*, 383–399.
18. Bjork, E.L.; Bjork, R. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*; Gernsbacher, M.A., Ed.; Worth: New York, NY, USA, 2011; pp. 59–68.
19. Chen, O.; Kalyuga, S.; Sweller, J. Relations between the worked example and generation effects on immediate and delayed tests. *Learn. Instr.* **2016**, *45*, 20–30. [CrossRef]
20. Sweller, J.; van Merriënboer, J.J.G.; Paas, F.G.W.C. Cognitive architecture and instructional design. *Educ. Psychol. Rev.* **1998**, *10*, 251–296. [CrossRef]
21. Sweller, J.; Chandler, P. Why Some Material Is Difficult to Learn. *Cogn. Instr.* **1994**, *12*, 185–233. [CrossRef]
22. Kaiser, I.; Mayer, J.; Malai, D. Self-generation in the context of inquiry-based learning. *Front. Psychol.* **2018**, *9*. [CrossRef]
23. Gardiner, J.M.; Hampton, J.A. Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* **1985**, *11*, 732. [CrossRef]
24. Nairne, J.S.; Pusean, C.; Widner, R.L. Representation in the mental lexicon: Implications for theories of the generation effect. *Mem. Cogn.* **1985**, *13*, 183–191. [CrossRef]
25. McDaniel, M.A.; Waddill, P.J.; Einstein, G.O. A contextual account of the generation effect: A three-factor theory. *J. Mem. Lang.* **1988**, *27*, 521–536. [CrossRef]
26. Lutz, J.; Briggs, A.; Cain, K. An examination of the value of the generation effect for learning new material. *J. Gen. Psychol.* **2003**, *130*, 171–188. [CrossRef] [PubMed]
27. McElroy, L.A.; Slamecka, N.J. Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *J. Verbal Learn. Verbal Behav.* **1982**, *21*, 249–259. [CrossRef]
28. Foos, P.W.; Mora, J.J.; Tkacz, S. Student study techniques and the generation effect. *J. Educ. Psychol.* **1994**, *86*, 567–576. [CrossRef]
29. Richland, L.E.; Bjork, R.A.; Finley, J.R.; Linn, M.C. Linking Cognitive Science to Education: Generation and Interleaving Effects. In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*; Bara, B.G., Barsalou, L.W., Bucciarelli, M., Eds.; Erlbaum: Mahwah, NJ, USA, 2005; pp. 1850–1855.
30. Clark, D.; Linn, M.C. Designing for Knowledge Integration: The Impact of Instructional Time. *J. Learn. Sci.* **2003**, *12*, 451–493. [CrossRef]
31. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory Explorations in the Learning Sciences, Instructional Systems and Performance Technologies*, 1st ed.; Springer: New York, NY, USA, 2011. Available online: <http://www.loc.gov/catdir/enhancements/fy1316/2011922376-d.html> (accessed on 28 September 2020).
32. Metcalfe, J.; Kornell, N. Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychon. Bull. Rev.* **2007**, *14*, 225–229. [CrossRef]

33. Kang, S.H.K.; McDermott, K.B.; Roediger, H.L. Test format and corrective feedback modify the effect of testing on long-term retention. *Eur. J. Cogn. Psychol.* **2007**, *19*, 528–558. [[CrossRef](#)]
34. Hattie, J.; Timperley, H. The Power of Feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [[CrossRef](#)]
35. Balzer, W.K.; Doherty, M.E.; O'Connor, R. Effects of cognitive feedback on performance. *Psychol. Bull.* **1989**, *106*, 410–433. [[CrossRef](#)]
36. Lysakowski, R.S.; Walberg, H.J. Instructional Effects of Cues, Participation, and Corrective Feedback: A Quantitative Synthesis. *Am. Educ. Res. J.* **1982**, *19*, 559–572. [[CrossRef](#)]
37. Harackiewicz, J.M.; Manderlink, G.; Sansone, C. Rewarding pinball wizardry: Effects of evaluation and cue value on intrinsic interest. *J. Personal. Soc. Psychol.* **1984**, *47*, 287–300. [[CrossRef](#)]
38. Kulik, J.A.; Kulik, C.L.C. Timing of Feedback and Verbal Learning. *Rev. Educ. Res.* **1988**, *58*, 79–97. [[CrossRef](#)]
39. Pashler, H.; Cepeda, N.J.; Wixted, J.T.; Rohrer, D. When does feedback facilitate learning of words? *J. Exp. Psychol. Learn. Mem. Cogn.* **2005**, *31*, 3–8. [[CrossRef](#)] [[PubMed](#)]
40. Cacioppo, J.T.; Petty, R.E.; Feinstein, J.A.; Jarvis, W.B.G. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychol. Bull.* **1996**, *119*, 197–253. [[CrossRef](#)]
41. Cazan, A.M.; Indreica, S.E. Need for cognition and approaches to learning among university students. *Procedia Soc. Behav. Sci. J.* **2014**, *127*, 134–138. [[CrossRef](#)]
42. Weißgerber, S.C.; Reinhard, M.-A.; Schindler, S. Learning the hard way: Need for cognition influences attitudes towards and self-reported use of desirable learning difficulties. *Educ. Psychol.* **2018**, *38*, 176–202. [[CrossRef](#)]
43. Faul, F.; Erdfelder, E.; Lang, A.G.; Buchner, A. G-Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **2007**, *39*, 175–191. [[CrossRef](#)]
44. Meier, M.; Wulff, C. *Daphnia magna* as the ultimate classroom organism: Implementing scientific investigations into school practice. In *Animal Science, Issues and Professions. Daphnia: Biology and Mathematics Perspectives*; El-Doma, M., Ed.; Nova Science Publ.: New York, NY, USA, 2014; pp. 225–244.
45. Flesch, R. A new readability yardstick. *J. Appl. Psychol.* **1948**, *32*, 221–233. [[CrossRef](#)]
46. Künsting, J. *Effekte von Zielqualität und Zielspezifität auf selbstreguliert-entdeckendes Lernen durch Experimentieren*. Ph.D. Thesis, Universität Duisburg-Essen, Essen, The Netherlands, 2007.
47. Preckel, F. Assessing Need for Cognition in Early Adolescence. *Eur. J. Psychol. Assess.* **2014**, *30*, 65–72. [[CrossRef](#)]
48. Lienert, G.; Raatz, U. *Testaufbau und Testanalyse*; Beltz: Weinheim, Germany, 1998.
49. Landis, J.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)]
50. Eid, M.; Gollwitzer, M.; Schmitt, M. *Statistik und –forschungsmethoden*; Beltz Verlagsgruppe: Weinheim, Germany, 2017.
51. Geiser, C. *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung*; VS Verlag für Sozialwissenschaften/Springer Fachmedien Wiesbaden GmbH Wiesbaden: Wiesbaden, Germany, 2011.
52. Renkl, A. Wissenserwerb. In *Springer-Lehrbuch. Pädagogische Psychologie*, 2nd ed.; Wild, E., Möller, J., Eds.; Springer: Berlin, Germany, 2015; pp. 3–24. [[CrossRef](#)]
53. Jacoby, L.L. Remembering the data: Analyzing interactive processes in reading. *J. Verbal Learn. Verbal Behav.* **1983**, *22*, 485–508. [[CrossRef](#)]
54. Rosner, Z.A.; Elman, J.A.; Shimamura, A.P. The generation effect: Activating broad neural circuits during memory encoding. *Cortex* **2013**, *49*, 1901–1909. [[CrossRef](#)] [[PubMed](#)]
55. Furtak, E.M.; Seidel, T.; Iverson, H.; Briggs, D.C. Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching. *Rev. Educ. Res.* **2012**, *82*, 300–329. [[CrossRef](#)]
56. Kant, J.M.; Scheiter, K.; Oschatz, K. How to sequence video modeling examples and inquiry tasks to foster scientific reasoning. *Learn. Instr.* **2017**, *52*, 46–58. [[CrossRef](#)]

