In: Trends in Neuroscience and Education Volume 19 (2020) , 100130 ; EISSN: 2211-9493; https://doi.org/10.1016/j.tine.2020.100130
Accepted Manuscript

Access to the learning material enhances learning by means of generating questions: Comparing open- and closed-book conditions

Mirjam Ebersbach

Abstract

Generating questions referring to the learning material is a powerful learning strategy. The present study investigated potential mechanisms behind this effect. Students (N = 231) read a text and then generated questions referring to the text in three conditions: (1) open-book (i.e., text accessible), (2) closed-book (i.e., text inaccessible), and (3) cued closed-book (i.e., text inaccessible but keywords provided). After one week, students' knowledge gain was larger in the open-book and cued closed-book conditions compared to a restudy control condition. No difference emerged between the closed-book and the restudy condition. The number of generated questions was positively correlated with students' knowledge gain, whereas question depth was not. Results suggest that the effect of question generation may not be attributed to retrieval practice or higher-order thinking. More probably, rephrasing and learners' active engagement with the learning material, as reflected by the number of generated questions, might contribute to the effect.

Key words: question generation; closed-book; open-book; long-term retention; retrieval practice

Access to the learning material enhances learning by means of generating questions: Comparing open- and closed-book conditions

Identifying effective learning strategies that promote comprehension as well as short- and long-term retention is a central aim in cognitive and educational psychology. Generating questions referring to the learning material by the learners themselves is one such strategy that yields robust effects (for a review, see Song, 2016). In previous studies, question generation was often trained first by introducing and practicing the strategy (e.g., King, 1992, 1994). Furthermore, they investigated the effect of generating questions on the comprehension of texts (e.g., Andre & Anderson, 1978; Nolan, 1991; King, 1990, 1991, 1992). Only more recently, positive effects were also reported on the recall of learning contents (e.g., Berry & Chew, 2008; Bugg & McDaniel, 2012; Byun, Lee, & Cerreto, 2012; Foos, 2016; Levin & Arnold, 2008). Authors (subm.), for instance, asked students at the end of a university lecture to generate questions and answers referring to the slides presented in the lecture that were still accessible to them. In a test after one week, students who generated questions clearly outperformed students who simply reread the material. The depth of the generated questions was not related to the learning outcome.

The positive effect of generating questions on recall can be explained in various ways. First, question generation may stimulate *transfer-appropriate processing* by evoking similar retrieval processes as the final test (e.g., Kolers & Roediger, 1984) because the studied content has first to be re-activated before a corresponding question can be formulated. Relatedly, question generation might promote *self-testing*, which, in turn, has beneficial effects on learning (i.e., *testing effect*; e.g., Roediger & Karpicke, 2006). Second, generating questions usually requires *rephrasing the original learning material*. Rephrasing or paraphrasing is in fact an effective tool to stimulate deeper processing, comprehension, and recall (e.g., Bui, Myerson, & Hale, 2013; Doctorow, Wittrock, & Marks, 1978; Hagaman, Casey, & Reid, 2012; Rosenshine & Meister, 1994; Wammes, Meade, & Fernandes, 2017;

Wittrock, 1974). Third, question generation might promote *higher-order thinking* (Bates, Galloway, Riise, & Homer, 2014; Papinczak et al., 2012). Although this concept is not well-defined (Song, 2016), it relates to the aspect of the *depth of the generated questions*, which might account for diverging effects of question generation: Generating deeper (i.e., more complex, conceptual) questions yielded benefits for conceptual knowledge acquisition, whereas generating more superficial, factual questions did not (Bugg & McDaniel, 2012; see also Levin & Arnold, 2008). Other potential explanations for the effect of question generation assume that self-testing in the context of question generation encourages learners to monitor their knowledge acquisition in terms of triggering *meta-cognitive processes* (King, 1989; van Blerkom & van Blerkom, 2004). Furthermore, generating questions is also related to the *generation effect* (e.g., McDaniel, Waddill, & Einstein, 1988), meaning that information is better recalled if it has to be, at least partially, generated by the learners themselves. Question generation can also be conceived as such a generating activity.

In the present study[1], two potential explanations for the effect of question generation were tested by introducing particular question generation conditions. First, if the assumptions of *transfer-appropriate processing* and *self-testing* hold – both referring to retrieval practice as the underlying mechanism – the benefit of question generation should disappear when no retrieval processes are required for generating questions. This condition can be realized by an “open-book” condition, originally realized in research on the testing effect (Adesope, Trevisan, & Sundararajan, 2017). The size of the testing effect strongly depends on retrieval success during testing (Rowland, 2014), which is larger when students have access to the learning material (i.e., “open-book”) than when they have no access (i.e., “closed-book”). Agarwal, Karpicke, Kang, Roediger, and McDermott (2008) reported an advantage of the open-book over the closed-book condition concerning the testing effect immediately after the

[1] The study is preregistered: https://aspredicted.org/blind1.php.

learning phase (i.e., a large effect of $d = 1.41$) and after one week (i.e., a medium to large effect of $d > .70$). Moreover, learners in the open-book condition performed similarly well as learners in a closed-book condition with feedback who could review the learning material after they had tried to retrieve the answers. All conditions were better than simple restudying. Thus, allowing learners to review the learning material while being tested or at least providing feedback after the retrieval attempt yields positive effects on memory compared to a pure retrieval phase without feedback (but see Rummer, Schweppe, & Schwede, in press). Roelle and Berthold (2017) showed that the advantage of open-book tests emerges in particular for complex learning material, as the possibility to review the learning material facilitates storing the correct content. Similar mechanisms might be assumed for generating questions as a learning strategy.

Second, if the assumption of *higher-order thinking* is true, the depth of the generated questions should be positively related to students' learning outcome, which was revealed in some studies involving question generation (e.g., Bugg & McDaniel, 2012; Levin & Arnold, 2004) but not in others (Authors, subm.; Berry & Chew, 2008).

The present study compares for the first time open- and closed-book conditions with regard to the generation of questions and thereby tests potential explanations for the question generation effect. If retrieval practice in terms of transfer-appropriate processing and self-testing does in fact account for the effect of question generation, the effect should disappear in an open-book condition, where learners can review the material for formulating questions. In contrast, one might also speculate that a pure closed-book condition is too difficult for question generators when their retrieval success is poor (Rowland, 2014). Thus, if the accessibility of learning material is important for learning success (see Agarwal et al., 2008, concerning the testing effect), then generating questions in an open-book condition should yield larger learning gains than generating questions in a closed-book condition. Authors (subm.) found in fact a significant advantage of generating questions in an open-book condition compared to restudying, but no closed-book condition was realized in this study.

One problem of having learners generate questions in closed-book conditions is that the questions they generate cannot be controlled: The generated questions might cover the whole learning material in its breadth or only a few aspects. However, cueing the generation of questions by providing learners with keywords their questions should refer to can solve this problem.

Thus, in the present study, the effects of question generation in an *open-book condition*, a *closed-book condition without cues*, a *closed-book condition with cues*, and a *restudy* condition were compared. As a central aim of learning is to retain the learned content across time, the learning outcome was tested one week after the learning phase. A positive effect of generating questions compared to restudying was expected, which should be – in line with previous results for the testing effect (Agarwal et al., 2008) – more pronounced in the open- than in the closed-book conditions. In addition, the depth of the generated questions as well as their number should be positively related to the learning outcome.

Method

Design

The study followed a one-factorial design with learning condition as between-subjects variable (i.e., generating questions in an open-book condition; generating questions in a closed-book condition with cues; generating questions in a closed-book condition without cues; restudying). Students' knowledge gain from pre- to posttest (after one week) served as dependent variable. In addition, the number of the self-generated questions and answers was considered as well as the depth of the self-generated questions.

Participants

Participants were students attending three university courses of developmental psychology. An a priori check of the data revealed that five participants in the question generation conditions did not generate any questions. They were therefore excluded from the analyses. The remaining sample consisted of 231 undergraduate students (168 females, 61 males, 2 did not report sex; mean age: M = 22.3 years, SD = 5.0)[2]. They had not been previously introduced to the topic of the learning content (i.e., dyscalculia) in the context of their studies. The students were randomly assigned within each course to the following conditions: open-book (n = 54), closed-book (n = 51), cued closed-book (n = 44), and restudy as control condition (n = 82). The unequal group sizes occurred due to an organizational error. Participation was voluntary with informed consent and could be terminated at any time. The learning content was announced to be relevant for the exam; in addition, participants could take part in a lottery.

Material

The learning material consisted of a text passage of 408 words, printed on one A4 page, comprising detailed facts on symptoms, diagnosis, potential causes, and consequences of as well as interventions for dyscalculia. Ten keywords in the text were printed in bold.

Procedure

About 30 min before the end of the lecture, students were informed that they may participate in a study taking place in the context of the lecture, involving learning material that would also be relevant for their exam. Participation would be anonymous and voluntary.

---

[2] An a priori sample size analysis by means of GPower (Faul, Erdfelder, Buchner, & Lang, 2009) suggested an optimal sample size of N = 179 to N = 279 for an ANCOVA, given a medium effect size and a power of .80 or .95, respectively. Similar sample sizes are required for a one-way ANOVA (i.e., N = 176 to N = 276).

If they decided to enroll, they could additionally take part in a lottery. Afterwards, the students who enrolled for the study (which were almost all) received ten questions referring to the text on dyscalculia that they had not yet read, to assess their prior knowledge (see Appendix). In addition, demographic data were assessed. Then, the text on dyscalculia and booklets were distributed. All students had 10 min to read the text. Then, they were randomly assigned via the instructions in their respective booklets to one of four learning conditions, that is either (a) rereading the text (i.e., restudying), or generating ten questions and answers referring to the text by (b) reviewing the text (i.e., open-book), (c) using provided cues without reviewing the text (i.e., cued closed-book), or (d) without cues and without reviewing the text (i.e., closed-book). As stated earlier, ten keywords were printed in bold in the text – these were the same keywords as those provided in the cued closed-book condition. This design made the cued closed-book condition comparable to the open-book condition, in which students had to generate questions and answers that referred to these keywords in the text (i.e., in order to assure that the same content was covered by the questions). To prevent students in the closed-book conditions from using the text, the text was collected from these students after the reading phase. The learning phase took about 20 min in all four conditions. Finally, all materials were collected and students were dismissed. One week later, an unannounced test took place at the beginning of the lecture. The test included the same ten questions as in the pretest addressing students' prior knowledge. The maximum score in this test was 38.5 points. The performance of about 20% of the sample was coded by a second rater, yielding interrater reliabilities of rs = .98 for both the pretest and main test. In addition, question depth was rated according to Levin and Arnold (2008) on four levels: (1) factual questions including definitions and isolated concepts, (2) questions that targeted the comprehension and distinction of related concepts, (3) questions that targeted the underlying knowledge structure including the analysis and synthesis of concepts, and (4) critical questions including implications (interrater reliability: $r = .99$)

Results

Originally, the data were to be analyzed by an ANCOVA with final test performance as dependent variable, learning condition as independent variable, and prior knowledge as covariate to test the effect of learning condition. However, preliminary analyses suggested that the homogeneity of regression coefficients as one requirement for conducting an ANCOVA was not met. In addition, a post-hoc power analysis revealed only a power of $1-\beta = .67$ (Faul et al., 2009). Thus, an ANCOVA was not suited to analyze the data.

Instead, a univariate ANOVA was calculated with knowledge gain as dependent variable (i.e., difference between test performance and prior knowledge) and learning condition as independent variable[3]. For this analysis, all requirements were met. Descriptive statistics are presented in Figure 1. The ANOVA revealed a main effect of learning condition, $F(3, 231) = 5.26$, $p = .002$, $\eta_p^2 = .07$. The post-hoc computed power of this analysis was $1-\beta = .94$ (Faul et al., 2009). Repeated comparisons (Bonferroni corrected) suggested that question generation in the open-book condition yielded a larger knowledge gain than restudy, $p =.002$, $d = 0.60$. The same was true for the cued closed-book condition compared to restudy, $p = .023$, $d = 0.56$. No other differences reached significance, *p*s < .44.

[3] Van Breukelen (2006) indicates that an ANOVA addressing the change from a baseline, as calculated here, is – like an ANCOVA with baseline performance as covariate – unbiased in randomized designs. In addition, Knapp and Schafer (2009) argue that an ANOVA is suited when one is merely interested in the performance gain from pre- to post-test, not in the particular role of pre-test performance on post-test performance, which also applies to the present study.

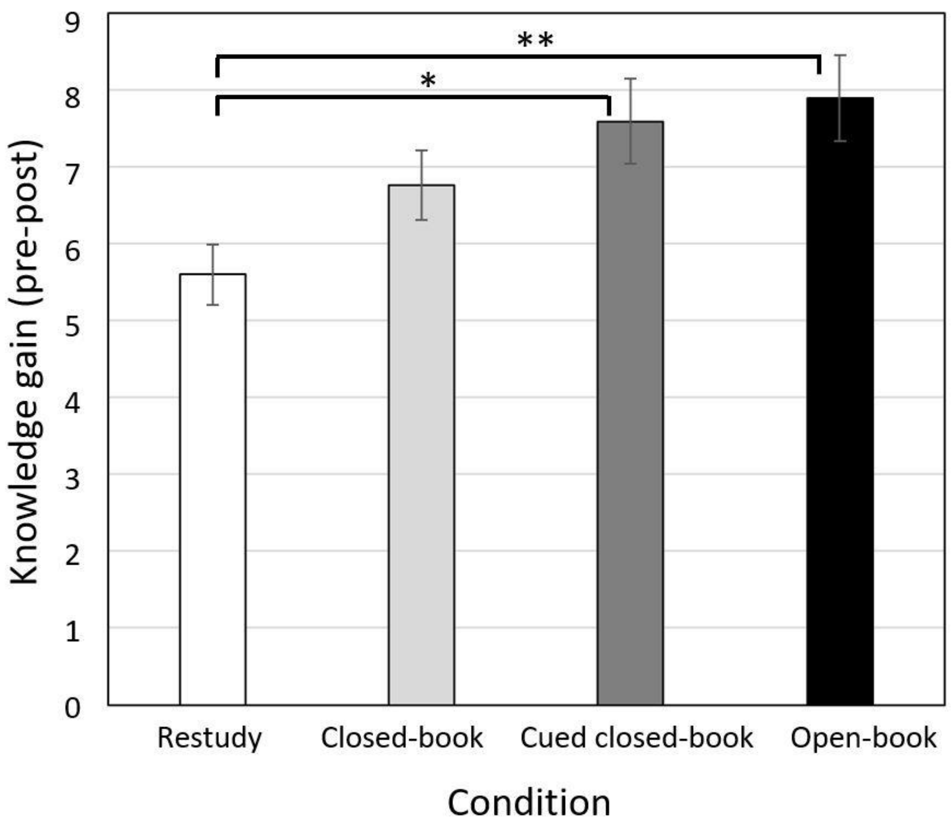

Figure 1.

Mean knowledge gain (i.e., difference between pre- and post-test) per learning condition.

Furthermore, it was explored whether the number of generated questions and answers and question depth were associated with knowledge gain (for descriptive statistics, see Table 1). Correlations were first computed across all three question generation conditions, yielding significant positive correlations between knowledge gain and the number of generated questions, $r(149) = .45$, as well as the number of generated answers, $r(149) = .48$, ps < .01, but not for question depth, $p = .28$ (Bonferroni corrected). Separate analyses for each question generation condition revealed similar patterns. Knowledge gain was significantly correlated in the closed-book condition with the number of generated questions, $r(51) = .61$, and the number of answers, $r(51) = .57$, ps < .01, but not with question depth, $p > .26$, and in the cued closed-book condition with the number of generated questions, $r(44) = .62$, and with the number of answers, $r(44) = .69$, ps < .01, but not with question depth, $p > .52$. In the open-book condition, however, no correlation reached significance, ps > .29 (Bonferroni corrected).

Table 1

Descriptive statistics of the assessed variables per condition

	Condition			
Measure	Restudy	Closed book	Cued closed book	Open book
Prior knowledge	2.4 (2.2)	2.0 (2.5)	1.5 (1.8)	1.2 (1.9)
Test performance	8.0 (3.7)	8.8 (3.6)	9.1 (3.1)	9.1 (2.6)
Number of generated questions	-	7.0 (2.0)	8.7 (1.8)	9.8 (0.8)
Number of generated answers	-	6.6 (2.4)	8.0 (2.5)	9.7 (0.8)
Mean question depth	-	1.1 (0.2)	1.1 (0.1)	1.1 (0.1)

Note. Standard deviations in parentheses. Maximum number of questions and answers: 10. Possible range of question depth: 1 to 4.

Discussion

In the present study, it was examined under which conditions the generation of questions yielded the largest effects on students' knowledge gain: in (a) an open-book condition, in which students could review the learning material, (b) a closed-book condition, in which students could not review the learning material but had to generate questions from memory, or (c) a cued closed-book condition, in which students could not review the learning material but were provided with keywords for their questions. Students' knowledge gain, tested one week after the learning session, benefited more from generating questions than from restudying the material, but only when students had access to the learning material (i.e., open-book condition) or when they were provided with keywords (i.e., cued closed-book condition). Generating questions in a pure closed-book condition yielded a similar knowledge gain as restudying. In addition, the number of generated questions and answers – but not question depth – correlated positively with students' knowledge gain.

Generating questions referring to the learning material is thus an effective learning strategy and is more beneficial than simple restudying. By manipulating the conditions, two potential explanations for the effect of question generation were tested (see Introduction). Retrieval practice accounts would predict that the effect would only be present in a closed-book condition but not in an open-book condition that does not involve retrieval. However, there was the other way around. Thus, retrieval practice alone cannot explain the effect of question generation.

Only students in the cued closed-book condition and in the open-book condition outperformed students who restudied the text. This leads to the assumption that rephrasing, which takes place even when a statement contained in the learning material is merely reworded as a question, might be one explanation for the effect of question generation. Rephrasing is facilitated when students can review the original material, such as in the open-book condition, or are at least provided with cues, such as in the cued closed-book condition. In contrast, recall and, as a consequence, rephrasing might be limited in the closed-book condition. This assumption is corroborated by the fact that only 54% of the questions generated in the closed-book condition referred to the target cues in the text, which were not provided in this condition. Thus, the covered contents might be limited in this condition. In contrast, the advantage of generating questions in an open-book condition over restudying the material replicates previous findings (Authors, subm.).

Another account for explaining the effect of question generation is that it promotes higher-order thinking (Papinczak et al., 2012), which might be reflected in the depth of the generated questions. However, question depth and knowledge gain were not related (for similar results, see Authors, subm.; Berry & Chew, 2008), which is probably due to the fact that most questions addressed simple facts and did not require deeper understanding (see

Table 1). Potentially, there was a floor effect of question depth, which might be overcome when learners are trained in advance to formulate deeper, comprehension-based or even critical questions or questions that address implications of the learned facts. They might profit even more from question generation (e.g., Rosenshine, Meister, & Chapman, 1996).

Nevertheless, the number of generated questions and answers correlated positively with students' knowledge gain. This relationship emerged in the two closed-book conditions (i.e., with and without cues), but not in the open-book condition – probably because almost all students generated all ten required questions and corresponding answers in this condition. This finding suggests that the extent of the involvement with the learning material and of the mental effort is central for learning.

The study also has some limitations. First, it refers to learning from texts but not to learning in the context of lectures (even if the study took place in a lecture). It might be worthwhile to replicate this study with actual lecture material (e.g., slides). However, as learning from texts is a common activity of students in particular in the context of self-regulated learning, the results are still instructive. Second, group sizes differed between conditions and were too small to directly test the moderating effect of the number of generated questions on students' knowledge gain. It could only be shown that the number of generated questions was positively associated with the knowledge gain in the two closed-book conditions, but it would be interesting to analyze how this number moderates the effects in each learning condition. Third, the overall performance was rather poor, which might be due to the fact that the exam was still about two months away and that the final test was unannounced to rule out that students would prepare for the test.

In order to test alternative accounts for the effect of question generation, other designs are required. For instance, to test whether the extent of generation activity (cf. McDaniel et al., 1988) in generating questions contributes to its effect, the to be generated questions might be prompted to different degrees (e.g., from providing one keyword to providing large parts of the questions, which only need to be completed by a single word). The question generation effect should vary depending on the extent of generation activity. In the present study, there was no advantage for the pure closed-book condition but an advantage for the cued closed-book condition over restudying – which might contradict the generation account. However, a graded provision with cues might yield a more detailed picture. In addition, one could manipulate the complexity of the learning material in order to find out whether an open-book condition in question generation outperforms a closed-book condition only with complex (or coherent) learning material (cf. Roelle & Berthold, 2017) or also with simple, incoherent learning material.

It can be concluded that fostering students' active involvement with the learning material by means of question generation enhances their long-term knowledge gains compared to restudying. However, during question generation they should have access to the learning material or should at least be provided with keywords referring to the content. Generating questions does not require prior training and is also a technique which students might incorporate into their self-regulated learning activities (e.g., while preparing for exams). In addition, teachers might encourage students in the context of lectures to generate questions.

Acknowledgements

We thank XXX for their support in the data collection, preparation and scoring, and XXX for proofreading the manuscript.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*, 659–701. doi:10.3102/0034654316689306

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861–876. doi:10.1002/acp.1391

Andre, M. L. D. A., & Anderson, T. H. (1978-1979). The development and evaluation of a self-questioning study technique. *Reading Research Quarterly, 14,* 605-623.

Bates, S. P., Galloway, R. K., Riise, J., & Homer, D. (2014). Assessing the quality of a student-generated question repository. *Physics Education Research, 10*, 020105. doi: 10.1103/PhysRevSTPER.10.020105.

Berry, J. W., & Chew, S. L. (2008). Improving learning through interventions of student-generated questions and concept maps. *Teaching of Psychology, 35,* 305–312. doi:10.1080/00986280802373841

Bugg, J. M., & McDaniel, M. A. (2012). Selective benefits of question self-generation and answering for remembering expository text. *Journal of Educational Psychology, 104,* 922–931. doi:10.1037/a0028661

Bui, D. C., Myerson, J., & Hale, S. (2013). Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology, 105*, 299-309. doi: 10.1037/a0030367

Byun, H., Lee, J., & Cerreto, F. A. (2014). Relative effects of three questioning strategies in ill-structured, small group problem solving. *Instructional Science, 42*, 229–250. doi:10.1007/s11251-013-9278-1

Doctorow, M., Wittrock, M. C., & Marks, C. (1978). Generative processes in reading comprehension. *Journal of Educational Psychology, 70*, 109-118. doi: 10.1037/0022-0663.70.2.109

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. doi: 10.3758/BRM.41.4.1149

Foos, P. W. (2016). Effects of student-written questions on student test performance. *Teaching of Psychology*, *16*, 77–78. doi:10.1207/s15328023top1602_10

Hagaman, J. L., Casey, K. J., & Reid, R. (2010). The effects of the paraphrasing strategy on the reading comprehension of young students. *Remedial and Special Education, 33,* 110–123. doi:10.1177/0741932510364548

King, A. (1989). Effects of self-questioning students' comprehension training on college of lectures. *Contemporary Educational Psychology, 14*, 366–381. doi: 10.1016/0361-476X(89)90022-2

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27*, 664-687. doi: 10.3102/00028312027004664

King, A. (1991). Improving lecture comprehension: Effects of a metacognitive strategy. *Applied Cognitive Psychology, 5*, 331–346. doi:10.1002/acp.2350050404

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29,* 303–323. doi: 10.2307/1163370

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*, 338–368. doi: 10.3102/00028312031002338

Knapp, T. R., & Schafer, W. G. (2009). From gain score *t* to ANCOVA *F* (and vice versa). *Practical Assessment, Research & Evaluation, 14*, 1-7.

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23,* 425–449. doi:10.1016/S0022-5371(84)90282-2

Levin, A., & Arnold, K.-H. (2008). Fragen stellen, um Antworten zu erhalten – oder Fragen generieren, um zu lernen? *Zeitschrift für Pädagogische Psychologie, 22,* 135–142. doi:10.1024/1010-0652.22.2.135

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27,* 521–536. doi: 10.1016/0749-596X(88)90023-X

Nolan, T. E. (1991). Self-questioning and prediction: Combining metacognitive strategies. *Journal of Reading, 35,* 132-138

Papinczak, T., Babri, A. S., Peterson, R., Kippers, V., & Wilkinson, D. (2011). Students generating questions for their own written examinations. *Advances in Health Sciences Education: Theory and Practice, 16*(5), 703–710. doi:10.1007/s10459-009-9196-9

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255. doi: 10.1111/j.1467-9280.2006.01693.x

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

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*, 479–530. doi:10.3102/00346543064004479

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*, 181–221. doi: 10.3102/00346543066002181

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*, 1432–1463. doi:10.1037/a0037559

Rummer, R., Schweppe, J., & Schwede, A. (in press). Open-book vs. closed-book tests in university classes: a field experiment. Special Issue on "How desirable are 'desirable difficulties' for learning in educational contexts?" *Frontiers in Educational Psychology*. doi: 10.3389/fpsyg.2019.00463

Song, D. (2016). Student-generated questioning and quality questions: A literature review. *Research Journal of Educational Studies and Review, 2*, 58–70.

van Blerkom, M. L., & van Blerkom, D. L. (2004). Self-monitoring strategies used by developmental and non-developmental college students. *Journal of College Reading and Learning, 34,* 45-60. doi: 10.1080/10790195.2004.10850161

van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies corrected. *Journal of Clinical Epidemiology, 59*, 920–925. doi: 10.1016/j.jclinepi.2006.02.007

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2017). Learning terms and definitions: Drawing and the role of elaborative encoding. *Acta Psychologica, 179*, 104–113. doi: 10.1016/j.actpsy.2017.07.008

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychology, 19*, 87-95. doi: 10.1080/00461527409529129

APPENDIX

Questions of the pre-test, assessing prior knowledge, and of the final test.

(Maximum score in parentheses; words in bold were the keywords printed in bold in the text and are identical to the keywords provided to the learners in the cued closed-book condition – they were not in bold in the pre- and final test)

(1) What is **dyscalculia**? (1.5)

(2) What needs to considered, according to the ICD-10, for a **diagnosis of dyscalculia**? (3)

(3) What is the **main characteristic of dyscalculia**? (1.5)

(4) What are **causes** of dyscalculia? (2.5)

(5) How high is the **incidence of dyscalculia**? (1)

(6) What are **symptoms of dyscalculia**? (9)

(7) Which **psychological deviations** can result from dyscalculia? (10)

(8) What are **potential effects of such psychological deviations**? (7)

(9) What factors contribute to a **positive prognosis** of dyscalculia? (2)

(10) What are the aims of **interventions with regard to dyscalculia**? (1)