

Journal of Experimental Psychology: Applied

Generating Mnemonics Boosts Recall of Chemistry Information

Jonathan G. Tullis and Jiahui Qiu

Online First Publication, May 3, 2021. <http://dx.doi.org/10.1037/xap0000350>

CITATION

Tullis, J. G., & Qiu, J. (2021, May 3). Generating Mnemonics Boosts Recall of Chemistry Information. *Journal of Experimental Psychology: Applied*. Advance online publication. <http://dx.doi.org/10.1037/xap0000350>

Generating Mnemonics Boosts Recall of Chemistry Information

Jonathan G. Tullis and Jiahui Qiu

Department of Educational Psychology, University of Arizona

Students frequently generate mnemonic cues to help them remember difficult or abstract information (Tullis & Maddox, *Metacognition and Learning*, 2020, 15, 129). Self-generated mnemonics have the potential to be particularly effective means of remembering target information because they can transform abstract information into meaningful units, connect information to existing schema, and create distinct retrieval routes to the targets. Across five experiments, we compared the effectiveness of self-generated mnemonics to mnemonics generated by others for remembering chemistry information. Generating one's own mnemonics consistently boosted recall for both the chemistry content and the mnemonic itself. However, experimentally boosting recall of mnemonics through retrieval practice did not affect recall of associated chemistry content. These results indicate that improved recall of chemistry content is not caused by better recall of the mnemonic itself; rather, generating a mnemonic involves deep and effortful processing of chemistry content that boosts recall more than reading someone else's mnemonic.

Public Significance Statement

Creating their own mnemonics can help students remember chemistry content and the mnemonic better than reading mnemonics generated by others. Generating mnemonics is an effective strategy to remember information because it prompts students to deeply process the material.

Keywords: mnemonics, generation, metacognitive control

Generating effective mnemonic cues may be one of the most valuable techniques by which students control and improve their long-term learning. Mnemonics have a long history of helping people remember important information. Cicero practiced the method of loci mnemonic by connecting items in a list to specific familiar locations in ancient times (Bower, 1970). During the Renaissance, students used mnemonic cues to help them learn grammar and the alphabet (Patten, 1990). To this day, people generate mnemonic cues to help remember important information (Worthen & Hunt, 2011). For example, 73% of those surveyed by Harris (1980) reported using the first letter of to-be-remembered items to create a simpler memorable word (e.g., list of Great Lakes = HOMES). Middle school, high school, and college students report generating mnemonic cues—rhymes, acronyms, songs, and stories—to remember connections among important ideas (Tullis & Maddox, 2020; Van Etten et al., 1997). Across five experiments, we tested the effectiveness of self-generated mnemonics for bolstering recall of chemistry information.

Creating mnemonic cues to organize and retrieve information may be a uniquely effective way to support long-term learning because mnemonic cues improve initial encoding *and* create strong retrieval routes to the target information (Crutcher & Ericsson, 2000). More specifically, mnemonics can transform difficult-to-

remember information into something meaningful (Worthen & Hunt, 2011), connect it to other information in learners' own long-term memory (Mastropieri et al., 2000), and create strong and personal retrieval routes to the target that make accessing it in memory easier (Atkinson & Raugh, 1975). Instructing learners to generate mnemonic cues can improve learning of foreign language vocabulary (Atkinson & Raugh, 1975), prose passages (Shriberg, 1982), state capitals (Levin et al., 1982), and even some biology information (Carney & Levin, 2003). Mnemonic use may be more effective for learning vocabulary than many other strategies, including presenting the new vocabulary in semantic context (McDaniel & Pressley, 1989). Further, some evidence suggests that students may remember more when they generate their own mnemonics than when they study mnemonics provided by a teacher (Bloom & Lamkin, 2006). The benefits of self-generated memory cues can sometimes persist over long retention intervals. For example, students who generated their own mnemonic for the names of the cranial nerves remembered more of the nerves than those who received the teacher's mnemonic 2 weeks after they studied the nerves, and the mnemonic advantage of the self-generated group grew after 10 weeks (Bloom & Lamkin, 2006; but see Wang et al., 1992 for an example where the advantages of mnemonics fade over long retention intervals). The use of mnemonics may have real-world consequences, as students who use mnemonics have higher GPAs than those who do not (Carlson et al., 1976).

Mnemonic cues may be especially helpful to recode information in domains where learners must master a wealth of unfamiliar vocabulary or abstract concepts. Indeed, students report generating mnemonic cues for classes that specifically involve learning a lot of new facts and terminology (McCabe et al., 2013), and relying on memory cues has advantages for complex domains with vast new

Jonathan G. Tullis  <https://orcid.org/0000-0001-5237-1776>

Correspondence concerning this article should be addressed to Jonathan G. Tullis, Department of Educational Psychology, University of Arizona, 1430 E. Second St., Tucson, AZ 85721, United States. Email: tullis@arizona.edu

vocabularies, including chemistry (Banks, 1941), physics (Gough, 1977), biology (Stagg & Donkin, 2016), and psychology (Richmond et al., 2011). One primary challenge for chemistry students is that chemistry concepts are abstract and unrelated to novices' existing semantic networks (Gabel, 1999). Students must learn an entirely new chemistry lexicon and countless new facts before they can develop more complex ideas (Balch, 2005; Battino, 1992). In fact, theorists have argued that memorizing facts and terminology may be the most appropriate learning method in introductory chemistry (Battino, 1992), especially for students who have little or no chemistry knowledge (Chittleborough et al., 2005). Given the breadth of new terminology and abstract knowledge expected of novice chemistry students, student-generated cues may be particularly effective for boosting chemistry learning by connecting novel facts to students' existing schema. Chemistry teachers have published mnemonic devices as early as 1941 (Banks, 1941) and as recently as 2010 (Stephens, 2010) to help students master the abstract new vocabulary. While countless mnemonics have been shared across chemistry teachers, no research has examined how student-generated chemistry mnemonics influence understanding and retention of students' knowledge as compared to those provided by teachers. Some prior research hints that student-generated mnemonic cues benefit related domains that suffer from large and abstract new lexicons that learners must master (e.g., psychology: Richmond et al., 2011).

Despite their general effectiveness, mnemonic cues can fail in two primary ways: A learner can fail to retrieve the cue (i.e., retrieval deficiency) or the learner can fail to accurately interpret the cue (i.e., decoding deficiency; Dunlosky et al., 2005). Generating one's own mnemonics may help learners both remember and decode their own mnemonics compared to cues provided by others (Tullis & Fraundorf, 2017). Generating a mnemonic cue may provide considerable benefit for remembering that cue later for two primary reasons. First, generation in and of itself is beneficial for memory. Generating an item to fit a specified context can require deep processing and that deep processing can bolster later retrieval (McFarland et al., 1980). As compared to reading a cue, generating a mnemonic cue should lead students to recall the cue more frequently during the final memory test (Dunlosky et al., 2005; Pyc & Rawson, 2010). Second, learner-generated cues may be memorable because they can be rooted in a learner's unique knowledge and experiences. Learners may rely upon their personal experiences to generate idiosyncratic and unique cues for themselves (Määntylä, 1986; Tullis & Benjamin, 2015b). Tying new knowledge into existing, idiosyncratic personal experiences elaborates the new information and is one of the most effective means of supporting memory (Symons & Johnson, 1997). Attaching abstract facts to a well-developed self-schema may help one's initial encoding and memory for chemistry information (Gabel, 1999). However, whether learners can find connections between abstract chemistry knowledge and their own personal schema remains untested. Inabilities to tie chemistry mnemonics to one's existing knowledge may limit the retrievability of self-generated mnemonics.

In addition to boosting recall of the mnemonic, generating one's own mnemonics may also support cue decoding because it can create distinct retrieval routes to the target. When generating mnemonic cues to support cued recall of simple materials for themselves, learners output distinct cues that limit the range of

possible target items (Tullis & Benjamin, 2015a, 2015b). Distinctiveness may be the most important factor for cueing recall because it limits the potential search space for the corresponding targets (Nairne, 2002). When learners generate cues for others, they decrease the distinctiveness of cues, and cues for others consequently do not support one's memory as much as cues generated for oneself (Tullis & Benjamin, 2015b). Cue decodability has been primarily assessed using relatively simple materials (e.g., Dunlosky et al., 2005). Decoding mnemonics may be challenging for chemistry content because the target chemistry information can be very complex, abstract, and multifaceted (Gabel, 1999). Across five experiments, we assessed the ability of learners to retrieve and decode self-generated mnemonics for chemistry content.

Experiment 1

In Experiment 1, learners studied a list of chemistry concepts. For half of the concepts, they generated their own mnemonic cue to help remember the idea; for the other half, they received a cue that a prior participant generated. The method of passing cues to the next participant has been used widely when testing the effectiveness of simple cues (e.g., Zhang & Tullis, 2021) because it controls for the content of the cues and the mnemonics are identical across conditions. This novel comparison between self-generated and peer-generated conditions tests whether self-generated cues have special properties (i.e., distinctiveness and connections to idiosyncratic knowledge) and processes (i.e., generation) that uniquely support memory, as suggested by prior theory (Tullis & Finley, 2018).

Method

Participants

Thirty-eight introductory educational psychology students at the University of Arizona participated in exchange for partial course credit. The first learner on each of four computers generated mnemonic cues for every concept so that cues could be passed along to later learners. These first four learners, however, could not receive any prior participants' cues (because there were no prior participants). Because they never studied another learner's cues, they did not take the chemistry test and are not included in any data analysis. The final sample, then, included 34 participants. A sample size of 34 was required to detect a Cohen's *d* effect size of 0.5 with alpha of .05 and power of .80.

Materials

Twenty-four chemistry concepts and corresponding questions were derived from the Next Generation Science Standards (Next Generation Science Standards Lead States, 2013). An example of a chemistry concept is: "In covalent bonds, electrons are shared between two atoms." The corresponding test question is: "What happens in a covalent bond?." The first author, who has a graduate degree in teaching high school science and taught high school chemistry classes, endorsed the content validity of the concepts and test questions. An experienced high school teacher who teaches AP and Honors chemistry and a chemistry professor at the University of Arizona who teaches introductory general chemistry reviewed the materials after the completion of the experiments and judged that the facts and questions accurately reflected material that

they teach in their classes. Chemistry concepts and data for all experiments can be found here: <https://osf.io/6z2xu/>

Procedure

The experiment was presented in MATLAB using the Psychophysics Toolbox (Brainard, 1997) and CogToolbox (Fraundorf et al., 2014). Learners completed the task on their own computer while up to three other learners completed the same task on other computers in the room. Learners were instructed to remember a list of chemistry concepts for an upcoming test. We told learners that they would sometimes generate their own mnemonic cues to help them remember the concepts and sometimes they would receive mnemonic cues that a different learner generated to help them remember the concepts. In the instructions, we provided four different examples of mnemonic cues for a variety of science facts (e.g., *Afferent neurons Arrive*, *Efferent neurons Exit*). Instructions specified that “Your cues can be images that you imagine, phrases you create, rhymes, or other connecting information that help you remember the concept.” During the mnemonic phase, learners studied the 24 chemistry concepts one at a time in a random order in black 40-point Arial font at the top of the computer screen. The first participant on each computer generated and typed in their mnemonic cues for every chemistry concept; their test data are not included because they never saw prior learners’ cues. All subsequent learners experienced both the self-generated and other-generated conditions. For a random half of the facts during this initial study, learners generated and typed in their own cues in a response box in the center of the screen. For the other half of the concepts, learners read the chemistry fact and the mnemonic cue that the prior learner on the computer had generated. When a concept was assigned to the other-generated condition, the computer displayed the mnemonic cue that was generated by the most recent prior participant in the middle of the screen. Concepts and conditions were displayed in a random order and the study/generate trials were entirely self-paced.

After studying or generating mnemonic cues for each of the chemistry concepts, participants immediately started the memory test. During the test, the list of 24 chemistry questions (e.g., “What happens in a covalent bond?”) was presented one at a time in a new random order. Participants answered these questions on their own; they were not explicitly told to use the mnemonics that they studied or generated during the earlier block. Finally, following the test for all the chemistry concepts, we tested memory for the chemistry mnemonics. The chemistry concepts were displayed one at a time on the screen (e.g., “In covalent bonds, electrons are shared between atoms.”), and participants were instructed to type in the corresponding mnemonic that was paired with each concept during the initial study phase. Test trials were entirely self-paced. On average, the complete experiment lasted 25.77 min ($SD = 7.39$).

Results

Data for all experiments in the manuscript are available at: <https://osf.io/6z2xu/> (Tullis & Qiu, 2021). For this and every following experiment, two independent coders rated each chemistry content test response and mnemonic cue memory test response as correct or not. Cohen’s Kappas were calculated to assess the agreement of coding between the two raters and are displayed in Table 1. Every disagreement was resolved through discussion between the raters.

Table 1

Cohen’s Kappa, a Measure of Reliability, for the Coding of Students’ Responses to Chemistry Questions, Tests of Mnemonic Memory During the Final Memory Test, and Tests of Mnemonic Memory During the Practice

Experiment	Content correct	Mnemonic memory	Practice mnemonic
Experiment 1	.87	.70	
Experiment 2	.86	.75	
Experiment 3	.83	.72	
Experiment 4	.56	.60	.66
Experiment 5	.93	.76	.79

Some examples of chemistry facts and learner-generated mnemonics are displayed in Table 2.

Learners remembered more chemistry concepts when they generated the mnemonic than when they read another learner’s mnemonic, $t(33) = 3.41$, $p = .002$, Cohen’s $d = 0.59$, as shown in Figure 1. Further, learners remembered more self-generated mnemonics than peer-generated mnemonics, $t(33) = 7.57$, $p < .001$, $d = 1.32$. When generating their own mnemonics, learners spent more time ($M = 40.89$ s, $SD = 18.01$) than when reading a peer’s mnemonic ($M = 24.53$, $SD = 11.29$), $t(33) = 7.67$, $p < .001$, $d = 1.34$.

Discussion

Experiment 1 showed that learners correctly answered more chemistry questions when they generated their own mnemonics than when they received others’ mnemonics. Further, learners remembered self-generated mnemonics better than mnemonics that their peers generated. Generation of mnemonics, then, can benefit recall of complex chemistry information. However, the benefit of generating one’s own mnemonics comes with a significant cost: Generating mnemonics takes substantially more time than reading others’ mnemonics.

Experiment 2

In Experiment 2, we tested whether the differential time spent between conditions in Experiment 1 caused the mnemonic differences between the conditions. To do so, we equated the study time between the self-generated condition and the peer-generated condition. If the differences in memory for the content and mnemonic persist even with study time equated, the differences between conditions likely reflect differential processing rather than differential total study time.

Method

Participants

As in Experiment 1, we aimed to recruit a sample size of 34 in order to detect a Cohen’s d effect size of 0.5 with alpha of .05 and power of .80. We recruited 10 participants from Amazon Mechanical Turk, who were paid \$10 for their participation. We recruited an additional 26 participants from introductory educational psychology courses, as in Experiment 1. The first participant-generated

Table 2*Examples of Chemistry Facts and Mnemonics Generated by Participants in Experiment 1*

Chemistry fact	Mnemonic generated
A valence electron is located in the outermost shell of an atom and participates in bonds with other atoms.	Valence electron = very edge (VE = VE)
Catalysts lower the activation energy needed to start a reaction.	Cats get down low to make you mad
Hydrogen bonds will only form with nitrogen, oxygen, and flourine.	fon (like phone)
In covalent bonds, electrons are shared between two atoms.	A couple celebrates their marriage bond by sharing buffalo wings
When acids and bases mix, a salt is formed.	A + B = S
J. J. Thomson discovered and identified the first subatomic particle: the electron.	J.J. Thomson electron

Note. All of the facts and mnemonics generated are available at the open science web link.

mnemonics for all of the chemistry facts and was not included in any further data analysis.

Materials

The materials were the same as in Experiment 1.

Procedure

The experiment was programmed using PsychoPy3 and learners completed the experiment online on their own. The instructions were the same as in Experiment 1. The first participant typed in their mnemonic cues for every chemistry concept; their test data are not included because they never saw prior learners' cues. All subsequent learners experienced both the self-generated and other-generated conditions. As in Experiment 1, learners generated and typed in their own cues in a response box in the center of the screen for a random half of the chemistry concepts. In the self-generated condition, participants advanced to the next screen by pressing enter after typing their mnemonic. The other half of the chemistry concepts were assigned to the peer-generated condition, in which learners read the chemistry fact and the mnemonic cue that a prior learner had

generated. The computer displayed the chemistry fact and the peer-generated mnemonic for the total amount of time that was taken for that fact by the prior participant to generate the mnemonic. When the time needed to generate the mnemonic by the prior participant elapsed, the screen automatically advanced. Yoking the time needed to generate the mnemonic with the study time for the subsequent participant who received it ensures that study times are equated between conditions, and this yoking procedure has been used in prior research (Tullis & Benjamin, 2011). After generating or reading each mnemonic, participants provided a judgment of how helpful they believed the mnemonic to be. They rated "How helpful is this mnemonic to remember this fact on a scale of 1 (not at all helpful) to 4 (very helpful)?" The test procedures were identical to Experiment 1.

Results

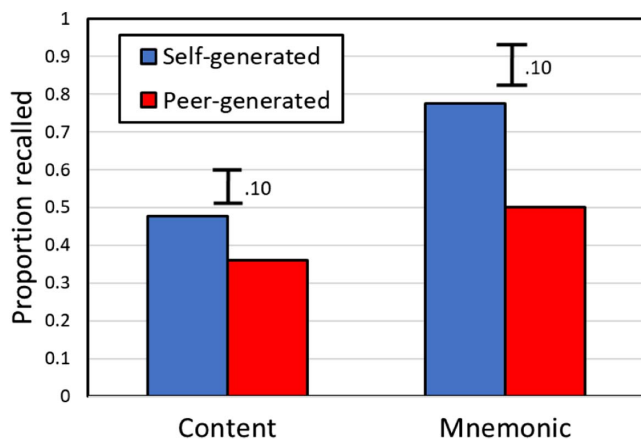
The patterns across MTurk participants and Ed Psych students were similar, so the data were combined across groups. Learners spent equivalent time-generating mnemonics ($M = 38.01$ s, $SD = 12.52$) as they did reading others' mnemonics ($M = 37.61$, $SD = 12.52$), $t(34) = .15$, $p = .89$, $d = 0.02$. This indicates that the yoking procedure ensured equivalent study time across the two conditions. Despite equivalent study time, learners remembered more chemistry concepts when they generated the mnemonic than when they read another learner's mnemonic cues, $t(34) = 2.79$, $p = .009$, $d = 0.48$, as shown in Figure 2. Learners also remembered more of the mnemonics they generated than mnemonics generated by others, $t(34) = 6.03$, $p < .001$, $d = 1.03$. Finally, learners rated their own mnemonics as more helpful for remembering the content ($M = 2.73$, $SD = .45$) than others' mnemonics ($M = 2.46$, $SD = .51$), $t(34) = 3.47$, $p = .001$, $d = 0.60$.

Discussion

Experiment 2 replicates the results from Experiment 1 while equating the study time between conditions. As in Experiment 1, learners correctly answered more chemistry questions (and remembered more of the mnemonics) when they generated their own mnemonics than when they received others' mnemonics. Even when time on task was controlled between conditions, generating one's own mnemonics boosted recall of complex chemistry information. Experiment 2 shows that time on task cannot explain the differences between the conditions because time was equated between conditions. Instead, we posit that the results indicate

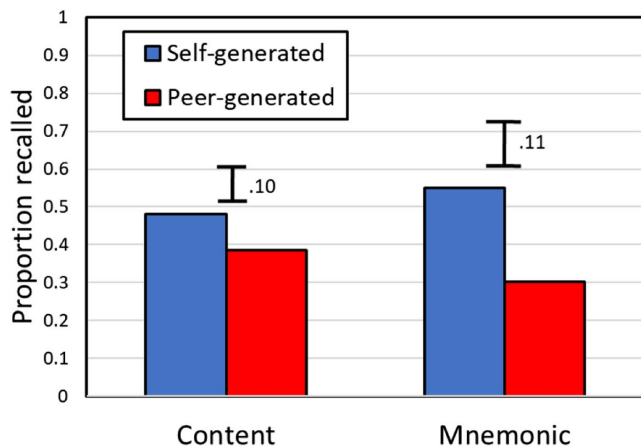
Figure 1

Proportion of Chemistry Content (Left) and Chemistry Mnemonics (Right) Recalled in Experiment 1



Note. The width of the error bars here and on all subsequent graphs indicate the within-subjects 95% confidence interval across conditions (Loftus & Masson, 1994). See the online article for the color version of this figure.

Figure 2
Proportion of Chemistry Content (Left) and Chemistry Mnemonics (Right) Recalled in Experiment 2



Note. See the online article for the color version of this figure.

that generating a mnemonic can prompt learners to identify important information in the chemistry fact, organize that information, and connect it to existing schema, which may all involve deep processing of the to-be-remembered information. Learners show metacognitive awareness that generating their own mnemonics is beneficial as they rate self-generated mnemonics as more useful for remembering than peer-generated mnemonics. We will return to these ideas in-depth after replicating and extending the results in three additional experiments.

Experiment 3

Experiment 3 was similar to the prior experiments, but it included an additional condition in which mnemonic cues were generated by chemistry experts. In Experiment 3, learners generated their own cue, received a cue generated by a peer, or received a cue generated by a chemistry graduate student. Mnemonic cues generated by experts may differ from those generated by novices. Deeper understanding of content may allow experts to identify aspects of the concepts that are crucial to remember. Experts may be able to generate cues that appropriately mirror the structure of the domain, and this mapping may be vital to supporting learning. Alternatively, expert-generated cues may not support students' memories as well as student-generated cues because experts may struggle to accurately take the perspective of novices. Experts' perspectives are far removed from novices' perspectives; consequently, experts may struggle to anticipate what novices will and will not understand. For example, teachers systematically overpredict student performance across a variety of age levels and domains (Berg & Brouwer, 1991), especially with physics and chemistry content (Friedrichsen et al., 2009; Halim & Meerah, 2002). These biases may arise because expertise in the content area may prevent them from accurately judging what novices know. Expertise biases may prevent experts from creating mnemonics useful for novices. Alternatively, expertise in chemistry content may not improve mnemonic creation because chemistry knowledge and mnemonic generation may tap different knowledge and skill sets.

Method

Participants

We first recruited as many expert mnemonic generators as we could. We eventually collected mnemonics from 23 chemistry graduate students at the University of Arizona; chemistry graduate students were paid \$20 to generate mnemonic cues for the novice learners. Similarly, twenty-three undergraduate introductory educational psychology students were recruited to generate mnemonics for their peers; these participants received partial course credit for their participation. After recruiting mnemonic generators, we recruited chemistry learners. An additional 43 introductory educational psychology students at the University of Arizona participated in exchange for partial course credit. A sample size of 39 was required to detect an f effect size of 0.295 (which was found in Experiment 1) with alpha of .05 and power of .80.

Materials

The same chemistry concepts that were used in the prior experiments were used in Experiment 3.

Procedure

Experiment 3 proceeded similar to Experiment 1 with slight changes. First, cues that were fed to others were always intended for others. In Experiments 1 and 2, learners generated mnemonics to help themselves remember the concepts; these cues were then sometimes forwarded to other learners. In Experiment 3, generators intentionally produced mnemonics that would help a different learner remember the chemistry concepts. Both graduate students and novice students generated mnemonic cues for novice learners in the participant pool. These novice learners were described as "freshmen from the University of Arizona who have not taken any college chemistry courses." The 23 graduate students and 23 peers generated mnemonic cues for each of the 24 chemistry concepts one at a time. All of the mnemonic generators participated before the learners were recruited.

Second, when the learners were studying the chemistry information, chemistry concepts were randomly, but evenly, divided among three conditions: self-generated mnemonics, mnemonics generated by peers, and mnemonics generated by experts. When studying the chemistry concepts, a learner generated their own cues (without seeing anyone else's cue), received a cue that a peer from the same participant pool generated, or received a mnemonic cue that a chemistry graduate student generated. Each novice learner was linked with one chemistry graduate student and with one peer generator. No distinctions were made between cues generated by peers or graduate students. The study/generate trials were self-paced.

As in Experiment 2, learners' judgments of their own learning (JOLs) were elicited. JOLs may be more sensitive to differences between mnemonic cues than the chemistry tests, so they are included as an additional measure to test differences between conditions. For each chemistry concept, learners predicted how helpful the cue would be in remembering the chemistry concept on a scale of 1 (*not helpful at all*) to 4 (*very helpful*). Learners reported their JOLs after studying others' cues or after inputting their own cue. Beyond these three changes, the instructions, study

session, chemistry content test, and mnemonic memory test proceeded as in Experiment 1.

Results

The Cohen's Kappas of coding agreement for the test correct and mnemonic memory are shown in Table 1. The proportion of test questions answered correctly by condition are shown in Figure 3. A repeated measures ANOVA on the proportion of test answers by condition showed a significant effect of condition, $F(1, 84) = 20.37, p < .001, \eta_p^2 = .33$. Follow-up paired t -tests showed that learners correctly answered more question in the self-generated condition than in the peer-generated condition, $t(42) = 4.81, p < .001, d = 0.74$. Learners answered more questions correctly in the self-generated condition than in the expert-generated condition, $t(42) = 4.81, p < .001, d = 0.86$. No significant difference was found between the peer-generated and expert-generated conditions, $t(42) = 1.23, p = .23, d = 0.19$.

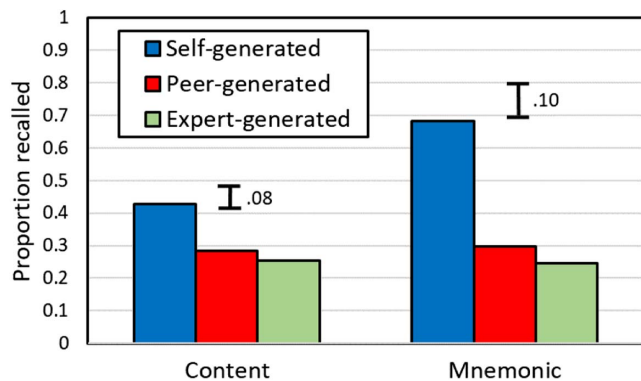
Second, we examined whether cue generation condition affected the proportion of mnemonic cues recalled, which is also displayed in Figure 3. A one-way repeated measures ANOVA on the proportion of mnemonic cues recalled revealed a significant effect of condition, $F(2, 84) = 87.35, p < .001, \eta_p^2 = .68$. Follow-up paired t -tests showed that learners remembered more self-generated mnemonics than peer-generated mnemonics, $t(42) = 10.09, p < .001, d = 1.55$, or expert-generated mnemonics, $t(42) = 11.84, p < .001, d = 1.83$. No significant difference was found between the recall of peer-generated and expert-generated mnemonics, $t(42) = 1.50, p = .14, d = 0.23$.

Next, we examined whether mnemonic condition affected learners' predictions of recall. A one-way repeated measures ANOVA on JOLs showed a significant effect of condition, $F(2, 84) = 13.91, p < .001, \eta_p^2 = .25$. As shown in Figure 4, learners predicted that self-generated cues would be more helpful for their memory than peer-generated, $t(42) = 4.95, p < .001, d = 0.76$, or expert-generated mnemonics, $t(42) = 4.98, p < .001, d = 0.77$. Learners' JOLs did not differ between peer-generated or expert-generated mnemonics, $t(42) = .36, p = .72, d = 0.06$.

Finally, we compared study times across the three conditions, as shown in Figure 5. A one-way repeated measures ANOVA on study

Figure 3

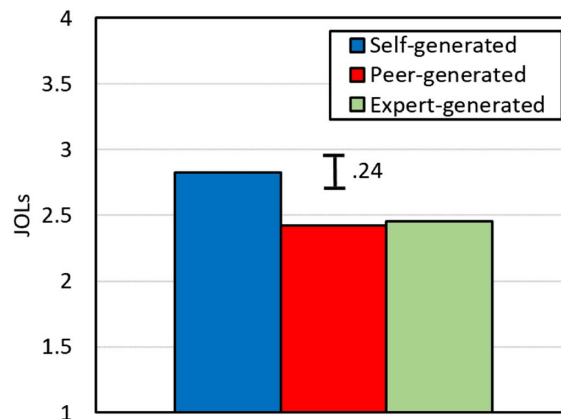
Proportion of Chemistry Content (Left) and Chemistry Mnemonics (Right) Recalled in Experiment 3



Note. See the online article for the color version of this figure.

Figure 4

Judgments of Learning (JOLs) From 1 to 4 Based Upon Condition Reported in Experiment 3



Note. See the online article for the color version of this figure.

time showed a significant effect of cue condition, $F(2, 84) = 87.97, p < .001, \eta_p^2 = .68$. Learners spent significantly more time generating cues than reading cues generated by peers, $t(42) = 9.65, p < .001, d = 1.49$, and by experts, $t(42) = 9.27, p < .001, d = 1.43$. Learners spent slightly less time reading cues generated by peers than by experts, $t(42) = 2.06, p = .046, d = 0.32$.

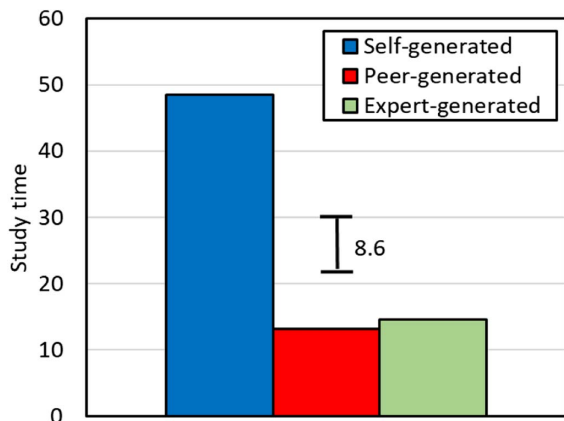
Discussion

Experiment 3 replicated the prior two experiments and extended the results to include an additional expert-generated condition. Generating one's own mnemonic cues bolstered memory for the content and mnemonic more than receiving a mnemonic that someone else generated. Recall of chemistry content and the mnemonic did not differ between cues generated by experts or by novices. The only difference found between mnemonics generated by experts and those generated by peers was the time taken by the learners to study the mnemonic; learners spent less time reading a peer's cues than an expert's cues. Having greater knowledge about the domain did not afford graduate students better abilities at creating mnemonics for novices. The evidence hints that graduate students are not able to take perspective of novice learners, as their mnemonics are equally as effective as peer-generated ones but require slightly more study time.

Experiment 4

While Experiments 1–3 showed important benefits of generating one's own mnemonic cues for later recall of chemistry content, they do not explain why mnemonic generation boosts recall. In the next two experiments, we tested whether better memory for self-generated mnemonic cues caused better memory for the associated chemistry content. To do so, we experimentally manipulated memory for the mnemonics by assigning mnemonics to differentially effective practice sessions. Half of the mnemonics were restudied and half underwent retrieval practice. Retrieval practice, in which learners practice retrieving information from long-term memory, consistently improves long-term memory for that information

Figure 5
Time Spent, in Seconds, During the Initial Generate/Read Phase-Based Upon Condition in Experiment 3



Note. See the online article for the color version of this figure.

compared to rereading (Roediger & Karpicke, 2006; Tullis et al., 2018). Prior research suggests that combining retrieval practice with cue generation can improve recall of the target information (Miyatsu & McDaniel, 2019; Pyc & Rawson, 2010; but see Karpicke & Smith, 2012). If memory for the mnemonics underlies differences between self-generated and other-generated conditions, improving memory of the mnemonics through retrieval practice should boost recall of chemistry content. Alternatively, if we boost memory for the mnemonic without affecting recall of the chemistry content, this indicates a significant limitation of the utility of mnemonic generation. We specifically used retrieval practice to target memory for the mnemonic itself (rather than the chemistry content) because we tested whether mnemonic memory causes greater recall of the associated content. Retrieval practice for the chemistry content could directly strengthen later retrieval for the content without relying on the mnemonic.

Further, we extended the retention interval to assess the longer term impact of mnemonic generation on memory. Prior research suggests that generating mnemonics can boost long-term retention compared to reading teacher-provided mnemonics (Bloom & Lamkin, 2006), but can also yield significantly greater forgetting rates than rote repetition (Wang et al., 1992). In the current research, participants generated or read mnemonics on 1 day, left the lab, and returned to the lab about 2 days later to take the final memory tests.

Method

Participants

We collected data from 52 participants in introductory Educational Psychology courses at the University of Arizona during day 1 because we could not foretell how many would return to complete the second day. Forty-three participants ultimately came back for the second day. We aimed to collect a minimum of 34 participants, which a G^* power analysis indicated would be needed to detect a medium effect (Cohen's $d = 0.5$) with .8 power and an alpha of .05 (Faul et al., 2007).

Materials

Thirty-two chemistry concepts, and corresponding questions, were derived from the guidelines of the next gen science standards.

Procedure

The procedure was very similar to that of Experiment 1, with two significant changes. First, learners practiced the mnemonic cues. After finishing the initial study/generate study block as in Experiment 1, half of self-generated and half of peer-generated mnemonics were randomly assigned to the rereading condition. For the rereading condition, the chemistry concept was displayed at the top of the screen (as during the initial study block) and the corresponding mnemonic cue was displayed in the center of the screen. Learners studied the concepts with the corresponding mnemonic cues for as long as they wanted. The other half of the concepts were assigned to the retrieval practice condition. In the retrieval practice condition, the chemistry concept was displayed at the top of the screen and learners attempted to recall the corresponding mnemonic cue from the initial study block with the instructions "What was the mnemonic cue that was associated with this concept?" Learners received no feedback about their answers. Learners spent as much time practicing retrieval and restudying each item as they wanted. Immediately following rereading or retrieval practice for each concept, learners made a JOL. For each chemistry concept, learners rated "How likely will you be able to remember the chemistry concept on a test when you come back to the lab?" on a scale of 1 (*not at all likely*) to 4 (*very likely*). After the restudy/retrieval practice block, participants left the lab. Participants returned to the lab 1–3 days later ($M = 36.7$ hr; range 21–75 hr). The chemistry and mnemonic tests proceeded as in Experiment 1.

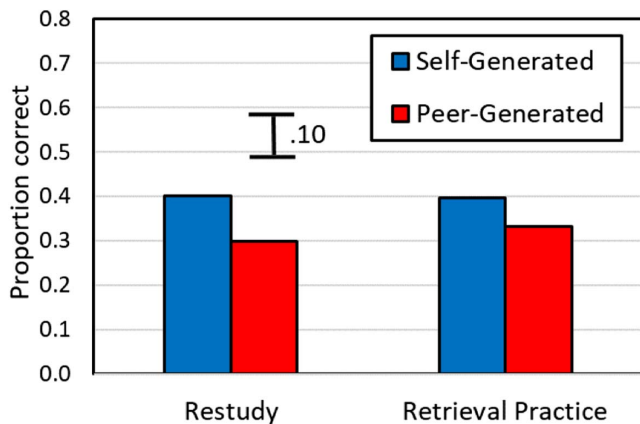
Results

The Cohen's Kappas of coding agreement for the test correct and mnemonic memory are shown in Table 1. First, we examined the initial read/generate study time based upon condition. When learners read peer-generated cues, they spent significantly less time ($M = 14.91$, $SD = 6.00$) than when they generated their own cues ($M = 48.71$, $SD = 22.97$), $t(38) = 11.12$, $p < .001$, $d = 1.80$. Next, we compared memory for the mnemonic during retrieval practice on the first day using a repeated measures t-test. Learners recalled more self-generated mnemonic cues during retrieval practice on the first day ($M = .81$, $SD = .18$) than peer-generated mnemonic cues ($M = .39$, $SD = .22$), $t(38) = 10.87$, $p < .001$, $d = 1.74$.

We compared chemistry questions correctly answered during the final test, as shown in Figure 6. A 2 (cue generator: self vs. peer) \times 2 (practice condition: restudy vs. retrieval practice) repeated measures ANOVA on recall of chemistry content showed a significant main effect of cue generator, $F(1, 38) = 14.44$, $p = .001$, $\eta_p^2 = .28$, such that content associated with self-generated mnemonics was recalled better than content with peer-generated mnemonics. Neither the interaction between generator and practice condition, $F(1, 38) = .67$, $p = .42$, $\eta_p^2 = .02$, nor the main effect of practice condition, $F(1, 38) = .46$, $p = .50$, $\eta_p^2 = .01$, reached significance.

Figure 6

The Proportion of Chemistry Content Correctly Recalled as a Function of Cue Condition and Practice Condition in Experiment 4



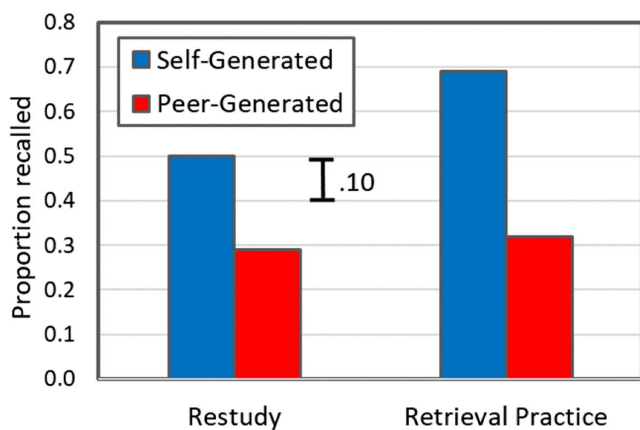
Note. See the online article for the color version of this figure.

We examined the proportion of mnemonic cues correctly recalled during the final test as shown in Figure 7. A 2 (cue generator) \times 2 (practice condition) repeated measures ANOVA showed a significant interaction, $F(1, 38) = 12.54, p = .001, \eta_p^2 = .25$, a significant main effect of cue generator, $F(1, 38) = 59.76, p < .001, \eta_p^2 = .61$, and a significant main effect of practice condition, $F(1, 38) = 18.69, p < .001, \eta_p^2 = .33$. Posthoc comparisons show that self-generated mnemonics were remembered better after retrieval practice than restudy, $t(38) = 5.19, p < .001, d = 0.84$. Memory for peer-generated mnemonics did not differ based upon the kind of practice they received, $t(38) = .72, p = .48, d = 0.12$.

We also examined learners' JOLs during the first day as shown in Figure 8. A 2 (cue generator) \times 2 (practice condition) repeated measures ANOVA showed a significant interaction, $F(1, 38) = 13.59, p < .001, \eta_p^2 = .26$, a significant main effect of cue generator, $F(1, 38) = 21.16, p < .001, \eta_p^2 = .36$, and a significant main effect of practice condition, $F(1, 38) = 5.40, p = .03, \eta_p^2 = .12$.

Figure 7

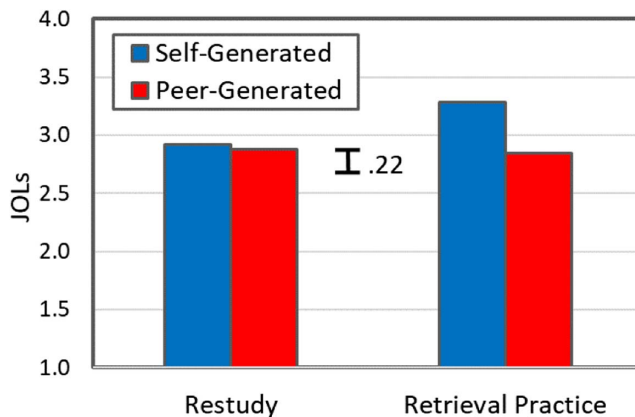
The Proportion of Mnemonic Cues Recalled as a Function of Cue Condition and Practice Condition in Experiment 4



Note. See the online article for the color version of this figure.

Figure 8

Learners' JOLs as a Function of Cue Condition and Practice Condition in Experiment 4



Note. See the online article for the color version of this figure.

Posthoc follow up repeated measures *t*-tests show that JOLs in the restudy condition did not differ between self- and peer-generated conditions, $t(38) = .63, p = .53, d = 0.10$, but, under retrieval practice, JOLs were significantly greater for self- than other-generated mnemonics, $t(38) = 5.33, p < .001, d = 0.86$.

Discussion

Generating one's own mnemonic cues enhanced memory for both the chemistry content and the mnemonic cue itself, as in the prior three experiments. Even with a retention interval of 2 days, recall of content and mnemonics in the self-generated condition was better than that in the peer-generated condition. Practicing retrieval of the mnemonics increased recall of the mnemonics for the self-generated cues but did not ultimately affect recall of chemistry content in either condition. These data provide strong evidence that memory for the mnemonic does not directly change memory for the chemistry content. Before interpreting these results, we replicated them in Experiment 5.

Finally, learners' predictions of their test performance (i.e., their JOLs) differentiated between self-generated and other-generated mnemonics under retrieval practice, but not under rereading conditions. Retrieval practice may provide clear diagnostic cues about the memorability of the mnemonics (i.e., successful or unsuccessful practice recall) that rereading does not. Learners likely predict better test performance for self-generated mnemonics because they recalled more self-generated mnemonics than peer-generated mnemonics during this practice test. JOLs track final recall of the mnemonics (but not final recall of the chemistry content) closely.

Experiment 5

In this final experiment, we sought to replicate Experiment 4 while modifying the mnemonic recall test conditions. In Experiment 4, learners' memory for mnemonics was tested while the chemistry concept was present. This procedure may allow learners to reconstruct or recreate the mnemonic from the visibly present chemistry concept, rather than retrieve it from memory. In the current

Experiment 5, learners' memory for the mnemonics was tested without showing the full chemistry concept. This procedure ensures that learners are recalling the mnemonic from memory, rather than recreating it from the concept itself. Recalling the chemistry mnemonic without the full chemistry concept present is a stricter test of mnemonic recall and better reflects actual test conditions (when the chemistry concept is not present).

Method

Participants

Forty-four participants in introductory educational psychology courses at the University of Arizona participated in day 1 study for partial course credit. Forty participants ultimately came back for day 2 study.

Materials

The same materials were used as in Experiment 4.

Procedure

The procedure occurred similar to that in Experiment 4, with two significant test changes. During the test on the second day, participants first answered each chemistry content question (e.g., "What happens in a covalent bond?"). Then, learners immediately answered the question "Did you use the mnemonic cue associated with this concept to help answer the question? Yes or No." Finally, regardless of whether they reported using the mnemonic or not, we asked them to recall the associated mnemonic. As learners tried to recall the corresponding mnemonic, the chemistry content question (e.g., "What happens in a covalent bond?") was still displayed on the screen. In prior experiments, the entire chemistry concept (e.g., "In covalent bonds, electrons are shared between atoms.") was displayed when learners attempted to retrieve the corresponding mnemonic. Here, we only showed the chemistry question and did not provide the answer. Withholding the full chemistry concept reduces the possibility that learners regenerate the mnemonic from the chemistry concept itself.

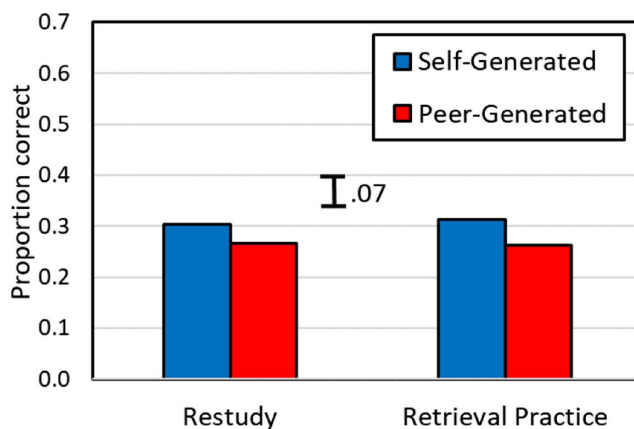
Results

The Cohen's Kappas of coding agreement for the test correct and mnemonic memory are shown in Table 1. First, we examined the initial study time based upon condition. When learners read peer-generated cues, they spent significantly less time ($M = 14.36$, $SD = 7.12$) than when they generated their own cues ($M = 50.17$, $SD = 20.74$), $t(39) = 12.13$, $p < .001$, $d = 1.94$. Next, we compared memory for the mnemonic cue during retrieval practice on the first day using a repeated measures t -test. Learners recalled more self-generated mnemonic cues on the first day ($M = .81$, $SD = .25$) than peer-generated mnemonic cues ($M = .39$, $SD = .23$), $t(39) = 9.77$, $p < .001$, $d = 1.56$.

Next, we compared the proportion of chemistry questions answered correctly during the final test by condition, as shown in Figure 9. A 2 (cue generator: self vs. peer) \times 2 (practice condition: restudy vs. retrieval practice) repeated measures ANOVA revealed a significant main effect of cue generator, $F(1, 39) = 4.89$, $p = .03$, $\eta_p^2 = .11$. Neither the interaction between generator and condition,

Figure 9

The Proportion of Chemistry Content Correctly Recalled as a Function of Cue Condition and Practice Condition in Experiment 5



Note. See the online article for the color version of this figure.

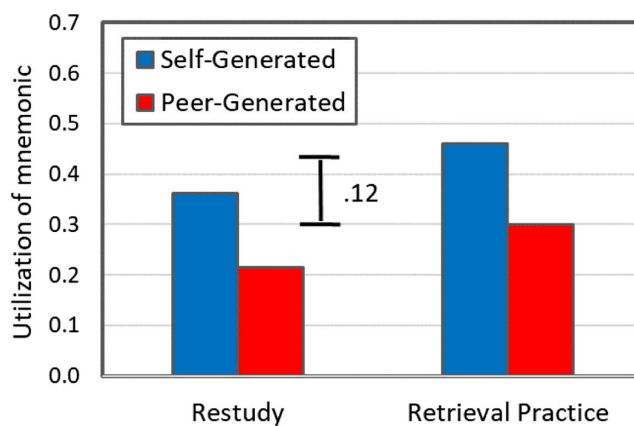
$F(1, 39) = .11$, $p = .75$, $\eta_p^2 = .003$, nor the main effect of practice condition, $F(1, 39) = .02$, $p = .88$, $\eta_p^2 = .001$, impacted the proportion of chemistry content recalled.

We analyzed the proportion of instances in which learners reported using the mnemonic cues to help recall the chemistry information the final test as shown in Figure 10. A 2 (cue generator) \times 2 (practice condition) repeated measures ANOVA showed a significant main effect of cue generator, $F(1, 39) = 25.02$, $p < .001$, $\eta_p^2 = .39$, and a significant main effect of practice condition, $F(1, 39) = 14.63$, $p < .001$, $\eta_p^2 = .27$. The interaction between generator and condition did not impact the reported use of mnemonic cues, $F(1, 39) = .06$, $p = .80$, $\eta_p^2 = .002$.

Next, we examined the proportion of mnemonic cues correctly recalled during the final test as shown in Figure 11. A 2 (cue generator) \times 2 (practice condition) repeated measures ANOVA

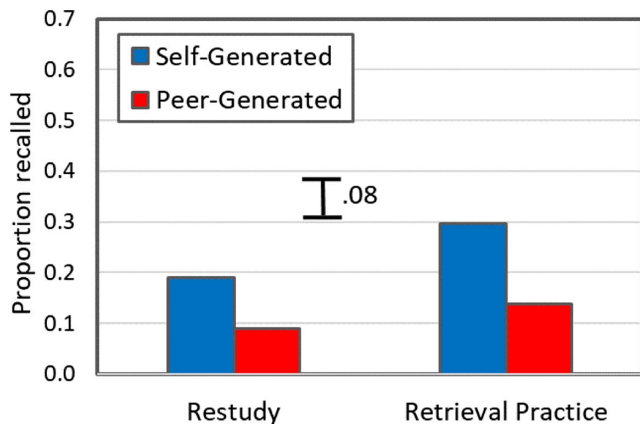
Figure 10

The Proportion of Mnemonics That Participants Say They Utilized to Answer the Chemistry Content Questions as a Function of Cue Condition and Practice Condition in Experiment 5



Note. See the online article for the color version of this figure.

Figure 11
The Proportion of Mnemonic Cues Recalled as a Function of Cue Condition and Practice Condition in Experiment 5



Note. See the online article for the color version of this figure.

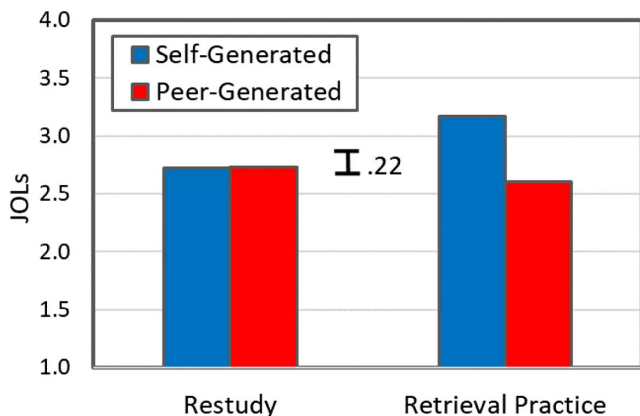
showed a significant main effect of cue generator, $F(1, 39) = 20.31, p < .001, \eta_p^2 = .34$, and a significant main effect of practice condition, $F(1, 39) = 13.69, p < .001, \eta_p^2 = .26$. Mnemonic recall was not significantly impacted by the interaction of generator and condition, $F(1, 39) = 2.17, p = .15, \eta_p^2 = .05$.

Finally, we examined learners' JOLs during the first day as shown in Figure 12. A 2 (cue generator) \times 2 (practice condition) repeated measures ANOVA showed a significant interaction, $F(1, 39) = 27.47, p < .001, \eta_p^2 = .41$, a significant main effect of cue generator, $F(1, 39) = 20.27, p < .001, \eta_p^2 = .34$, and a significant main effect of practice condition, $F(1, 39) = 5.49, p = .02, \eta_p^2 = .12$.

Discussion

Consistent with the prior experiments, generating one's own mnemonics boosted memory for both the chemistry content and the mnemonics compared to receiving mnemonics generated by

Figure 12
Learners' JOLs as a Function of Cue Condition and Practice Condition in Experiment 5



Note. See the online article for the color version of this figure.

peers. Practice retrieving the mnemonic improved memory for the mnemonic but did not affect memory for the chemistry content. Retrieval practice also boosted self-reported utilization of the mnemonic cue. Despite the increased reported use and recall of the mnemonic, practice retrieving the mnemonic did not affect recall of the chemistry content, which replicates results from Experiment 4. These data suggest that benefits of self-generated mnemonics on remembering chemistry content do not arise solely because generation improves memory for the mnemonic. Instead, these results suggest that a significant obstacle to the effectiveness of self-generated mnemonics is their decodability (i.e., the ability to recall the appropriate target information from the given mnemonic). If mnemonics were decodable, increasing recall of mnemonics would increase recall of the corresponding chemistry content.

So, if memory for the mnemonic themselves does not cause improvements, why does generating mnemonics boost recall for the associated chemistry content? There are several possibilities. Learners spend a greater amount of time encoding when required to create their own mnemonic cues. Additional study time itself may allow stronger encoding and underlie this difference. However, Experiment 2 Likely rules out this explanation because equating study time between conditions does not reduce the differences between self-generated and other-generated mnemonics. A second explanation is that, when learners try to generate mnemonics, they deeply process the target information. Even if they do not successfully create an effective or decodable mnemonic, learners may try to identify important information, restructure the information, connect it to their own experiences, and transform it into something more memorable. These attempts at creating mnemonics may entail deep processing that can help recall the content even without the mnemonic. The process of generating the mnemonic (rather than the mnemonic product) may be useful for promoting long-term retention of the information. For example, learners may even remember some aspects of the cue generation process which could boost recall of the chemistry content.

Relationship Between Content and Mnemonic Recall

Finally, as a measure of mnemonic decodability, we examined the relationship between ability to recall each mnemonic and the associated chemistry content. To do so, we computed the gamma correlation between recall of the mnemonic (0 or 1) and recall of the corresponding chemistry concept (0 or 1) for self-generated and other-generated mnemonics across the five experiments. The average within-participant gamma correlations are shown in Table 3. These correlations reflect how decodable the mnemonic cues are because they track the correspondence between memory for the mnemonic and content. The results show no significant differences (and small differences) between the decodability of self-generated and other-generated mnemonics. Generating a mnemonic, then, does not guarantee a significantly greater ability to decode that mnemonic than receiving a mnemonic.

General Discussion

Generating one's own chemistry mnemonics boosted recall of the corresponding chemistry content and the mnemonic cues when compared to receiving someone else's mnemonics. These results consistently replicate across five experiments; reproducibility is

Table 3

Within Participant Gamma Correlations (and Standard Deviations) Between Recall of Each Chemistry Concept and Recall of the Associated Mnemonic, Which Reflects Cue Decodability Gamma Correlations Can Only Be Calculated If Participants Have Some Variability in Each of the Two Measures, So Averages and Repeated Measure t-Tests Are Computed Only for Participants That Can Provide Data in Both Conditions

Experiment	Self-generated	Other-generated	Inferential Statistics
Experiment 1	.29 (.76)	.22 (.65)	$t(19) = .71, p = .49, d = 0.09$
Experiment 2	-.13 (.72)	.05 (.31)	$t(29) = 1.14, p = .26, d = -0.22$
Experiment 3	.30 (.90)	-.16 (.86)	$t(27) = 1.78, p = .09, d = 0.34$
Experiment 4	.17 (.66)	.15 (.72)	$t(31) = .09, p = .92, d = 0.02$
Experiment 5	.67 (.49)	.59 (.62)	$t(21) = .46, p = .65, d = 0.10$
All participants	.24 (.77)	.15 (.70)	$t(128) = .98, p = .33, d = 0.09$

vital for psychological science (Camerer et al., 2018; Zwaan et al., 2017) and may be particularly important for novel findings (Pashler & Harris, 2012; Roediger, 2012; Schimmack, 2012; Simons, 2014). Self-generated mnemonics may be remembered better than other-generated mnemonics either because (a) the participant generates them and generation is mnemonically beneficial or (b) self-generated mnemonics are rooted in personal schema (and self-reference is beneficial for memory). Creating one's own mnemonics also consistently boosted recall of the associated chemistry content. Our data suggest that generating one's own mnemonics may engender deep processing of the associated content, including selecting information, organizing information, connecting it to existing knowledge, and potentially transforming the information. However, the data across our five experiments show that generating one's own mnemonics boosts recall of the mnemonics, but does not enable greater decoding of that mnemonic. When we experimentally manipulated memory for the mnemonics, recall of corresponding chemistry content was not affected. Increasing recall of the mnemonic without increasing recall of the corresponding chemistry content indicates that learners fail to decode their remembered mnemonics.

Creating mnemonics that are memorable and decodable for complex and abstract real-world chemistry information may be very difficult. Much prior research has examined the efficacy of self-generated mnemonics using simple mediators (i.e., single words or images) with simple to-be-learned materials (e.g., foreign language translations: Bellezza & Poplawski, 1974; Jamieson & Schimpf, 1980; Kuo & Hooper, 2004; Saber & Johnson, 2008). Effective chemistry mnemonics need to connect the multifaceted cueing information (i.e., the question) with the complex target information (i.e., the answer). Improvements in recall of the mnemonic without improvements in chemistry content recall suggests that decodability (or the link from the mnemonic to the target information) limits the effectiveness of self-generated mnemonics for complex chemistry content. Self-generated mnemonics do not always contain enough information to recall the chemistry target information. For example, when a learner generates a mnemonic to remember that "ionic bonds involve a transfer of electrons from one atom to another," one participant generated "The word 'ionic' sounds like 'Sonic' which is the name of my aunt's dog and Sonic doesn't share, he just TAKES toys from other dogs." This is a mnemonic that ties into their personal knowledge and may be easily remembered. However, the mnemonic does not specify what ionic

bonds are taking, so they do not fully connect the query with all of the important target information. Further, some kinds of information may be more conducive to creating effective mnemonics than others. Our materials comprised a variety of different kind of to-be-remembered facts, including some lists ("Electrons fill up orbitals in the sequence of s, p, d, f, g, h, i") and some definitions ("Valence electrons are in the outermost shell of an atom and bond with other atoms"). Future research can evaluate whether different kinds of stimuli benefit from mnemonic generation equally.

Retrieval practice of the mnemonic improved recall of the mnemonic but did not affect recall of the associated chemistry content. Retrieval practice consistently boosts recall of the recalled content (e.g., Roediger & Karpicke, 2006), but the impact of retrieval practice on memory for related information is debated (Butler, 2010; McDaniel et al., 2007; Pan & Rickard, 2017; Rohrer et al., 2010; Tran et al., 2015). Our data suggest that practice recall of self-generated (or other-generated) mnemonics does not yield mnemonic benefits for the related chemistry content. Retrieval practice of the mnemonic may fail to improve memory for the content because learners struggle to appropriately decode the mnemonics.

Self-generated mnemonics elicited greater predictions of learning than other-generated mnemonics. In Experiments 2 and 3, learners made greater mnemonic predictions when they generated their own mnemonics than when they received a mnemonic generated by someone else. The advantage for self-generated over other-generated mnemonics is somewhat unexpected because generating a mnemonic likely requires more cognitive effort (and more time) than reading a prior participant's mnemonic, and learners typically view cognitive effort as indicating poor learning (Miele et al., 2011; but see Peng & Tullis, 2020). However, if learners struggle to understand a prior participant's mnemonic, the effort needed to understand a prior participant's mnemonic (or the failure to understand it) would decrease predictions. Alternatively, learners could value constructing their own mnemonics because they can connect to-be-remembered information with their existing schema. The ability to connect new information with one's existing schema may drive increased metacognitive predictions (e.g., Koriat, 1997).

In Experiments 4 and 5, learners predicted greater learning for self-generated mnemonics over other-generated mnemonics under retrieval practice but showed little difference between conditions during restudy. Retrieval practice reveals differences in recallability of the mnemonic that restudying cannot, and learners use outcomes

of practice retrieval to guide predictions of future performance (e.g., Nelson & Dunlosky, 1991; Tullis et al., 2013). Consequently, retrieval practice (in which learners recall many more self-generated mnemonics than other-generated mnemonics) exacerbates differences between conditions. Restudying a mnemonic prevents the learner from trying to retrieve the mnemonic [and retrieval is often a diagnostic cue about future recallability (Koriat, 1997)]; thus, restudying may mask the differences between self-generated and other-generated mnemonics (Dunlosky & Nelson, 1992).

We designed five experiments to assess within-participants differences between self- and other-generated conditions in order to increase our power to detect differences. Some research suggests that experimental design impacts the benefits that generation provides. More specifically, between-participants research designs weaken (but do not eliminate) the advantages of generation over reading when compared to within-participants research designs (Bertsch et al., 2007). Experimental design impacts the benefits of generation because generation strengthens item processing at the expense of relational (or between-item) processing (McDaniel & Bugg, 2008). Relational processing guides free recall; consequently, generation across an entire list of items (which impairs relational processing) can impair the free recall of generated items (McDaniel & Bugg, 2008). Relational processing across different chemistry facts is likely not crucial to succeeding on the current task (or on real-world tests of chemistry content) because our tests do not rely upon connecting between different chemistry facts. We anticipate that the benefits of self-generated mnemonics would persist in between-participants designs (because relational processing is not vital to success on our test), but future research can specifically test this question.

Extensive practice generating mnemonics or specific guidance about how to generate effective mnemonics could prompt learners to generate mnemonics with improved decodability (and ultimately usefulness). For example, extended practice using the method of loci can improve the method of loci's efficacy (Baltes & Kliegl, 1992). Similarly, people who compete in memory competitions repeatedly practice using memory cues to succeed at daunting memory challenges (Foer, 2012). Giving more specific instructions to learners about how to generate effective mnemonics should also improve the quality of self-generated cues. Asking learners to generate "focused" cues (i.e., ones that they would likely generate again) prompts them to produce cues that they are more likely to regenerate and yields more effective cues (Mäntylä & Nilsson, 1988). Future research can investigate whether instructing learners to generate cues that connect to the target information (to improve decodability) would improve later recall of complex chemistry information.

Individual differences may also play a role in who benefits from self-generating cues. Our sample comprised largely novice learners, who have been shown to benefit from creating mnemonics for abstract information (McCabe et al., 2013). Additional research suggests that requiring self-generated memory cues is particularly beneficial for students with learning disabilities (Mastropieri et al., 2000) and for low-performing students, who spontaneously generate fewer memory cues than others do (Scruggs et al., 1986). However, low-performing individuals may struggle to create effective mnemonics for complex information. Understanding the impact of mnemonic generation for a wider variety of learners will be crucial to supporting students in and out of the classroom.

If learners can be prompted to generate effective mnemonic cues, the advantages of self-generated memory cues extend beyond memory tasks. Mnemonics can boost students' use and comprehension of new vocabulary (McDaniel & Pressley, 1989; Pressley et al., 1981), help students to integrate information and make inferences (Levin & Levin, 1990), and allow students to solve higher order thinking problems (Richmond et al., 2011). Students can even use mnemonics to apply and manipulate factual biology and psychology knowledge (Rosenheck et al., 1989). While our experiments have only tested *recall* of the target chemistry information, self-generated mnemonics have the potential to benefit application of that knowledge, problem-solving, and inferences.

Learning increasingly occurs in contexts where students must self-regulate their study, including out-of-class reading, homework, and distance learning (Bjork et al., 2013; Tullis, 2020). Effective learning requires good decisions about how, when, and what to study (Metcalfe, 2009). When students make good study decisions, learning is substantially enhanced (Thiede et al., 2003; Tullis & Benjamin, 2012). The strategic control that people exercise over their encoding and retrieval processes largely contributes to individual differences in memory capabilities (Benjamin, 2008). Substantial research has examined metacognitive control of *what* learners study (Kornell & Metcalfe, 2006; Tullis & Benjamin, 2012), *when* they study (Benjamin & Bird, 2006; Toppino et al., 2009), and even *which* strategies they use to practice information (e.g., self-testing: Karpicke & Roediger, 2008; Tullis et al., 2018). But, little metacognitive research has examined how effectively learners use strategies to *encode* new information, which is crucially important for how well they remember that information (Craik & Lockhart, 1972). Understanding when and how successfully students use encoding strategies—like generating mnemonic cues—may enable us to support students so that their learning is more efficient and effective.

References

- Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, 1, 126–133.
- Balch, W. R. (2005). Elaborations of introductory psychology terms: Effects on test performance and subjective ratings. *Teaching of Psychology*, 32, 29–34.
- Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology*, 28(1), 121–125.
- Banks, B. B. (1941). Chemical mnemonic devices. *Journal of Chemical Education*, 18, Article 97.
- Battino, R. (1992). On the importance of rote learning. *Journal of Chemical Education*, 69(2), 135–137.
- Bellezza, F. S., & Poplawski, A. J. (1974). The function of one-word mediators in the recall of word pairs. *Memory & Cognition*, 2, 447–452.
- Benjamin, A. S. (2008). Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48, pp. 175–223). Academic Press.
- Benjamin, A. S., & Bird, R. D. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language*, 55, 126–137.
- Berg, T., & Brouwer, W. (1991). Teacher awareness of student alternate conceptions about rotational motion and gravity. *Journal of Research in Science Teaching*, 28, 3–18.

- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*, 201–210.
- Bloom, C. M., & Lamkin, D. M. (2006). The Olympian struggle to remember the cranial nerves: Mnemonics and student success. *Teaching of Psychology*, *33*(2), 128–129.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444.
- Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist*, *58*, 496–510.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmeld, A., Buttrick, N., Chan, T. Z., Chen, Y. L., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, *2*, 637–644.
- Carlson, R. F., Kincaid, J. P., Lance, S., & Hodgson, T. (1976). Spontaneous use of mnemonics and grade point average. *The Journal of Psychology: Interdisciplinary and Applied*, *92*, 117–122.
- Carney, R. N., & Levin, J. R. (2003). Promoting higher-order learning benefits by building lower order mnemonic connections. *Applied Cognitive Psychology*, *17*, 563–575.
- Chittleborough, G., Treagust, D. F., & Mocerino, M. (2005). Non-major chemistry students' learning strategies: Explaining their choice and examining the implications for teaching and learning. *Science Education International*, *16*, 5–22.
- Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671–684.
- Crutcher, R. J., & Ericsson, K. A. (2000). The role of mediators in memory retrieval as a function of practice: Controlled mediation to direct access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1297–1317.
- Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, *41*, 389–400.
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374–380.
- 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*(2), 175–191.
- Foer, J. (2012). *Moonwalking with Einstein: The art and science of remembering everything*. Penguin.
- Fraundorf, S. H., Diaz, M. I., Finley, J. R., Lewis, M. L., Tooley, K. M., Isaacs, A. M., Lam, T. Q., Trude, A. M., Brown-Schmidt, S., & Brehm, L. (2014). *CogToolbox for MATLAB*. <http://www.scottfraundorf.com/cogtoolbox.html>
- Friedrichsen, P. J., Abell, S. K., Pareja, E. M., Brown, P. L., Lankford, D. M., & Volkman, M. J. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching*, *46*, 357–383.
- Gabel, D. (1999). Improving teaching and learning through chemistry education research: A look to the future. *Journal of Chemical Education*, *76*(4), 548–554.
- Gough, W. (1977). The use of mnemonics in mathematics and physics. *Physics Education*, *12*(6), 385–386.
- Halim, L., & Meerah, S. M. M. (2002). Science trainee teachers' pedagogical content knowledge and its influence on physics teaching. *Research in Science & Technological Education*, *20*, 215–225.
- Harris, J. E. (1980). Memory aids people use: Two interview studies. *Memory & Cognition*, *8*, 31–38.
- Jamieson, D. G., & Schimpf, M. G. (1980). Self-generated images are more effective mnemonics. *Journal of Mental Imagery*, *4*, 25–33.
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.
- Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*(1), 17–29.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 609–622.
- Kuo, M. L. A., & Hooper, S. (2004). The effects of visual and verbal coding mnemonics on learning Chinese characters in computer-based instruction. *Educational Technology Research and Development*, *52*(3), 23–38. <https://doi.org/10.1007/BF02504673>
- Levin, M. E., & Levin, J. R. (1990). Scientific mnemonics: Methods for maximizing more than memory. *American Educational Research Journal*, *27*, 301–321.
- Levin, J. R., Kessler Berry, J., Miller, G. E., & Bartell, N. P. (1982). More on how (and how not) to remember the states and their capitals. *The Elementary School Journal*, *82*(4), 379–388.
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490.
- Määntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 66–71.
- Määntylä, T., & Nilsson, L. G. (1988). Cue distinctiveness and forgetting: Effectiveness of self-generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 502–509.
- Mastropieri, M. A., Sweda, J., & Scruggs, T. E. (2000). Putting mnemonic strategies to work in an inclusive classroom. *Learning Disabilities Research & Practice*, *15*(2), 69–74.
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159–163.
- McCabe, J. A., Osha, K. L., Roche, J. A., & Susser, J. A. (2013). Psychology students' knowledge and use of mnemonics. *Teaching of Psychology*, *40*(3), 183–192.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4–5), 494–513.
- McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and unifying explanation. *Psychonomic Bulletin & Review*, *15*, 237–255.
- McDaniel, M. A., & Pressley, M. (1989). Keyword and context instruction of new vocabulary meanings: Effects on text comprehension and memory. *Journal of Educational Psychology*, *81*, 204–221.
- McFarland, C. E., Jr., Frey, T. J., & Rhodes, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning & Verbal Behavior*, *19*, 210–225.
- Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered? It depends on your beliefs about intelligence. *Psychological Science*, *22*, 320–324. <https://doi.org/10.1177/0956797610397954>

- Miyatsu, T., & McDaniel, M. A. (2019). Adding the keyword mnemonic to retrieval practice: A potent combination for foreign language vocabulary learning? *Memory & Cognition*, *47*, 1328–1343.
- Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*, 389–395.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science*, *2*, 267–270.
- Next Generation Science Standards Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.
- Pan, S. C., & Rickard, T. C. (2017). Does retrieval practice enhance learning and transfer relative to restudy for term-definition facts? *Journal of Experimental Psychology: Applied*, *23*(3), 278–292.
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*, 531–536. <https://doi.org/10.1177/1745691612463401>
- Patten, B. M. (1990). The history of memory arts. *Neurology*, *40*, 346–352. <https://doi.org/10.1212/WNL.40.2.346>
- Peng, Y., & Tullis, J. G. (2020). Theories of intelligence influence self-regulated study choices and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(3), 487–496.
- Pressley, M., Levin, J. R., & Miller, G. E. (1981). How does the keyword method affect vocabulary comprehension and usage? *Reading Research Quarterly*, *16*, 213–226.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*, Article 335.
- Richmond, A. S., Carney, R. N., & Levin, J. R. (2011). Got neurons? Teaching neuroscience mnemonically promotes retention and higher-order thinking. *Psychology Learning & Teaching*, *10*, 40–45.
- Roediger, H. L. (2012). Psychology's woes and a partial cure: The value of replication. *Observer*, *25*(2), 27–29.
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 233–239.
- Rosenheck, M. B., Levin, M. E., & Levin, J. R. (1989). Learning botany concepts mnemonically: Seeing the forest and the trees. *Journal of Educational Psychology*, *81*(2), 196–203.
- Saber, J. L., & Johnson, R. D. (2008). Don't throw out the baby with the bathwater: Verbal repetition, mnemonics, and active learning. *Journal of Marketing Education*, *30*, 207–216.
- Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple study articles. *Psychological Methods*, *17*, 551–566.
- Scruggs, T. E., Mastropieri, M. A., Jorgensen, C., & Monson, J. (1986). Effective mnemonic strategies for gifted learners. *Journal for the Education of the Gifted*, *9*, 105–121.
- Shriberg, L. K. (1982). *Comparison of two mnemonic encoding strategies in children's recognition and recall of abstract prose information* [Unpublished doctoral dissertation]. University of Wisconsin.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76–80.
- Stagg, B. C., & Donkin, M. E. (2016). Mnemonics are an effective tool for adult beginners learning plant identification. *Journal of Biological Education*, *50*(1), 24–40.
- Stephens, C. E. (2010). A simple mnemonic for tautomerization mechanisms in organic chemistry. *Journal of Chemical Education*, *87*, 1186–1187.
- Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394.
- Thiede, K. W., Anderson, M. C., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*, 66–73.
- Toppino, T. C., Cohen, M. S., Davis, M. L., & Moors, A. C. (2009). Metacognitive control over the distribution of practice: When is spacing preferred? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1352–1358.
- Tran, R., Rohrer, D., & Pashler, H. (2015). Retrieval practice: The lack of transfer to deductive inferences. *Psychonomic Bulletin & Review*, *22*, 135–140.
- Tullis, J. G. (2020). *E-learning: The opportunities and challenges of online instruction*. Routledge Encyclopedia of Education.
- Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, *64*, 109–118.
- Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review*, *19*, 743–749.
- Tullis, J. G., & Benjamin, A. S. (2015a). Cue generation: How learners flexibly support future retrieval. *Memory & Cognition*, *43*, 922–938.
- Tullis, J. G., & Benjamin, A. S. (2015b). Cuing others' memories. *Memory & Cognition*, *43*, 634–646.
- Tullis, J. G., Fiechter, J. L., & Benjamin, A. S. (2018). The efficacy of learners' testing choices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 540–552.
- Tullis, J. G., & Finley, J. R. (2018). Self-generated memory cues: Effective tools for learning, training, and remembering. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 179–186.
- Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition*, *41*, 429–442.
- Tullis, J. G., & Fraundorf, S. H. (2017). Predicting others' memory performance: The accuracy and bases of social metacognition. *Journal of Memory and Language*, *95*, 124–137.
- Tullis, J. G., & Maddox, G. B. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition and Learning*, *15*, 129–154.
- Tullis, J. G., & Qiu, J. (2021). *Chemistry mnemonics*. <https://osf.io/6z2xu/>
- Van Etten, S., Freebern, G., & Pressley, M. (1997). College students' beliefs about exam preparation. *Contemporary Educational Psychology*, *22*, 192–212.
- Wang, A. Y., Thomas, M. H., & Ouellette, J. A. (1992). Keyword mnemonic and retention of second-language vocabulary words. *Journal of Educational Psychology*, *84*(4), 520–528.
- Worthen, J. B., & Hunt, R. R. (2011). *Mnemonology: Mnemonics for the 21st century*. Taylor & Francis.
- Zhang, D., & Tullis, J. G. (2021). Self-generated reminders boost memory more than normative ones. *Memory & Cognition*. Advance online publication. <https://doi.org/10.3758/s13421-020-01120-7>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2017). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, 1–50.

Received August 3, 2020

Revision received January 4, 2021

Accepted January 5, 2021 ■