Original Article

QJEP

Where and how to learn: The interactive benefits of contextual variation, restudying, and retrieval practice for learning

Quarterly Journal of Experimental Psychology
1–12

DOI: 10.1177/1747021820968483
qjep.sagepub.com
SAGE

Megan N Imundo, Steven C Pan, Elizabeth Ligon Bjork and Robert A Bjork

Abstract
Students are often advised to do all of their studying in one good place, but restudying to-be-learned material in a new context can enhance subsequent recall. We examined whether there are similar benefits for testing. In Experiment 1 ($n = 106$), participants studied a 36-word list and 48 hr later—when back in the same or a new context—either restudied or recalled the list without feedback. After another 48 hr, all participants free-recalled the list in a new context. Experiment 2 ($n = 203$) differed by having the testing-condition participants restudy the list before being tested. Across both experiments, testing in a new context reduced recall, which carried over to the final test, whereas restudying in a new context did not impair (and in Experiment 2, significantly enhanced) recall. These findings reveal critical interactions between contextual-variation and retrieval-practice effects, which we interpret as consistent with a distribution-of-memory-strengths framework.

Keywords
Context; contextual variation; retrieval practice; test-enhanced learning; bifurcation model

Received: 26 May 2020; revised: 13 August 2020; accepted: 3 September 2020

"Try to study in the same place at the same time every day" (Rasmussen, 2019). That popular recommendation (e.g., Hopper, 2003; Joubert, 2018) implies that learning is optimal in a fixed environmental context. Research has demonstrated, however, that varying study contexts yields substantial learning benefits. Smith et al. (1978; Experiment 1), for example, demonstrated that studying a word list across two sessions in two physically different contexts yielded better recall than if both sessions were in identical contexts. That study has been cited over 800 times and suggests a remarkable property of human memory: *contextual variation* enhances learning (Glenberg, 1979; Smith & Rothkopf, 1984; Smith & Vela, 2001).

Prominent accounts of context effects on learning focus on incidental encoding of environmental cues. Given the cue dependency of memory, encoding under varied contexts—which presumably include more diverse cues than non-varied contexts—yields increased cue support that enhances recall (Tulving, 1983; Tulving & Pearlstone, 1966). More varied environmental cues across encoding contexts also increases the likelihood that some of those cues will be present to assist later retrieval in a different context. Relatedly, everyday memory processes often occur in a variety of different contexts; thus, the human brain may have evolved to capitalise on the varied cues that are available when, for example, information is encoded in one location but needs to be retrieved in another.

Research on contextual variation, however, has not consistently revealed memory benefits (e.g., Eich, 1985; Godden & Baddeley, 1980; Pessin, 1932). Such results suggest two moderating factors. First, benefits may be reduced when participants can easily mentally reinstate the

Department of Psychology, University of California, Los Angeles, Los Angeles, CA, USA

Corresponding author:
Megan N Imundo, Department of Psychology, University of California, Los Angeles, 502 Portola Plaza, 1285 Pritzker Hall, Box 951563, Los Angeles, CA 90095-1563, USA.
Email: meganimundo@g.ucla.edu

initial learning context (Smith, 1979). The literature on *context reinstatement* (e.g., Godden & Baddeley, 1975) indicates that having the same contextual information present during learning and at test enhances recall (although the effect has in some cases failed to materialise, for example, Fernandez & Glenberg, 1985; Saufley et al., 1985). Second, benefits may be reduced when learners focus on information intrinsic to the to-be-learned materials at the expense of environmental cues during encoding (Matzel et al., 1985) or retrieval (Smith, 1994).

In addition, the learning activity in which one engages has rarely been considered in contextual variation research. Nearly all prior studies have used repeated studying (*restudying*), but an alternate method, practice testing (*retrieval practice*), is substantially more effective (e.g., Bjork, 1975; Roediger & Butler, 2011) and is widely endorsed by researchers over restudying (e.g., Dunlosky et al., 2013; Pan & Bjork, in press). Whether the benefits of retrieval practice (i.e., the *testing effect*) interact with the effects of contextual variation has, however, been largely unexplored (cf. Smith & Handy, 2014). Several studies have shown that retrieval practice does not need to occur in the same location or use the same format as a final test in order for a testing effect to manifest (e.g., Orr & Foster, 2013; Wellington et al., 2015; see also Pan & Rickard, 2018), suggesting that the testing effect is robust to some types of contextual variation. Furthermore, the evidence as to the effects of retrieval practice on the encoding of contextual features is currently mixed: Enhanced memory for spatial context (e.g., Akan et al., 2018) has been observed, but memory for stimulus source (e.g., speaker voice as in Brewer et al., 2010) or font colour information (e.g., Hong et al., 2019) does not appear to be enhanced.

Beyond investigating the role of learning activity, no contextual variation studies have featured multi-day intervals as is common in numerous learning circumstances (e.g., sessions were only separated by 3 hr in Smith et al., 1978). If benefits of contextual variation dissipate quickly, then that would constitute a serious limitation. Conversely, if other learning tasks—most notably, studying versus retrieval practice—amplify its benefits, then that would increase its pedagogical utility. The present research addressed these critical issues.

Experiment 1

Over two sessions separated by 2 days, participants learned a word list using retrieval practice or restudy in identical or varied environmental contexts. Two days after the second session, memory for the word list was assessed on a final recall test administered in a novel context.

Method

Participants. We recruited 112 participants from the psychology participant pool at a large research university in the United States in exchange for partial course credit. Data from six participants were not analysed, one owing to prior familiarity with the materials and five owing to non-compliance with instructions, leaving 106 participants (*same context-restudy:* $n=22$; *same context-test:* $n=27$; *varied context-restudy:* $n=31$; *varied context-test:* $n=26$) in the final sample (79% female; $M_{age}=19.78$ years). The desired sample size was determined using G*Power 3.1 software (Faul et al., 2007), which computed that 104 participants were required to detect an average testing effect ($g=0.50$; Rowland, 2014) as a main effect or interaction in a 2×2 analysis of variance (ANOVA) with 70% power (which was chosen due to limitations of the available participant pool and laboratory space; we acknowledge, however, that this threshold is somewhat lower than current standards for desired power in psychological research). This sample size exceeded those of similar contextual variation studies (e.g., Smith et al., 1978, Experiment 1: $n=16$; Smith & Handy, 2014, Experiment 1: $n=45$).

Design. A fully-crossed 2 (context: same or varied) $\times$ 2 (learning activity: restudy or test) between-participants design was used. The dependent measure was the proportion of words correctly recalled on the test during the final session of the experiment.

Materials. The learning materials were 36 English nouns (e.g., *cream, fluid, monument*) from Kornell, Rhodes, et al. (2011). These words had, on average, a frequency of 20.5 occurrences in 1,014,000 written words (Kucera & Francis, 1967), 5.9 letters, and 2.0 syllables.

Physical contexts. The locations employed are shown in Figure 1.

Location A (testing room). Location A was a small laboratory testing room in the university's Psychology Building. The room contained five desktop computers arranged across several tables, comfortable chairs, and several white storage cabinets. Participants used the first four computers; the experimenter used the fifth computer. Other features were bare white walls, a small window, and smooth grey flooring.

Location B (patio). Location B was a patio adjacent to a university student centre containing square metal tables with heavy metal chairs, which is frequently crowded with students studying and eating, as well as tour groups. Participants were seated at one table, facing each other and using laptops, with the experimenter watching the table from an extra chair. Other features included natural light, stone tiling, and substantial ambient noise.

Location C (seminar room). Location C was a medium-sized seminar room used for meetings or small classes in a

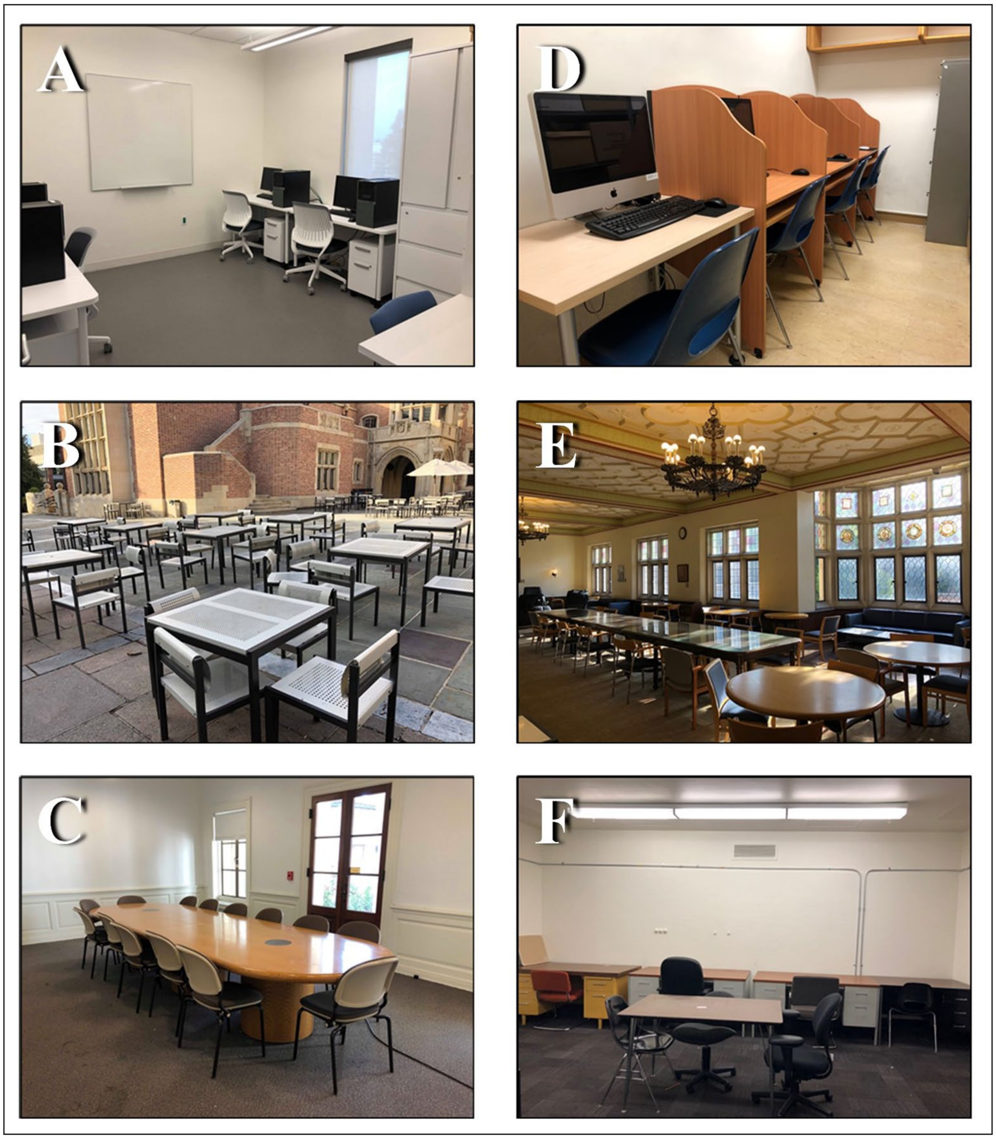

Figure 1. Physical contexts. Experiment 1 involved Locations A, B, and C (testing room, patio, and seminar room). Experiment 2 involved Locations D, E, and F (testing room, study lounge, and office space).

historical university building. The room contained a large wooden oval table surrounded by plush cloth chairs and no computers. Participants were seated around the table and spaced at every other chair to prevent viewing of others' responses. Other features included wood decoration, white walls, carpeting, and natural light.

Procedure. As depicted in Figure 2, participants completed three sessions spaced 48 hr apart. Each session occurred at the same time of day (within 45 min) on Monday, Wednesday, and Friday of a given week. Across sessions, contexts could vary by location, room, experimenter, cohort, and mode of stimulus presentation (depending on condition). The order of locations was counterbalanced such that half of the participants completed Session 1 in Location A and Session 2 in Location B, while the other half did so in the reverse order. All participants completed Session 3 in Location C.

Session 1 (initial study). All participants studied the word list on a desktop or laptop computer seven times. During each study cycle, each word was presented individually for 5 s. Word order was randomised anew prior to each study cycle. To facilitate random assignment of participants to condition and to allow for variation in cohort for the varied-context conditions in Session 2, eight participants per hour participated in Session 1 in two separate 30-min blocks of four participants each (with participants randomly assigned to the first or second block and the first block preceding the second block).

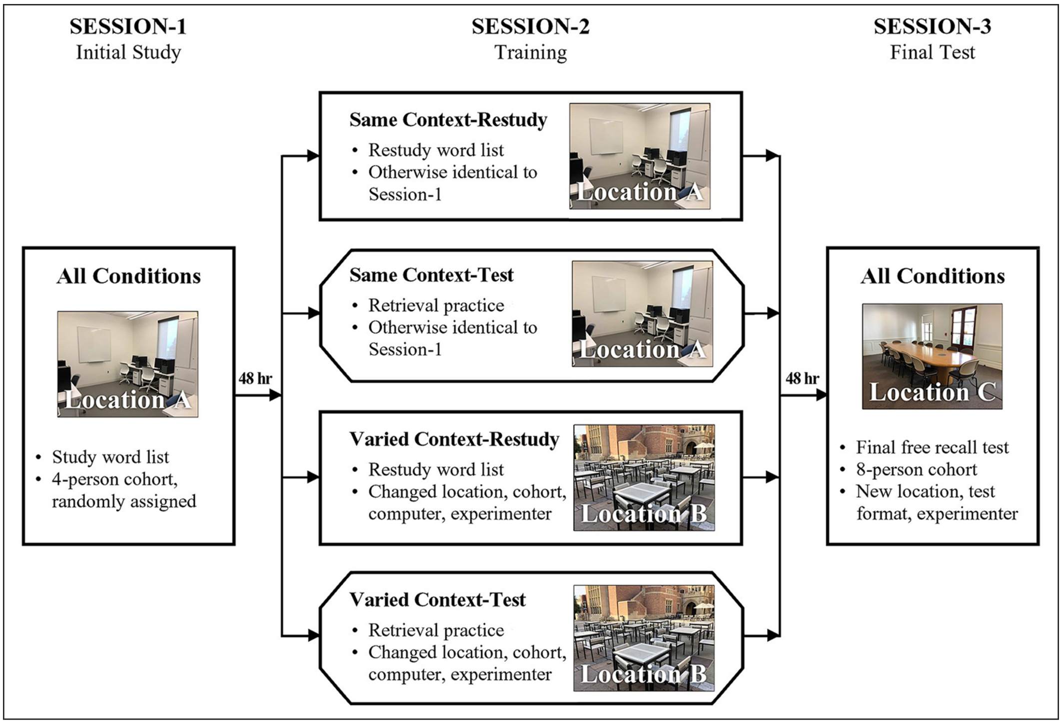

Figure 2. Schematic of the three-session procedure used in the same context-restudy, same context-test, varied context-restudy, and varied context-test conditions of Experiments 1 and 2 (example Experiment 1 arrangement shown). Details of the (re)study and retrieval practice procedures appear in the text. Note a critical difference between Experiments 1 and 2: though Experiment 1 employed pure restudy or retrieval practice during Session 2, that session in Experiment 2 always began with restudy. Thus, the restudy conditions engaged in two cycles of restudy and the test conditions completed one restudy cycle followed by retrieval practice.

Session 2 (training). Participants engaged in restudy (restudy conditions) or retrieval practice (test conditions). In the restudy conditions, participants restudied the word list one time, with the words presented in a random order for 5 s each, for a total of 3 min. In the test conditions, participants took a free recall test wherein they attempted to type all the words that they could remember within a 3-min period. Importantly, participants in the same-context conditions completed Session 2 in the same room, at the same computers, next to the same individuals, and with the same experimenter as in Session 1, whereas participants in the varied-context conditions experienced changes across all of those dimensions relative to Session 1.

Session 3 (final test). All participants completed a handwritten final free-recall test wherein they wrote all the words that they could remember within a 5-min period. Afterwards, they completed an exit survey (results of which are summarised in the 'General discussion' section) and were dismissed.

Results

Session 2 recall. Session 2 recall for the varied context-test and same context-test conditions was compared using an independent-samples *t*-test, with equal variances not assumed due to a significant Levene's test of homogeneity of variances, F (1, 51) = 9.98, p = .003. The *t*-test indicated that participants in the same context-test condition (M = 0.36, SD = 0.26) recalled a significantly greater proportion of the words—more than twice as much—relative to the different context-test condition (M = 0.17, SD = 0.15) during Session 2, t (42.78) = 3.22, p = .002, d = 0.89, 95% confidence interval (CI) = [0.07, 0.30].

Session 3 final recall. Correct recall proportions on the final test by participants in the same versus the varied-context conditions as a function of whether they restudied or were tested in Session 2 are illustrated in Figure 3. A two-way bootstrapped ANOVA with 10,000 samples was conducted with context (same or varied) and learning activity (test or restudy) as the independent factors and proportion of words recalled on the final test as the dependent variable. Bootstrapping was used because Levene's test of equality of error variances was significant, F (3, 102) = 5.22, p = .002. As indicated in Figure 3, a significant interaction between context and study activity was obtained, F (1, 102) = 7.43, MSE = 0.44, p = .008, $\eta_p^2 = .07$. Given that highly significant interaction, main effects were not considered and follow-up tests of simple effects were conducted instead.

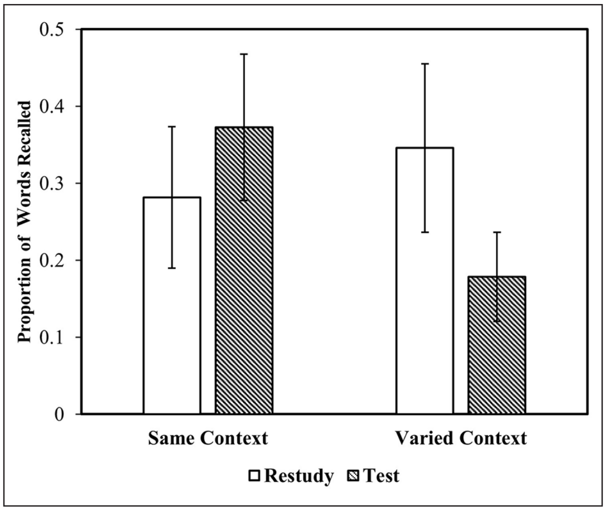

Figure 3. Correct recall proportions on the final test by participants in the same-context versus the varied-context conditions of Experiment 1 as a function of whether they restudied or were tested in Session 2. Error bars represent 95% confidence intervals.

All simple effects were assessed using an independent-samples *t*-test (10,000 samples) with equal variances not assumed.

Effects of contextual variation. Final test recall performance did not differ significantly for participants who restudied in the same context ($M=0.28$, $SD=0.22$) compared to those who restudied in a varied context ($M=0.35$, $SD=0.29$), t (50.61) = −0.92, $p=.36$, $d=0.31$, 95% CI = [−0.20, 0.08], although performance was numerically better in the latter condition. A null effect of context on the study conditions was further indicated by a Bayesian independent-samples *t*-test (conducted via the online tool by Rouder et al., 2009, and using the jeffreys-zellner-siow [ZS] prior), which showed $BF_{10}=2.53$ (suggesting that the data are approximately two-and-a-half times more likely under the null hypothesis than the alternative hypothesis). For participants who were tested rather than allowed to restudy, however, final test performance for those in the same-context condition ($M=0.37$, $SD=0.27$) was significantly better than for those in the varied-context condition ($M=0.18$, $SD=0.15$), t (41.15) = 3.21, $p=.003$, $d=1.42$, 95% CI = [0.07, 0.32].

Effects of learning activity. The same context-restudy condition ($M=0.28$, $SD=0.22$) yielded comparable final test performance as did the same context-test condition ($M=0.37$, $SD=0.27$), indicating no significant effect of learning activity within same context, t (46.99) = −1.29, $p=.21$, $d=0.18$, 95% CI = [−0.23, 0.05], although testing did yield numerically better recall (a Bayesian *t*-test similar to that performed above yielded $BF_{10}=1.78$, suggesting that, although the data are more likely under the null hypothesis than the alternative hypothesis, this evidence is fairly weak). In contrast, restudying in an identical context in Session 2 ($M=0.36$, $SD=0.29$) yielded significantly better final test performance than testing in a varied context ($M=0.18$, $SD=0.15$), t (47.47) = 2.83, $p=.007$, $d=1.25$, 95% CI = [0.05, 0.29].

Session 3 recall conditionalised on Session 2 recall. We examined whether words recalled on the final test by participants in the test conditions were primarily ones that had previously been recalled during retrieval practice in Session 2. This analysis indicated that the words recalled at final test were overwhelmingly ones that had been recalled during Session 2 (same context-test condition, $M=0.90$, $SD=0.13$; different context-test condition, $M=0.85$, $SD=0.23$) and that few words that had failed to be recalled during Session 2 were spontaneously recalled during the final test.

Discussion

Restudying in varied contexts produced numerically better recall than restudying in the same context—a partial replication of Smith et al. (1978) with longer between-session intervals. Furthermore, testing in varied versus the same contexts had deleterious effects. Specifically, contextual variation in Session 2 reduced retrieval success, which in turn lowered final test performance (it might even be argued that contextual variation “reversed” the testing effect). Although contextual variation and retrieval practice have been shown, independently, to promote long-term retention, we observed circumstances wherein their combination can have sub-additive effects on later recall. Finally, it is notable that a potent context reinstatement effect was observed during Session 2 in the same context-test condition relative to the varied context-test condition.

Experiment 2

The findings of Experiment 1 surprised us in two ways. First, the benefit of contextual variation for restudying on later recall was modest relative to prior findings. Although longer between-session intervals may have been a factor, participants also spent unequal time in the Session 1 and Session 2 contexts (30 min vs 5 min, respectively), unlike in prior studies. That disparity could have reduced opportunity for, and thereby weakened, encoding of Session 2 environmental cues. Thus, although we continued to use sessions separated by 48 hr, Session 1 and Session 2 had equal durations in Experiment 2. Second, we found that recalling information in a new Session 2 context negatively affected final test performance. To test whether this decrement was due to increased retrieval difficulty and reduced retrieval success in a new context (possibly due to reduced context reinstatement), we added a restudy cycle prior to retrieval practice in Session 2. We expected this restudy cycle to improve Session 2 retrieval success, and consequently final test performance, particularly for the varied context-test condition.

Method

The design and analysis plan was preregistered at Open Science Framework (https://osf.io/8spdw/?view_only=c97efe84a0c34885a3e4d336dc9232ae).

Participants. We recruited 211 participants in the same manner as in Experiment 1. The target sample size, determined using an a priori power analysis in G*Power 3.1 (Faul et al., 2007), was larger in an effort to ensure sufficient power to detect a contextual variation effect. Based on an expected effect size of $d=0.45$ (Smith & Vela, 2001), 48 participants per group is needed for 70% power in a between-participants design. Data from eight participants were not analysed, three owing to a stated lack of English proficiency and five owing to noncompliance with instructions, leaving 203 participants (same context-restudy: $n=52$; same context-test: $n=51$; varied context-restudy: $n=50$; varied context-test: $n=50$) in the final sample (72% female; $M_{\text{age}}=20.08$ years).

Design and materials. The design of Experiment 2, with the exceptions noted below, and word list stimuli were the same as in Experiment 1.

Physical contexts. Although logistical considerations precluded using the same locations, we again used three locations on the same university campus that were distinct from one another across multiple dimensions, and the photos for these locations are shown in Figure 1.

Location D (testing room). Location D was a small rectangular room in the university's Psychology Building, which contained a row of four desktop computers in wooden test booths (which participants used), metal chairs, and stacks of storage boxes. Features included bare white walls, no windows, bright fluorescent lights, and tile flooring.

Location E (study lounge). Location E was a large room in a university student centre that was typically full of students quietly studying. Features included high, ornate ceilings, carpeted flooring, lighting from chandeliers and large windows accented with stained glass, oil paintings, sofas, and plush chairs arranged around wooden tables.

Location F (office space). Location F was a midsize office room in the university's Life Sciences building that is temporarily in use for graduate-student office hours. The room contained large metal desks in different colours along its walls, fabric chairs, dirty carpet, no windows, a low ceiling, bare white cinderblock walls, dim fluorescent lighting, broken appliances, and had a distinct, musty smell.

Procedure. The three-session procedure was largely the same as that of Experiment 1 (i.e., involved three sessions, each separated by 48 hr), except that (a) the locations were different from that of Experiment 1 and (b) there were small changes within each session as indicated below.

Session 1 (initial study). All participants studied the word list two times (as opposed to seven as in Experiment 1) and did so in Location D using provided desktop computers.

Session 2 (training). Participants in the restudy conditions restudied the word list two times; participants in the test conditions first restudied the word list one time and then took a 3-min free recall test. Consequently, the duration of Session 2 was identical to that of Session 1 at 6 min. The same-context conditions occurred in Location D using desktop computers and with participants in groups; the varied-context conditions occurred in Location E using pen-and-paper materials (i.e., flashcards for restudying and free recall worksheets for testing) and participants were run individually.

Session 3 (final test). The final test occurred in Location F and was administered using laptop computers (a novel format for all participants relative to Sessions 1 and 2). No exit survey was administered.

Results

Session 2 recall. An independent-samples t-test revealed that participants in the same context-test ($M=0.38$, $SD=0.20$) and different context-test ($M=0.38$, $SD=0.17$) conditions recalled nearly identical proportions of the word list, $t(99)=-0.06$, $p=.96$, $d=-0.01$, 95% CI=[−0.07, 0.07] during Session 2.

Session 3 final recall. Correct recall proportions on the final test by participants in the same- versus the varied-context conditions as a function of whether they were in the restudy or the test conditions of Experiment 2 are shown in Figure 4. The analysis procedures mirrored those used for the data obtained in Experiment 1, as Levene's test of homogeneity of variances was significant, $F(3, 199)=6.10$, $p=.001$. The two-way ANOVA again revealed a significant interaction as suggested in Figure 4, $F(1, 199)=5.30$, $MSE=0.16$, $p=.02$, $\eta_p^2=.03$. Simple effects analyses were then conducted according to the procedures outlined in Experiment 1.

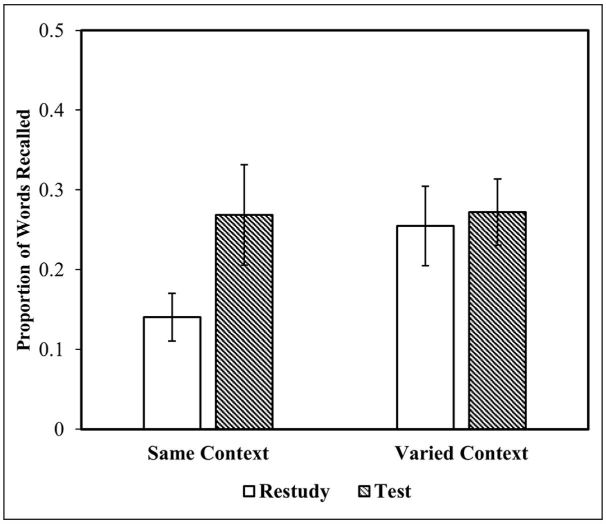

Figure 4. Correct recall proportions on the final test in Experiment 2 as a function of whether participants were in the restudy or test conditions of Experiment 2. Error bars represent 95% confidence intervals.

Effects of contextual variation. A significant effect of contextual variation within restudy conditions was observed, with those who restudied in the same context in Session 2 (M=0.14, SD=0.11) performing worse on the final test than those who restudied in a varied context in Session 2 (M=0.26, SD=0.18), t (78.47)=−3.91, p < .001, d=−0.78, 95% CI=[−0.17, −0.06]. In contrast, no significant effect of context was observed for the test conditions, with participants in the same context-test condition (M=0.27, SD=0.23) and in the varied context-test condition (M=0.27, SD=0.15) recalling comparable proportions of the word list at final test, t (86.40)=−0.10, p=.92, d=−0.02, 95% CI=[−0.08, 0.07]. Evidence for a null effect of context on the test conditions was indicated by a Bayesian t-test, with BF_{10}=4.74, providing substantial evidence for the null hypothesis.

Effects of study activity. A significant effect of restudy versus retrieval practice for same-context conditions was observed, with participants in the same context-restudy condition (M=0.14, SD=0.11) recalling a significantly lower proportion of the word list than did participants in the same context-test condition (M=0.27, SD=0.23), t (69.81)=−3.62, p=.001, d=−0.72, 95% CI=[−0.20, −0.06]. In contrast, participants in the varied context-restudy (M=0.26, SD=0.18) and participants in the varied context-test (M=0.27, SD=0.15) conditions demonstrated comparable recall rates, t (95.04)=−0.52, p=.61, d=−0.10, 95% CI=[−0.08, 0.05]; a Bayesian t-test yielded BF_{10}=4.21.

Session 3 recall conditionalised on Session 2 recall. Supplemental analyses again indicated that most of the words recalled on the final test had previously been recalled during Session 2 retrieval practice for both the same context-test (M=0.93, SD=0.12) and varied context-test (M=0.96, SD=0.07) conditions.

Discussion

During Session 2, testing rather than restudying in the initial learning context yielded better final test performance. In addition, restudying in a new location yielded better final test performance than restudying in the same location. Thus, Experiment 2 replicated the contextual variation memory benefit across multi-day intervals and a testing effect. We did not, however, find greater benefits from engaging in retrieval practice in a new context during Session 2. Instead, testing in a new or the initial learning context was equally effective. Although that finding differs from results obtained in Experiment 1, there was again a strong correspondence between Session 2 and final test performance for both test conditions, with Session 2 retrieval success predictive of Session 3 recall.

General discussion

We investigated the effects of contextual variation and testing on later recall across multi-day intervals and with a novel focus on interactions between such effects. In Experiment 1, contextual variation numerically enhanced the effectiveness of restudying and significantly impaired the effectiveness of retrieval practice. In Experiment 2, with Session 1 and Session 2 durations equated via an added restudy cycle, we fully replicated the benefit of contextual variation when Session 3 testing took place in a new environment. We also found that the greater recall of items in Experiment 1 during Session 2 when the Session 2 environment matched the Session 1 study environment could be completely offset by providing a single restudy opportunity in Session 2 before the items were tested.

Our results support two major conclusions. First, contextual variation can enhance the effectiveness of restudying, as measured on a subsequent test. Second, varying the environmental context from a study session to a test session (administered without feedback) reduces the level of recall, which then, depending on the level of recall, can result in poorer performance on a subsequent test in a new context—compared to providing a single restudy opportunity in the new context.

A strengths-of-memory-distributions account of contextual variation and retrieval practice effects

Our results align with the *strength-of-memory-distributions framework* proposed by Kornell, Bjork, and Garcia (2011; see also Halamish & Bjork, 2011, and Storm et al., 2014),

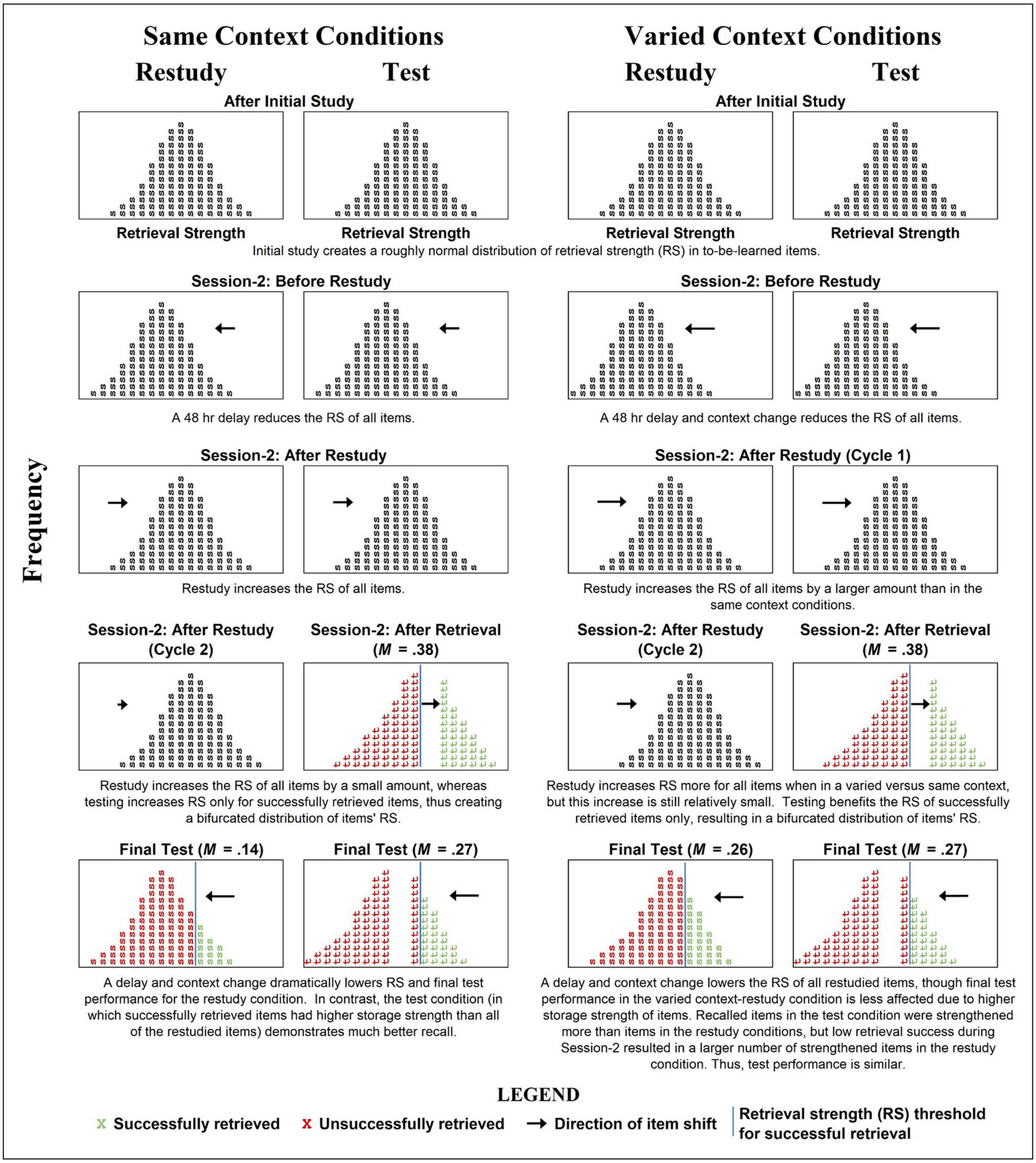

Figure 5. Distribution of the retrieval strength (RS) of 100 hypothetical items at each phase of Experiment 2. Distribution of items match actual Experiment 2 data, where appropriate, and are interpreted within the new theory of disuse framework.

a framework that derives from Bjork and Bjork's (1992) *new theory of disuse.* The basic idea is that pre-existing differences between items with respect to factors such as frequency or concreteness, coupled with moment-to-moment differences in attention or processing efficiency, result in the studied items being distributed in a roughly normal-distribution way on a memory strength dimension. It is then assumed that restudying to-be-learned items shifts the entire distribution, whereas testing (without feedback) shifts the strength of only the recalled items, bifurcating the distribution. Because recalling is a more powerful event than is restudying, the recalled items gain a bigger boost in memory strength than do corresponding restudied items. These dynamics are shown in a simplified way in Figure 5.[1]

Another simplification shown in Figure 5 is what Bjork and Bjork (1992) referred to as *retrieval strength*, which is assumed to reflect the activation or accessibility of to-be-recalled information and is heavily influenced by factors such as current situational cues and recency of study or recall. What is not shown explicitly is *storage strength*, which is assumed to reflect how entrenched or inter-associated a memory representation is with related knowledge and skills. Current performance is assumed to be entirely a function of current retrieval strength, but storage strength is assumed to slow the loss (forgetting) and enhance the gain (relearning) of retrieval strength. Importantly, the theory assumes that increases in storage strength are a decreasing function of current retrieval strength (thus, little additional learning—increases in storage strength—occur when highly accessible information is recalled or restudied). The length of the arrows in Figure 5 reflects the assumed interaction of storage strength and retrieval strength (for a graphical representation of that interaction, see Storm et al., 2014).

Interpretation of the present findings

Experiment 1. When tested in Session 2, there was a large advantage of reinstating the Session 1 context (36% recall vs 17% recall), which reflects that retrieval strength is highly dependent on current cues. For the items actually recalled, the act of recalling has a major impact on both the storage strength and retrieval strength of the recalled items, meaning that the items recalled 48 hr later in a new context consist virtually entirely of items that were recallable in Session 2.

When items are restudied, however, the lowering of retrieval strength via a change in contextual cues enhances the gain in storage strength, which then leads to less forgetting (loss of retrieval strength) across the 48 hr until the final test. The numerical, but not significant, advantage of varying the restudy context (35% vs 28%) is consistent with Smith et al.'s (1978) finding that restudying in a new context can help later recall in a novel/new context.

Experiment 2. When items are restudied twice in a new environmental context, versus being restudied twice when back in the Session 1 context, the change of environmental context lowers retrieval strength, as shown in Figure 5, enhancing the gain of storage strength when items are restudied, which then leads to less forgetting across the 48 hr until the final test in a novel context. Thus, restudying twice in a new context, versus restudying twice when back in the original context, results in better recall 48 hr later in a novel context (26% vs 14%)—more clearly replicating Smith et al. (1978).

When items are restudied and then tested during Session 2, the change in environmental context from Session 1 to Session 2 reduces retrieval strength during the restudy cycle, enhancing the gain of storage strength, which then reduces the forgetting (loss of retrieval strength) from the restudy cycle to the following test cycle, contributing to about the same level of recall (38%) during Session 2 independent of whether participants are back in the Session 1 context or in a new context.

When, in a Session 2 context that differs from the original context, items are restudied twice, recall 48 hr later in a novel Session 3 context is about the same (38%) as when the items are restudied once and then tested without feedback during Session 2. As illustrated in Figure 5, such equivalent recall during Session 3 appears to reflect a trade-off between the effects of all items being strengthened via restudying, versus fewer items being strengthened via testing, but strengthened to a greater extent in the study-test condition. Given this interpretation, it should then be the case that at a still longer retention interval an advantage of the test condition over the restudy condition would appear (cf. Storm et al., 2014).

Other considerations with respect to the current findings

Given evidence that unsuccessful retrieval followed by feedback can improve learning over restudying (e.g., Kornell et al., 2009) and testing alone (e.g., Kornell, Bjork, & Garcia, 2011), adding feedback might have reversed the deficit in the varied context-test condition in Experiment 1. Supporting evidence comes from Smith and Handy (2014), which in three out of four experiments found a benefit of testing with feedback in varied digital contexts versus the same context when a final test occurred in a neutral context. Multiple practice test trials with feedback may also be necessary for a benefit of varied versus same-context conditions to manifest (as was observed in Smith & Handy, 2016, for the case of five vs fewer repetitions; see also Schwoebel et al., 2018, for another case of repeated testing with feedback). A further consideration is the interval between initial study and retrieval practice. In the prototypical testing effect experiment, retrieval practice occurs in the same session as initial studying of to-be-learned information (Pan & Rickard, 2018), whereas in the present experiments it occurred 48 hr later. Feedback, repetition, and restudy opportunities may be advisable in cases where retrieval practice and initial study are separated across learning sessions. The delay between initial study and restudy may have also contributed to the large context variation benefit for the restudy conditions observed in Experiment 2, possibly through incorporating variation in temporal context as well as environmental context. This added layer of variation may have strengthened the context variation effect and may explain why we found a substantial context variation benefit even though it has not always been observed when contextual variation occurs within one experimental session (Smith & Handy, 2014, Experiment 4).

Practical implications

The present results support studying in varied environmental contexts—contrary to the recommendation to study in one fixed location—and extend the contextual variation benefit to multi-day intervals. They also, however, reveal that the benefits of contextual variation interact with one of the most effective learning techniques known to date: retrieval practice. Our findings suggest that testing in varied environmental contexts may not confer learning benefits over doing so in a fixed location and may, in fact, be detrimental when the level of recall is low and feedback or a restudy opportunity is not available.

It is important to add, though, that creating a level of learning that is sufficient to ensure a high rate of successful retrieval in a changed environment may make the act of retrieval more potent in terms of supporting later recall in a new environment. There is considerable evidence that the more involved or difficult the act of retrieval, provided it succeeds, the greater the learning benefit as measured by the likelihood of later recall (e.g., Appleton-Knapp et al., 2005; Bjork, 1988; Landauer & Bjork, 1978). Said differently, retrieving information from long-term memory is a fallible and probabilistic process—a kind of skill that, like other skills, profits from practice—and the more difficult or involved the act of retrieving to-be-learned materials during the learning phase, the more that act exercises processes that will be needed later when that material again needs to be recalled, perhaps in a different context.

In conclusion, we offer two recommendations to students who want to optimise their performance on an exam, which, of course, is not likely to be administered in a prior study location. First, study in more than one environmental setting. Second, if engaging in some type of retrieval practice of the material, either try to create a level of initial learning that will produce a high level of successful retrieval (at a delay) whether back in the setting of initial learning or not, or be sure that, via cooperative learning or some other mechanism, feedback after errors or omissions is available.

Acknowledgements

The authors would like to thank the team of undergraduate research assistants in the Bjork Learning and Forgetting lab, led by Inez Zung, Rosaline Chow, and Riley Sandberg, that helped run the study; the CogFog lab group and especially Jordan A. Brabec, Saskia Giebl, Stefany Mena, and Erin Sparck for their insightful comments; and Amanda K. Montoya and Steven P. Reise for advice on data analysis procedures.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Megan N Imundo https://orcid.org/0000-0003-4599-4777

Data accessibility statement

The data and materials from the present experiment are publicly available at the Open Science Framework website: OSF.IO/8SPDW.

Note

1. In Figure 5, the length of the arrows indicates the assumed strengthening of the recalled or restudied items. The dynamics are simplified by showing that recalled and restudied items move a constant amount (though a larger amount for recalled items), rather than some proportion of the strength that is left to gain, but that simplification does not change the implications of the distribution model framework.

References

Akan, M., Stanley, S. E., & Benjamin, A. S. (2018). Testing enhances memory for context. *Journal of Memory and Language*, *103*, 19–27.

Appleton-Knapp, S., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes and their interaction. *Journal of Consumer Research*, *32*, 266–276. https://doi.org/10.1086/432236

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123–144). Lawrence Erlbaum.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory II* (pp. 396–401). John Wiley & Sons.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Lawrence Erlbaum.

Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory*, *18*(4), 385–393.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 764–770. https://doi.org/10.1037/0278-7393.11.1-4.764

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. https://doi.org/10.3758/BF03193146

Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition*, *13*(4), 333–345.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112. https://doi.org/10.3758/BF03197590

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology*, *71*, 99–104. https://doi.org/10.1111/j.2044-8295.1980.tb02735.x

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 801–812. https://doi.org/10.1037/a0023219

Hong, M. K., Polyn, S. M., & Fazio, L. K. (2019). Examining the episodic context account: Does retrieval practice enhance memory for context? *Cognitive Research: Principles and Implications*, *4*, e46. https://doi.org/10.1186/s41235-019-0202-3

Hopper, C. H. (2003). *Practicing college learning strategies* (7th ed.). Cengage Learning.

Joubert, S. (2018, March 7). *6 tips for online learning success*. Web log post. https://www.northeastern.edu/bachelors-completion/news/six-tips-online-learning-success/

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998. https://doi.org/10.1037/a0015729

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794. https://doi.org/10.1177/0956797611407929

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). Academic Press.

Matzel, L. D., Schachtman, T. R., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation*, *16*(4), 398–412. https://doi.org/10.1016/0023-9690(85)90023-2

Orr, R., & Foster, S. (2013). Increasing student success using online quizzing in introductory (majors) biology. *CBE: Life Sciences Education*, *12*(3), 509–514.

Pan, S. C., & Bjork, R. A. (in press). Acquiring an accurate mental model of learning: Towards an owner's manual. In A. Wagner & M. Kahana (Eds.), *Oxford handbook of learning & memory: Foundations and applications*. Oxford University Press.

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, *144*(7), 710–756. https://doi.org/10.1037/bul0000151

Pessin, J. (1932). The effect of similar and dissimilar conditions upon learning and relearning. *Journal of Experimental Psychology*, *15*(4), 427–435. https://doi.org/10.1037/h0075537

Rasmussen, H. (2019, May 24). Best ways to study for economics exams. *ThoughtCo*. https://www.thoughtco.com/how-to-study-for-your-economics-exam-1146330

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

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Saufley, W. H., Otaka, S. R., & Bavaresco, J. L. (1985). Context effects: Classroom tests and context independence. *Memory & Cognition*, *13*(6), 522–528.

Schwoebel, J., Depperman, A. K., & Scott, J. L. (2018). Distinct episodic contexts enhance retrieval-based learning. *Memory*, *26*(9), 1291–1296.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(5), 460–471. https://doi.org/10.1037/0278-7393.5.5.460

Smith, S. M. (1994). Theoretical principles of context-dependent memory. In P. Morris & M. Gruneberg (Eds.), *Aspects of memory: Theoretical aspects of memory* (2nd ed., Vol. 2, pp. 168–195). Routledge.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342–353. https://doi.org/10.3758%2FBF03197465

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1582–1593. https://doi.org/10.1037/xlm0000019

Smith, S. M., & Handy, J. D. (2016). The crutch of context-dependency: Effects of contextual support and constancy on acquisition and retention. *Memory*, *24*(8), 1134–1141.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *1*(3), 341–358. https://doi.org/10.1207/s1532690xci0103_4

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220. https://doi.org/10.3758/BF03196157

Storm, B. C., Friedman, M. C., Murayama, K., & Bjork, R. A. (2014). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 115–124. https://doi.org/10.1037/a0034252

Tulving, E. (1983). Ecphoric processes in episodic memory. *Philosophical Transactions of the Royal Society of London*, *302*(1110), 361–371. https://doi.org/10.1098/rstb.1983.0060

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 381–391. https://doi.org/10.1016/S0022-5371(66)80048-8

Wellington, W., Hutchinson, D., & Faria, A. J. (2015). The effectiveness of online quizzing as a repetitive learning tool in a marketing class: A field study of the testing effect. In *Developments in business simulation and experiential learning: Proceedings of the annual ABSEL conference* (Vol. 42). https://absel-ojs-ttu.tdl.org/absel/index.php/absel/article/view/2907