WILEY

Does generation benefit learning for narrative and expository texts? A direct replication attempt

Julia Schindler[1] | Tobias Richter[1] | Raymond Mar[2]

[1]Department of Psychology IV, University of Würzburg, Würzburg, Germany

[2]Department of Psychology, York University, Toronto, Ontario, Canada

Correspondence
Julia Schindler, Department of Psychology IV, University of Würzburg, Röntgenring 10, 97070 Würzburg, Germany.
Email: julia.schindler@uni-wuerzburg.de

Abstract

Generated information is better recognized and recalled than information that is read. This so-called *generation effect* has been replicated several times for different types of material, including texts. Perhaps the most influential demonstration was by McDaniel et al. (1986, *Journal of Memory and Language, 25*, 645–656; henceforth MEDC). This group tested whether the generation effect occurs only if the generation task stimulates cognitive processes not already stimulated by the text. Numerous studies, however, report difficulties replicating this text by generation-task interaction, which suggests that the effect might only be found under conditions closer to the original method of MEDC. To test this assumption, we will closely replicate MEDC's Experiment 2 in German and English-speaking samples. Replicating the effect would suggest that it can be reproduced, at least under limited conditions, which will provide the necessary foundation for future investigations into the boundary conditions of this effect, with an eye towards its utility in applied contexts.

KEYWORDS
expository texts, generation effect, learning, narrative texts, replication

1 | INTRODUCTION

Generated information is often better recognized and recalled than information that is passively encoded. This phenomenon is called the *generation effect* and it was first reported for the learning of word pairs (e.g., McDaniel et al., 1988; Slamecka & Graf, 1978). The classical paradigm consists of two experimental conditions in which learners are presented with two words: (1) a context word (e.g., WINTER), and (2) a related target word (e.g., SNOW) which they should memorize for a learning test. In the generation condition, the fragmented target word (e.g., S_ _ W) is presented, and the learner must generate the target word (SNOW). In the reading condition, the same word is presented intact. For both conditions, participants must write down the target words during the learning phase. Typically, learners recognize and recall the generated target words better than the target words they merely read. This effect has been replicated several times for

different types of stimulus material, including numbers (e.g., Gardiner & Rowley, 1984), words (e.g., McDaniel et al., 1988; Slamecka & Graf, 1978), sentences (e.g., Graf, 1980, 1981; Lutz et al., 2003), and even rich textual information such as song lyrics (Goldman & Kelley, 2009) and recipes (Goverover et al., 2008, 2010, 2013, 2014). The effect has also been demonstrated using a diverse range of generation tasks, including fill-in-the-blanks (Abel & Hänze, 2019), letter completion (e.g., Einstein et al., 1984; McDaniel & Kerwin, 1987), unscrambling (e.g., Graf, 1982; McDaniel et al., 1986), and mental rotation (e.g., Graf, 1982). Moreover, the memory benefit has also emerged for a wide range of measures, including recognition, cued and free recall (e.g., McDaniel et al., 1988; Slamecka & Graf, 1978), and for cloze tasks (DeWinstanley & Bjork, 2004). Finally, the generation effect has been demonstrated with different retention intervals and for between- and within-subject designs (see the meta-analysis by Bertsch et al., 2007).

1.1 | The generation effect for narrative and expository texts

One form of the generation effect, which is particularly relevant to educational contexts, is improving memory for complex narrative and expository texts (Einstein et al., 1990; McDaniel et al., 1986; McDaniel et al., 1994; McDaniel et al., 2002). Text generation comprises activities for creating the learning material—or at least parts of it—instead of being presented with an intact text. Two generation paradigms in particular that have been empirically shown to be beneficial for learning with texts are letter completion (filling in letters deleted throughout the text) and sentence unscrambling (arranging randomly-ordered sentences into a meaningful order) (e.g., Einstein et al., 1984, 1990; McDaniel et al., 1986, 2002).

However, in contrast to word learning studies, generation does not seem to consistently enhance learning for full texts. Some studies, for example, have found no generation effect for narrative texts (Einstein et al., 1990, Exp. 1; McDaniel et al., 1986, Exp. 1). Others have failed to replicate the effect for learners with good comprehension skills (Abel & Hänze, 2019; Exp. 2a of McDaniel et al., 2002) or for expository texts (Maki et al., 1990, Exp. 1; Schindler et al., 2017; Thomas & McDaniel, 2007, Exp. 1).

These mixed findings suggest that text generation fails to provide a consistent advantage over reading for all learners across different conditions. Generation can be understood as an example of *desirable difficulties,* which are educational measures that make learning intentionally more difficult to improve outcomes (R. A. Bjork, 1994; E. L. Bjork & Bjork, 2011). One contextual framework advanced by McDaniel and Butler (2010), describes the outcomes of desirable difficulties as a complex interaction of learner characteristics, type of test, learning materials, and tasks (see also Einstein et al., 1990; McDaniel & Einstein, 1989, 2005). According to this framework, learning can be improved only when difficulties stimulate unique cognitive processes that are not already elicited by the learners and when the test requirements match the processes stimulated by the generation task (Schindler et al., 2019).

The material by processing-task interaction proposed by this framework may also explain a phenomenon observed for narrative and expository texts. Learning with narrative texts can be enhanced by letter completion, and learning with expository texts can be enhanced by unscrambling sentences (Einstein et al., 1990; McDaniel et al., 1986, 2002). However, the inverse appears to not to have an effect. That is, unscrambling does not benefit learning from narratives and letter completion does not benefit learning from expository texts (Einstein et al., 1990; McDaniel et al., 1986). A potential explanation for this divergence was first proposed in the *Material Appropriate Processing* (MAP) framework (McDaniel et al., 1986; McDaniel & Einstein, 1989), now one of the components of the contextual framework proposed by McDaniel and Butler (2010). According to McDaniel and colleagues, narrative and expository texts have qualitatively different encoding demands, which interact with the type of generation task. A learning benefit can only be observed when the generation task is appropriate for the learning material such that it stimulates cognitive processing that was not already elicited by the material content.

Narratives typically possess the regular and familiar structure of a story schema (Rumelhart, 1975). This story schema stimulates relational processing of the narrative's propositions, aiding organization and integration (McDaniel et al., 1986). Generating texts by unscrambling sentences also stimulates relational processing because rearranging the sentences into a meaningful order requires organization and integration of the sentence propositions. According to the MAP framework, unscrambling has no effect on learning from a narrative because unscrambling elicits a process already present during narrative comprehension as opposed to eliciting a novel cognitive process. Letter completion, in contrast, stimulates individual-item or proposition-specific processing, that is, processing of lexical concepts or relations between the concepts of a proposition (McDaniel et al., 1986). This process is not present during narrative comprehension. Thus, the letter completion task stimulates unique cognitive processes and thereby enhances learning.

Expository texts, in contrast to narrative texts, stimulate individual-item or proposition-specific processing. This type of text directly focuses on the comprehension of new or unfamiliar concepts instead of stimulating organizational and integrative processing between propositions (Einstein et al., 1990; McDaniel et al., 1986). According to the MAP framework, learning with expository texts is improved by unscrambling because this task stimulates cognitive processes not already elicited by the expository texts. Letter completion, however, has no effect on learning with expository texts because this form of generation task stimulates individual-item or proposition-specific processing, which is already elicited by simply reading the expository text.

1.2 | Replication difficulties

Despite this appealing explanation, a considerable number of studies report inconsistent findings regarding this genre by generation-type interaction. Letter completion, for example, had no effect on recall for narratives in some studies (Einstein et al., 1990, Exp. 1; McDaniel et al., 1986, Exp. 1; McDaniel et al., 2002; Exp. 2A), and benefited recall for expository texts in other studies (E. L. Bjork & Storm, 2011, Exp.1–4; Burnett & Bodner, 2014; Exp. 1 and 2; DeWinstanley & Bjork, 2004, Exp. 1A-3; Maki et al., 1990, pilot study and Exp. 1), both findings of which were unexpected according to the MAP framework. Similarly, sentence unscrambling had no effect on recall for expository texts in some studies (McDaniel et al., 2002, Exp. 1B; Thomas & McDaniel, 2007, Exp.1). Also, in some studies, sentence unscrambling unexpectedly benefitted learning from narratives (Einstein et al., 1990, Exp. 2; McDaniel et al., 1994, Exp. 1–3). Moreover, Schindler and Richter (in preparation) ran a series of six experiments on this topic under ecologically-valid and methodologically-stringent conditions. The generation effect was found in only one of the experiments. This particular experiment most closely resembled the original studies by McDaniel and colleagues (e.g., McDaniel et al., 1986, 2002) because learners were not informed of the subsequent test and were allowed to take as much time as needed to read or generate the texts. These findings suggest that the generation effect, though difficult to replicate under ecologically-valid and methodologically-stringent

conditions, might be replicable under conditions closer to the setting of the original studies.

These conflicting findings suggest that the text-generation effect is either unreliable (and thus not useful for educational contexts) or it is moderated by contextual factors (McDaniel & Butler, 2010). However, a replication of the original interaction effect is an important and necessary precondition before investigating potential moderating contextual factors. A complete replication of the effect under these limited conditions will open the door to methods that control for contextual factors under which the text generation effect might emerge. Thus, this replication is necessary before text generation should be utilized as a learning intervention in pedagogical contexts.

1.3 | The present study

The aim of the present study is to replicate the genre by generation-task interaction found by McDaniel et al. (1986, MEDC). Their study was the first to propose, test, and support the idea that different generation tasks (letter completion vs. sentence unscrambling) work differently in combination with different types of texts (narratives vs. expository). The present study will replicate Experiment 2 of MEDC because the findings provided more convincing evidence for the framework compared to Experiment 1. Our study will employ an English-speaking sample to keep the replication as close as possible to the original method, and will also investigate a German-speaking sample (using translated materials) to examine whether the generation effect generalizes to another language.

Participants will read or generate a short narrative ("The Just Reward" by Guterman, 1945) or an expository text passage ("The Frozen Country," modified from Levy, 1981, by MEDC). Participants in the generation conditions will either fill in missing letters or reorder scrambled sentences. Subsequently, in an unannounced learning test, they will be asked to write down as much information from the texts as they can remember. We are in close contact with the first author of MEDC to ensure that our method matches as closely as possible to the original study method (i.e., material, study design, instructions, and analyses). We expect to find the generation effect, showing a higher proportion of free recall for narrative texts when missing letters are completed (compared to the reading control condition) and for expository texts when scrambled sentences are reordered (compared to the reading control condition). No generation effect or a substantially smaller effect is predicted for unscrambling narratives and for completing expository texts.

2 | METHOD

2.1 | Ethics statement

The study will be conducted in full accordance with the Ethical Guidelines of the German Psychological Society (DGPs), the Canadian Tri-Council Research Ethics guidelines and the American Psychological Association (APA), and it has been already approved by the local ethics committees.

2.2 | Transparency

The data will be made available on the Open Science Framework.

2.3 | Participants

As in the original study, participants will be enrolled in introductory psychology classes (psychology and teaching students) and receive extra course credit for their participation. In the original study, a total of 72 students were randomly assigned to six experimental groups in a 2 (narrative vs. expository text) × 3 (letter completion vs. sentence unscrambling vs. reading control) between-subjects design, resulting in 12 participants per group.

In a meta-analysis of the generation effect, Bertsch et al. (2007) found that the effect was about half the size in between-subjects designs (d = 0.28) compared to within-subjects designs (d = 0.50). However, this meta-analysis included no studies that used full text generation. Thus, whether these findings can be generalized to text generation is still an open question. To date, the text-generation effect has been demonstrated with both between-subjects designs (e.g., Abel & Hänze, 2019; Einstein et al., 1984; MEDC) and within-subjects designs (E. L. Bjork & Storm, 2011; Goldman & Kelley, 2009; McDaniel et al., 1989).

Using a between-subjects design, MEDC reported main effects and interactions of considerable size. As predicted, the authors reported a large effect showing that participants who completed letters for a narrative recalled more information compared to those who read it, $F(1, 66)$ = 63.77 (equivalent to η_p^2 = .49). Participants who unscrambled sentences for an expository text also recalled more information compared to those reading it, $F(1, 66)$ = 30.57 (η_p^2 = .32). Lastly, for the overall interaction, letter completion led to the best recall with the narrative text, and unscrambling with the expository text, $F(2, 66)$ = 33.57 (η_p^2 = .50).

Despite these large effects in the original study, we entered a medium-sized effect (η^2 = .06; Cohen, 1988) in our power analysis using G*Power (Faul et al., 2007), which revealed a required sample size of N = 251, with power (1-β) set to .95 and an α-level of .05. We increased the target sample size to N = 300 (for both the English and German versions, a total of 600) to account for the substitution of participants who will report that they have expected the learning test with participants who will report that they had not, as in the MEDC study. Participants will be tested in small groups of one to four and be required to provide written consent before testing.

2.4 | Materials and procedure

The materials and procedure in the present study are as described in MEDC. Half of the participants in the English-speaking sample will be

presented with the English version of the Russian narrative, "The Just Reward" (Guterman, 1945), and the other half with the expository text "The Frozen Continent" (modified from Levy, 1981, by MEDC). Both texts will be presented with titles. Each text contains 20 sentences with 83 idea units in the narrative text and 69 idea units in the expository text. For the German-speaking sample, both texts will be translated to German. Texts will be translated by a speaker with native fluency in both languages, and the quality of translation will be assessed with back-translation using a second translator. In the letter completion condition, 18% of the letters will be randomly deleted and replaced with blanks of which 40% will be vowels. In the sentence unscrambling condition, participants will be randomly assigned to one of two conditions, each with text consisting of 20 sentences randomly ordered.

Participants will not be informed of the learning test. Instead, they will be told that the aim of the study is to investigate their text comprehension. Processing time to read or generate the texts will be recorded, and time on task will not be limited. Participants will then provide a comprehensibility rating for the text on a 5-point Likert scale (1 = *did not comprehend the passage at all*, 5 = *comprehended the passage very well*). After a distraction task (i.e., working on math problems for 5 min), the memory test will be administered. Participants will be asked to write down as much information about the text that they can remember in as much time as they need.

2.4.1 | Post-experimental questionnaire

In a post-experimental questionnaire, participants will be asked to indicate whether they had expected the memory test or not. In addition (and as an extension of the original study), demographics such as gender, age, first language, field of study (psychology or teaching), prior knowledge of the expository text content, and familiarity with the narrative will be assessed. These variables will be analyzed to check for comparability of the experimental groups. In case of significant differences between groups, these variables will be statistically controlled in additional analyses.

3 | ANALYSES AND EXPECTED RESULTS

Separate analyses will be conducted for the German and the English-speaking samples. As in the original study by MEDC, participants who report that they had expected the memory test will be excluded from analyses. We will run additional analyses that include the whole sample, if we find that the difference between the memory-test means of the excluded participants and the included participants is nonsignificant. All statistical significance tests will be based on a Type-I error rate of .05. Statistically significant results will also be subjected to a Bonferroni–Holm correction for multiple testing.

3.1 | Generation accuracy

As in the original study, the mean proportions of letters that are generated correctly in the letter completion condition will be computed for both texts. Mean deviation scores (i.e., the deviation of the sentences from their original position in the unscrambled text) will also be calculated for the narrative and the expository text in the unscrambling condition and compared using a two-sample t test. Mean deviation scores for both texts will be compared with the deviation scores of the randomly scrambled texts, separately for both random orders, using one-sample t tests. If the participants carefully order the scrambled sentences, the ordered texts should approximate the proper order of sentences and should show lower deviation scores than the two random sentence orders.

3.2 | Processing time

Effects of genre (narrative vs. expository) and learning condition (letter completion vs. sentence unscrambling vs. reading control) on processing time will be analyzed in a two-factor between-subjects analysis of variance (ANOVA), with processing time as dependent variable. According to MEDC, participants in the letter completion and sentence unscrambling conditions are predicted to have longer processing times for both texts because of increased encoding difficulty compared to those in the reading control condition.

3.3 | Comprehensibility ratings

Effects of genre and learning condition on text comprehensibility will be analyzed in a two-factor between-subjects ANOVA with comprehensibility ratings as the dependent variable. Although it seems plausible to assume that the reported comprehensibility of the two texts will differ for the letter completion and sentence unscrambling conditions (compared to the reading control condition), MEDC found no effect of learning condition on the comprehensibility ratings. They found, however, a main effect of genre, indicating better comprehensibility of the narrative compared to the expository text. Thus, in the present study, the narrative is expected to be rated as more comprehensible than the expository text.

3.4 | Free recall

As in the original study, recall accuracy for each idea unit in the two texts will be scored (correctly recalled = 1, not mentioned or incorrectly recalled = 0; no partial scoring). About 9% of the protocols will be scored by two different raters. If inter-rater reliability is high (Cohen's κ > .75), all of the remaining protocols will be scored by one of the two raters. Effects of genre and learning condition on free recall will be analyzed in a two-factor between-subjects ANOVA with proportion of correctly recalled information as the dependent variable. In

line with MEDC, an interaction is expected. Participants should recall more information correctly when they complete letters in the narrative and when they reorder the sentences of the expository text (compared to the reading control condition). However, recall will not improve or will improve less when unscrambling sentences of the narrative or completing letters in the expository text compared to the reading control condition. This interaction should persist even if processing time is included as a covariate in an analysis of covariance (ANCOVA).

To account for the multilevel structure of the data (idea units and participants), an additional Generalized Linear Mixed Model (GLMM) with a logit link function will be estimated for recall accuracy of idea units (Dixon, 2008). Genre (-1 = Narrative, 1 = Expository text) will be included as contrast-coded predictor variable in the GLMM. Learning condition will be included as two contrast-coded predictor variables, each comparing one of the generation conditions with the reading control condition (-1 = Reading, 1 = Letter completion, 0 = Sentence unscrambling; -1 = Reading, 0 = Letter completion, 1 = Sentence unscrambling). Test expectancy will be included as dummy-coded predictor variable with participants who report that they had not expected a test being the reference group (test not expected = 0, test expected = 1). Also, interaction terms will be included. Processing time will be included as a grand-mean centered control variable. Intercepts for participants and idea units will be allowed to vary randomly. Hypotheses are the same as for the ANCOVA reported by MEDC. In the case that the experimental groups are not comparable with respect to participants' prior knowledge and familiarity, both variables will be statistically controlled in the GLMM.

3.5 | Timeline

Data collection in the laboratory will likely not be possible before Spring 2021 because of the COVID-19 pandemic. Once started, we aim for data collection and analysis to be completed within 5 months so that the manuscript can be resubmitted for Stage 2 review in autumn 2021.

4 | CONCLUSION

The existing research has produced inconsistent findings regarding the reliability and generalizability of the text generation effect. The present study aims to closely replicate the most influential demonstration of the genre by generation-task interaction, Experiment 2 of McDaniel et al. (1986), using both a German- and an English-speaking sample. Replicating this interaction in these two samples would provide valuable evidence that the effect can be reliably reproduced, at least under limited conditions. It would also provide the foundation for directed investigations into the boundary conditions of this effect, with an eye towards its utility in applied contexts.

ORCID
Julia Schindler https://orcid.org/0000-0002-5833-1334
Tobias Richter https://orcid.org/0000-0002-0467-9044

REFERENCES

Abel, R., & Hänze, M. (2019). Generating causal relations in scientific texts: The long-term advantages of successful generation. *Frontiers in Psychology*, 10, 199. https://doi.org/10.3389/fpsyg.2019.00199

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35, 201–210. https://doi.org/10.3758/BF03193441

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bjork, E. L., & Storm, B. C. (2011). Retrieval experience as a modifier of future encoding: Another test effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1113–1124. https://doi.org/10.1037/a0023549

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Burnett, A. N., & Bodner, G. E. (2014). Learnin' 'bout my generation? Evaluating the effects of generation on encoding, recall, and metamemory across study-test experiences. *Journal of Memory and Language*, 75, 1–13.

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

DeWinstanley, P. A., & Bjork, E. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, 32, 945–955. https://doi.org/10.3758/BF03196872

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, 59, 447–456. https://doi.org/10.1016/j.jml.2007.11.004

Einstein, G. O., McDaniel, M. A., Bowers, C. A., & Stevens, D. T. (1984). Memory for prose: The influence of relational and proposition-specific processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 133–143. https://doi.org/10.1037/0278-7393.10.1.133

Einstein, G. O., McDaniel, M. A., Owen, P. D., & Coté, N. C. (1990). Encoding and recall of texts: Importance of material appropriate processing. *Journal of Memory and Language*, 29, 566–581. https://doi.org/10.1016/0749-596X(90)90052-2

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. https://doi.org/10.3758/BF03193146

Gardiner, J. M., & Rowley, J. M. C. (1984). A generation effect with numbers rather than words. *Memory and Cognition*, 12, 443–445. https://doi.org/10.3758/BF03198305

Goldman, B. A., & Kelley, M. R. (2009). The generation effect in the context of lyrical censorship. *Psi Chi Journal of Undergraduate Research*, 14, 72–78. https://doi.org/10.24839/1089-4136.JN14.2.72

Goverover, Y., Chiaravalloti, N., & DeLuca, J. (2008). Self-generation to improve learning and memory of functional activities in persons with multiple sclerosis: Meal preparation and managing finances. *Archives of Physical Medicine and Rehabilitation*, 89, 1514–1521. https://doi.org/10.1016/j.apmr.2007.11.059

Goverover, Y., Chiaravalloti, N., & DeLuca, J. (2010). Pilot study to examine the use of self-generation to improve learning and memory in people with traumatic brain injury. *American Journal of Occupational Therapy*, 64, 540–546. https://doi.org/10.5014/ajot.2010.09020

Goverover, Y., Chiaravalloti, N., & DeLuca, J. (2013). The influence of executive functions and memory on self-generation benefit in persons with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 35, 775–783. https://doi.org/10.1080/13803395.2013.824553

Goverover, Y., Chiaravalloti, N., & DeLuca, J. (2014). Task meaningfulness and degree of cognitive impairment: Do they affect self-generated learning in persons with multiple sclerosis? *Neuropsychological Rehabilitation: An International Journal*, 24, 155–171. https://doi.org/10.1080/09602011.2013.868815

Graf, P. (1980). Two consequences of generating: Increased inter- and intraword organization of sentences. *Journal of Verbal Learning and Verbal Behavior*, 19, 316–327. https://doi.org/10.1016/S0022-5371(80)90248-0

Graf, P. (1981). Reading and generating normal and transformed sentences. *Canadian Journal of Psychology*, 35, 293–308. https://doi.org/10.1037/h0081193

Graf, P. (1982). The memorial consequences of generation and transformation. *Journal of Verbal Learning and Verbal Behavior*, 21, 539–548. https://doi.org/10.1016/S0022-5371(82)90764-2

Guterman, N. (Ed.). (1945). *Russian fairy tales*. Pantheon.

Levy, B. A. (1981). Interactive processes during reading. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. l–35). Erlbaum.

Lutz, J., Briggs, A., & Cain, K. (2003). An examination of the value of the generation effect for learning new material. *The Journal of General Psychology*, 130, 171–188. https://doi.org/10.1080/00221300309601283

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 609–616. https://doi.org/10.1037/0278-7393.16.4.609

McDaniel, M. A., & Butler, A. C. (2010). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 175–198). Psychology Press.

McDaniel, M. A., & Einstein, G. O. (1989). Material-appropriate processing: A contextualist approach to reading and studying strategies. *Educational Psychology Review*, 1, 113–145. https://doi.org/10.1007/BF01326639

McDaniel, M. A., & Einstein, G. O. (2005). Material appropriate difficulty: A framework for determining when difficulty is desirable for improving learning. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 73–85). American Psychological Association. Retrieved from https://doi.org/10.1037/10895-006

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, 25, 645–656. https://doi.org/10.1016/0749-596X(86)90041-0

McDaniel, M. A., Hines, R. J., & Guynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language*, 46, 544–561. https://doi.org/10.1006/jmla.2001.2819

McDaniel, M. A., Hines, R. J., Waddil, P. J., & Einstein, G. O. (1994). What makes folk tales unique: Content familiarity, causal structure, scripts, or superstructures? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 169–184. https://doi.org/10.1037/0278-7393.20.1.169

McDaniel, M. A., & Kerwin, M. L. E. (1987). Long-term prose retention: Is an organizational schema sufficient? *Discourse Processes*, 10, 237–252. https://doi.org/10.1080/01638538709544674

McDaniel M. A., Ryan E. B., & Cunningham C. J. (1989). Encoding difficulty and memory enhancement for young and older readers. *Psychology and Aging*, 4(3), 333–338.

McDaniel, M. A., Waddil, 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. https://doi.org/10.1016/0749-596X(88)90023-X

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 211–236). Academic Press. Retrieved from https://doi.org/10.1016/B978-0-12-108550-6.50013-6

Schindler, J., Richter, T., & Eyßer, C. (2017). Mood moderates the effect of self-generation during learning. *Frontline Learning Research*, 5(4), 76–88. https://doi.org/10.14786/flr.v5i4.296

Schindler, J., & Richter, T. (in preparation). *An ecologically valid test of the text generation effect*.

Schindler, J., Schindler, S., & Reinhard, M.-A. (2019). Effectiveness of self-generation during learning is dependent on individual differences in need for cognition. *Frontline Learning Research*, 7(2), 23–39. https://doi.org/10.14786/flr.v7i2.407

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, 4, 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study–test processing in memory and metacomprehension. *Memory & Cognition*, 35, 668–678. https://doi.org/10.3758/BF03193305