BRIEF REPORT

Drawing your way to an A: Long-lasting improvements in classroom quiz performance following drawing

Shaela T. Jalava[1] · Jeffrey D. Wammes[1,2] · Ken Cheng[3]

Abstract

Generating a visual representation of a concept through drawing has been established as a valuable strategy for improving memory. While this has been demonstrated in both the laboratory and the real world, the findings are mixed in educational settings, perhaps due to variable operationalization of the drawing task. Participants are often provided additional scaffolding for their drawings, including instructions, training, or pre-drawn backgrounds. In the current work, we investigated whether the drawing advantage persists when participants in an introductory university class ($N = 145$) were permitted to create free-form drawings without additional scaffolding, akin to the technique often studied in the laboratory. In addition, because the effect has only been established at relatively short timescales, we tested whether benefits persist over multiple weeks. In the tutorial sessions of an undergraduate Biology course, we asked participants to either draw or copy a set of definitions for terms derived from the course textbook. One and 3 weeks later we quizzed participants by showing them the definition and asking them to recall the associated term. The evidence indicated not only that there was better recall for drawn than copied items after 1 week (even with no training or scaffolding), but also that this effect persisted after a 3-week delay. Our results establish that no external guidance is necessary for drawing to enhance memory, and that the benefits are robust over time, providing a foundation for further research on the durability of the benefits of drawing in the laboratory and the real world.

Keywords Memory · Human memory and learning · Education · University

Introduction

The ability to store useful information in memory and maintain it over long periods is fundamental to our everyday lives. So, it is not surprising that decades of research have been devoted to discovering means through which we might augment our mnemonic capabilities. Interestingly, creating a drawing of the information one is trying to learn tends to dramatically increase one's ability to retrieve it from memory later (Paivio & Csapo, 1973; Wammes et al., 2016). While this memorial enhancement has been established with

basic materials in the lab, its robustness and applicability is yet to be determined. The present work is concerned with determining (1) if drawing improves memory for educational information in a real-world setting, and (2) If the beneficial effects of drawing are robust over extended periods.

Drawing reliably improves learning and memory performance across a variety of materials, participant populations, encoding durations, comparison tasks, and testing formats (Fernandes et al., 2018; Jonker et al., 2019; Meade et al., 2018; Paivio & Csapo, 1973; Van Meter & Garner, 2005; Wammes et al., 2016, 2017, 2018a, b). It has been proposed that drawing enriches our memories by imparting them with vivid contextual information that can be later called upon as a cue for retrieval (Wammes et al., 2017). This contextual information has been characterized as being derived from multiple sources, or modes, of information present at encoding (i.e., pictorial, motoric, and elaborative). Together, these promote the formation of a multimodal memory trace that is more robust and accessible during later retrieval (Wammes et al., 2019), as evidenced by enhanced recollection and source memory (Wammes et al., 2018a). In principle, these

✉ Shaela T. Jalava
19stj@queensu.ca

[1] Department of Psychology, Queen's University, Kingston, Canada

[2] Centre for Neuroscience Studies, Queen's University, Kingston, Canada

[3] Biological Sciences, Macquarie University, Sydney, NSW, Australia

additional sources of contextual information *should* enhance memory regardless of the material being learned, but the evidence remains unclear.

The available evidence is mixed on whether drawing-related techniques are applicable in improving learning of real-world educational materials (for reviews, see Ainsworth & Scheiter, 2021; Van Meter & Garner, 2005). For the most part, the evidence bodes well. However, while some studies show drawing-related benefits (e.g., Alesandrini, 1981; Leopold & Leutner, 2012; Mason et al., 2013; Van Meter, 2001; Zhang & Linn, 2011), many seem to have shown no effect at all (e.g., Hall et al., 1997, Kulhavy et al., 1985), and some findings suggest an impairment in comprehension following drawing (Leutner et al., 2009; Ploetzner & Fillisch, 2017). The lack of consensus probably arises not only due to actual variability across materials in terms of how effective drawing might be, but also due to considerable methodological variability across studies, which often includes entirely different operationalizations of drawing. For instance, many studies move beyond the creation of pure, free-form drawings, adding additional implements or instructions. As a set of illustrative examples, some drawing manipulations provide students with all the components that they should include in their pictorial representation, including pre-drawn backgrounds (Schwamborn et al., 2010) and cut-outs (Lesgold et al., 1977; Lesgold et al., 1975). Other studies provide extensive training in how to draw (Van Meter, 2001), include a comparison condition of only silent reading (Dean & Kulhavy, 1981), or give explicit instructions on the type of drawing to make (such as 'analytic' vs. 'holistic'; Alesandrini, 1981). While these methods often support later memory, it is not clear whether free-form drawing without the additional instructions or implements would similarly afford memory benefits.

These design variations make it difficult to clarify the role of basic free-form drawing on learning, and, indeed, the application of drawing as a learning tool is a decision that should be made based on the learning materials and the subject area (Ainsworth & Scheiter, 2021). We do not highlight these design variations to be critical, but rather to point out the differences between the aims of the previous work and the goal of the present experiment. The body of work in the educational literature is primarily concerned with how to most profitably incorporate the generation of pictorial representations into learning environments, whether it be through free-form drawing or a coached assembly of pictorial representations. We are primarily concerned with extending the cognitive literature on drawing as a multimodal encoding method, and exploring how this activity in its simplest form – where students choose how they would like to draw the material without any external guidance – might affect memory for content learned in the classroom.

In addition to this extension to real-world materials, one major outstanding question regarding the association between drawing and memory is whether the effect persists over extended periods. In both the educational and psychological literatures on drawing-related benefits to learning, retrieval often occurs either immediately or after a very short retention interval. The typical time between learning and test tends only to be a few minutes (Alesandrini, 1981; Dean & Kulhavy, 1981; Meade et al., 2018; Schwamborn et al., 2010; Wammes et al., 2016, 2017). To be useful in educational institutions, therapeutic settings, or everyday life, the benefit of drawing must persist beyond the short delays that are customary in laboratory research. The benefits of other prominent encoding strategies have been shown to persist after 4 days (survival processing, Clark & Bruno, 2016), 1 week (production effect, Ozubko et al., 2012; matching encoding operations at test, Dewhurst & Knott, 2010), 1 month (retrieval practice, Butler & Roediger, 2007), and up to 5 months (retrieval practice and self-explanation, Larsen et al., 2013). With this, the other primary goal of the present research was to test whether the memory trace produced by drawing would similarly endure long retention intervals. In the present study, we tested memory after 1 and 3 weeks, with the additional opportunity to test a few of the studied items after 13 weeks.

While the beneficial effect of drawing on memory seems to be robust in the laboratory context, the results are mixed in the educational literature. If drawing is to be taken up as an encoding technique by students and other laypeople, there must be solid evidence that the beneficial effect of basic free-form drawing on memory translates to the real world, and persists across extended delays. There is reason to suspect that, like other elaborative encoding techniques, drawing will help preserve a memory trace for a longer retention interval than less elaborative encoding techniques. It is possible that memories encoded through drawing are less susceptible to forgetting, which would result in an interaction between encoding condition and retention interval where the benefit of drawing is more pronounced after long delays compared to short delays. This hypothesis, however, has not been formally tested. Thus, the goal of the present research was to test (1) whether the beneficial effects of drawing on memory extend to content learned in a real-world university course setting, and (2) whether these effects would persist over an entire semester.

Method

We preregistered the data collection and analysis plan for this experiment at https://aspredicted.org/QGI_DVJ. The data necessary to reproduce the analyses associated with this experiment can be downloaded from the Open Science Framework at: https://osf.io/7m4e6/.

Participants

All participants for this study were enrolled in BIOL122 (Biological Basis of Behaviour) in the second half of 2019 (Semester 2) at Macquarie University in Sydney, Australia. Because the experiment took place during the students' regularly scheduled practical sections, all students took part and were simply given the option to consent to our use of their data as part of our study. We intended to collect data from as many of the students enrolled in the course ($N = 215$) as possible. Participants were assessed three times for this experiment – on two practical session quizzes and the final exam. Due to students dropping the class or failing to attend one of the relevant practical sessions, the final number of students who provided data for at least one out of three tests was $N = 168$ (64 men, 103 women, one other). The age of participants ranged from 17 to 59 years ($M = 20.34$, $SD = 4.00$).

We set different exclusion criteria for the analysis of quiz performance and the analysis of final exam performance, as we knew that the exam was not designed with this experiment in mind. For quiz performance, we required that all included participants provided data for all within-participants conditions (i.e., they attended the initial encoding session and completed both quizzes). Out of 168 participants who provided data for at least one test, 144 met these criteria. For the analysis of the final exam, we only required that included participants had attended the initial encoding session. A total of 145 participants met those criteria; therefore, we included 145 participants in the paired-samples t-test analyzing final exam performance. This experiment was approved by the Humanities and Social Sciences Research (HASS) Human Research Ethics Committee at Macquarie University (Project ID: 5144).

Materials

As part of the course requirements, students participated in mandatory practical laboratory sessions each week, where they gather in smaller groups and engage in hands-on, interactive activities. During the practical session in Week 2 of the course, the students were given a list of 32 terms and definitions to learn. The terms and definitions were derived from the first two chapters of their online textbook, *Biological Basis of Behaviour* (Cheng, 2018), which corresponded to the first 2 weeks of lectures presented by author K.C., who was also the instructor delivering the course. The concepts covered included prominent figures in biopsychology, key ideas in the area of memory, and aspects of the scientific process. The terms ranged from nine to 43 letters in length ($M = 17.19$, $SD = 7.76$) and from three to 14 syllables ($M = 5.94$, $SD = 2.59$). The associated definitions ranged from 14 to 34 words in length ($M = 22.56$, $SD = 4.9$). The full list of 32 terms and

their associated definitions are included in the data repository: https://osf.io/7m4e6/. The term–definition pairs were arbitrarily divided into two sets of 16. Within each practical section, one set of term–definition pairs was assigned to the draw condition, where participants were tasked with generating a picture inspired by the definition, while the other set was assigned to the write condition, where participants were tasked with copying out the definition. Which set of term–definition pairs was assigned to which condition was counterbalanced across the six sections of the class.

As a secondary analysis, we also extracted questions on the final exam that were relevant to the 32 studied term–definition pairs for use as a secondary outcome variable. However, the final exam for this course was constructed independently of the present research. For this reason, the number of items was not matched across conditions, and very few of the terms and definitions were tested on the exam. The exam consisted of 50 multiple-choice questions and was intended to last 120 min plus 10 min of reading time. There were six questions on the final exam that corresponded to six of the 32 term–definition pairs. Four of these were in one set of definitions, and two were in the other. As a result of this and of our counterbalancing during encoding, half of the students drew two pairs and copied four pairs that were addressed on the final exam, and the other half of the students did the opposite.

Design and procedure

Encoding phase

During a regularly scheduled practical session in Week 2 of the course, students were given a list of 32 term–definition pairs to study. They were told that these items were for practice, to help them learn course material, and that while these items would be quizzed in Weeks 3 and 5, their performance on these quizzes did not count toward course grades. Participants were first given an instruction slide that informed them about the two experimental conditions (i.e., draw and copy), and what they were to do for each (see Fig. 1). Term–definition pairs were then presented one at a time via a slideshow. Each slide contained a bolded term centered at the top, the bolded study condition (i.e., draw or copy) centered and in parentheses underneath, then a definition, not bolded, and both horizontally and vertically centered. For half of the terms, students were asked to copy the definition. For the other half, they were asked to draw something based on the definition. They were given 40 s to complete the task specified with each item before the next slide was revealed.
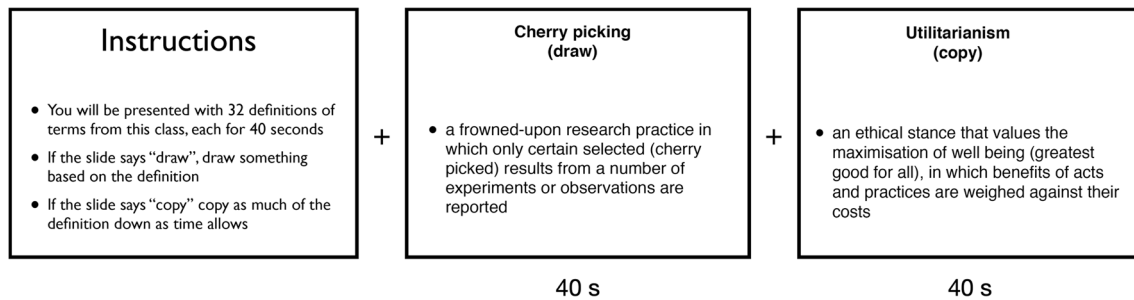
Fig. 1 Example slides from the practical session (encoding phase). The Instruction slide participants viewed at the start of the session (**left**), and two example term–definition slides (**middle** and **right**). Terms and definitions were shown for 40 s each

Delayed testing phase

Following this initial encoding session, participants were quizzed after two different retention intervals, 1 week later and 3 weeks later. Students who were absent from the encoding phase in Week 2 were told not to take the quizzes. One week after encoding, in Week 3's practical session, the students were tested on half of the term–definition pairs that they studied, divided evenly between drawn and copied items. The quiz was hosted on iLearn (an online Learning Management System used in the course), and each question provided the definition and asked the students to provide the corresponding term as a typed entry. The Week 3 quiz took an average of 12 min 39 s to complete. Three weeks after encoding (i.e., 2 weeks after the first quiz), in Week 5's practical session, students were given a quiz with an identical format, testing the remaining half of the term–definition pairs that they studied (i.e., those that were not tested in Week 3). The Week 5 quiz took an average of 13 min to complete. The students were not given feedback on their quiz results. Because we were also interested in how the encoding method impacted students' final exam performance as another outcome variable, we extracted the six final exam questions that directly covered any of the 32 term–definition pairs. The final exam was given in Week 15, 13 weeks after the students initially studied the term–definition pairs.

Response coding

Responses to the quiz were semi-automatically coded for correctness using tools provided by the iLearn system, which hosted the course and its quizzes. Author and course instructor K.C. developed a set of correct answers for each item that was coded into the software. If a given answer matched this set of correct answers, it was auto-coded as correct. The set of correct answers included the exact term from the Encoding phase, as well as acceptable alternates. For example, if the correct answer was "peer review", then "peer reviewing" was entered as an alternate correct answer and would also be auto-coded as correct. Similarly, "mindfulness" was accepted in lieu of "mindfulness

practice". This method of automatic scoring, however, left open the possibility that an answer would be coded incorrectly due to minor typos or spelling mistakes. All answers that were marked as incorrect were reviewed by a member of the research team and manually marked as correct in the case of minor acceptable errors. The researcher in charge of marking was blind to experimental condition. That is, they did not know if the answer was produced by a student who had originally drawn the term–definition pair or copied the term–definition pair.[1]

Results

Quiz performance

Following our preregistered analysis plan, we first assessed whether drawing a definition improved memory for a term–definition pair after a 1- or 3-week delay. Results (Fig. 2, Fig. 3) suggested a small advantage gained from drawing as opposed to copying, and a large decrease in performance with increasing retention interval. The results of a two-way repeated-measures ANOVA (Encoding Condition: copy or draw × Retention Interval: 1 week or 3 weeks) revealed, perhaps not surprisingly, that the proportion of term–definition pairs recalled was higher after 1 week ($M = 0.48$, $SD = 0.21$) compared to 3 weeks ($M = 0.31$, $SD = 0.21$) $F(1, 143) = 135.57$, $MSE = 4.28$, $p < .001$, $\eta_p^2 = .49$, indicating that substantial forgetting occurred over time. The proportion of term–definition pairs recalled that were initially drawn

---

[1] An independent-samples *t*-test revealed that the percentage of quiz items *manually* graded as correct in Week 3 did not differ as a function of whether they were initially drawn ($M = 9.40$, $SD = 12.16$) or copied ($M = 9.20$, $SD = 12.04$), $p = .96$. An identical *t*-test performed for Week 5 also indicated that the percentage of items *manually* graded as correct did not differ as a function of whether they were initially drawn ($M = 4.77$, $SD = 8.74$) or copied ($M = 3.74$, $SD = 7.89$), $p = .73$. This indicates that there was no systematic difference between conditions in grading.
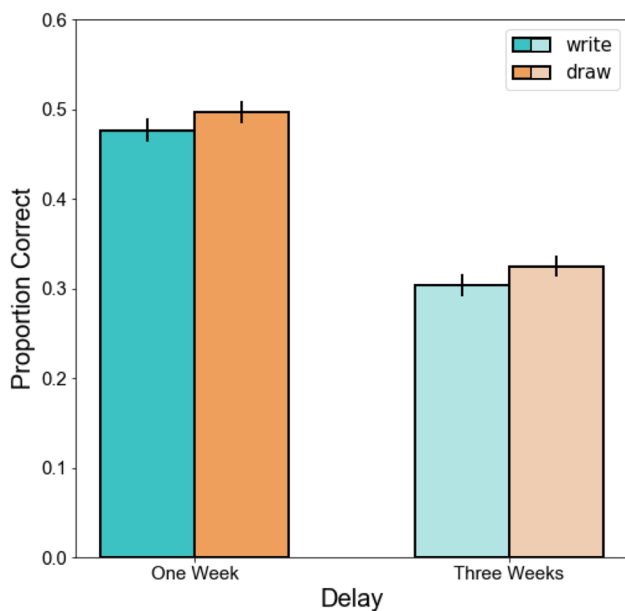
Fig. 2 Average proportion correct on practical session quizzes as a function of encoding condition and delay. Error bars represent the within-participant standard error of the mean (SEM)

Fig. 3 Each line represents participants' change in proportion correct between encoding condition at each time point, with greater magnitudes of change represented by increasing color intensity. Darker orange lines represent a greater difference score favoring drawn items, and darker teal lines represent a greater difference score favoring copied items. The width of the line increases proportionally with the number of participants that are represented by that line. Vertical boxplots display the dispersion of the data at each encoding condition and time point

($M = 0.40$, $SD = 0.23$) was also higher than pairs that were initially copied ($M = 0.38$, $SD = 0.23$), $F(1, 143) = 3.96$, $MSE = 0.06$, $p = .048$, $\eta_p^2 = .03$. Thus, our first two hypotheses were supported, indicating that while there was a reduction in performance as a function of increasing delay from initial study, the beneficial effects of drawing persisted for at least 3 weeks after initial learning. The two-way interaction between delay and encoding condition, however, was not statistically significant, $F(1, 143) = .001$, $MSE = .00$, $p = .97$, $\eta_p^2 = .00$. indicating that, counter to our predictions, the benefit of drawing was not larger after a longer delay. In fact, the magnitude of the effect was near identical 1 or 3 weeks after learning.[2]

---

[2] We performed an exploratory analysis to determine if there were differences in the types of errors participants made as a function of encoding condition. We coded errors as a near miss (semantically or syntactically related to the correct answer, e.g., academic review instead of peer review), a far miss (unrelated to the correct answer, e.g., time management instead of distributed practice), a swap (the correct term for a different quizzed definition, e.g., Ivan Pavlov instead of B. F. Skinner), or a blank response. Two independent raters, blind to condition, coded each error, reaching 92.35% agreement. A third rater, also blind to condition, resolved any disagreements. We calculated the proportion of errors made in each category and each encoding condition within participants, and averaged those proportions across the sample. We performed four paired-samples t-tests on each error category comparing the draw and write conditions. Participants made a greater proportion of near misses when the definition was copied ($M = 0.10$, $SD = 0.11$) compared to when it was drawn ($M = 0.07$, $SD = 0.07$), $t(143) = -2.96$, 95% CI [-0.05, -0.01], $p = .004$. No other comparison reached statistical significance, all ps > .05. We would like to thank an anonymous reviewer for this suggestion.
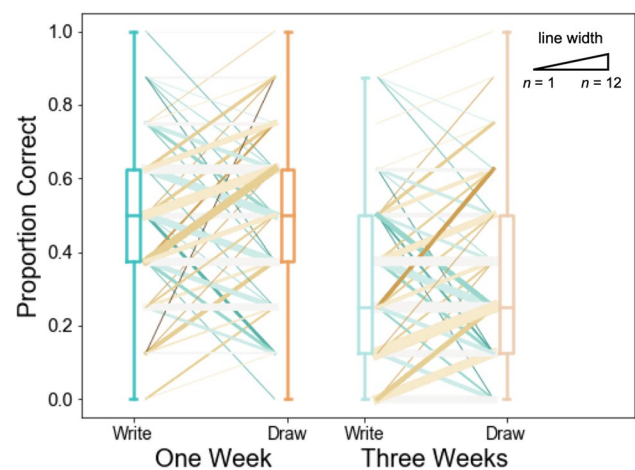
Final exam performance

Next, we assessed whether drawing a definition improved memory for a term on a course final exam given 13 weeks after encoding. The results of a paired-samples t-test (Encoding Condition: copy or draw) revealed that performance on exam questions related to items that were initially drawn ($M = .77$, $SD = .31$) was no better than performance on exam questions related to items that were initially copied ($M = .79$, $SD = .27$), $t(144) = -0.65$, $p = .52$, $d = 0.054$, 95% CI [–.0379, .0759]. Of course, these results should be interpreted with caution as they are based on only six total items.

Discussion

Although the power of drawing to enhance memory has been repeatedly demonstrated in the laboratory (Paivio & Csapo, 1973; Wammes et al., 2016; Wammes et al., 2017; Wammes et al., 2019), the effect of simple free-form drawing in classroom settings is less clear, often due to additional variations, trainings or instructions (Alesandrini, 1981; Lesgold et al., 1977; Lesgold et al., 1975; Schwamborn et al., 2010). Few have examined the impact of a pure drawing instruction in a classroom setting or tested memory after retention intervals long enough to be beneficial in educational settings. Thus, the present findings fill a gap in the body of literature by demonstrating that drawing material learned in a university course can lead to enhanced memory on tests administered weeks later.

Although the exact mechanism driving the drawing effect is not known, other work has proposed that the benefit is driven by the rich, multimodal memory representation produced by the act of drawing (Fernandes et al., 2018; Wammes et al., 2016), an idea that has been supported by empirical work (Wammes et al., 2019). Specifically, creating a drawing of to-be-remembered material incorporates multiple distinct sensory components known to boost memory into one simple technique, including elaborative thought, motoric processes, and pictorial coding (Paivio et al., 1968; Wammes et al., 2019). The current work demonstrates that this sort of multimodal representation can be produced by the free-form drawing technique after just one short study session, and that the benefits persist for weeks. While it is possible that the same benefits might be achieved through the aggregated effects of studying using unimodal techniques over multiple sessions, drawing afforded memory benefits after a single session, in one, unified activity, and was incorporated into the classroom with very few costs.

Although drawing significantly improved quiz performance, the size of the effect did not increase over time. That is, while drawn definitions were better remembered after 1 week, they were forgotten at the same rate. Also, we did not observe a significant improvement on the final exam. One possibility is that the benefit of drawing does not last over such extended retention intervals (~13 weeks). However, there is a more plausible explanation for this pattern of results. First, because the creation of the final exam preceded the incorporation of our study in the course, we had no control over how many exam questions corresponded to the studied term–definition pairs. By happenstance, just six of the exam questions matched our set of term–definition pairs. This limited sample of questions, combined with the fact that the majority of students scored quite high on these six questions overall ($M = 0.79$, $SD = 0.21$; approaching five out of six correct), suggests the presence of a ceiling effect. Future research could include numerous questions across a range of difficulties when testing over long retention intervals to increase power and avoid a possible Type-II error. Alternatively, the information included in the encoding manipulation could be designed based entirely on content known to be appraised on the final exam.

Another limitation of the present study, and indeed any study conducted in real-world settings, is that we could not control for additional time spent studying the term–definition pairs. Many students, being aware that classroom evaluations are a part of the course requirements, would have spent additional time outside of class reviewing course material. While this applies most clearly to performance on the final course exam, it certainly influenced performance on the quizzes as well. Participants may have preferentially studied copied definitions outside of classroom hours, either because they were conscious that they might form a more distinctive memory trace via drawing that was lacking when copying,

or because they devoted more time to studying content that they remembered less well. This lack of control of what happens outside of the classroom might also help to explain why the benefit of drawing was not as pronounced as typically observed in the laboratory, and why there was no difference between copied and drawn items on the final exam. However, it is also possible that participants could have chosen to study the drawn definitions, depending on their intuitions about their retention. We have no evidence to suggest that participants preferentially studied definitions from one condition over another, but future work could explore the factors that drive students' choices about what material to restudy outside the classroom setting. Critically, given the amount of additional noise added from re-exposure outside of our controlled experiment and the small amount of encoding time (and hence drawing time) per item (just 40 s), the fact that there was a reliable benefit of drawing at all for quiz performance speaks to its potency as a learning tool.

Although this study is an important step in applying pure free-form drawing techniques in the classroom (i.e., without the additional effort of training or detailed instructions), there are many opportunities for future research to expand upon these methods. While our materials were pulled directly from the course textbook and are ecologically valid, course content is not always delivered as a series of term–definition pairs in isolation. It will be important to explore the effects of drawing as an encoding technique when the content is a class lecture, where the material is delivered via a continuous narrative with accompanying slides and further context. This type of study design may allow students to use drawing to undergo an even more generative and multifaceted encoding experience than would be achieved by drawing term–definition pairs in isolation. Time in a practical session was limited in this study, resulting in a short time for each item. The results provide reasons for future studies to examine the possible benefits of more elaborate drawing of course materials. Future research can also design a carefully crafted final exam that contains more opportunities to test memory for drawn versus copied items. Drawing is also not a technique that will work in all circumstances, so other fruitful avenues of research will include extending this paradigm to different areas of study (e.g., an English or History course), age groups (e.g., elementary or high school), and difficulty levels (e.g., simpler or more challenging content).

Conclusion

The present study provides a critical step forward in the extension of the beneficial effects of drawing on memory from the laboratory into the real world, demonstrating that student-driven free-form drawing can be used as a strategy in a university classroom to improve test performance. Further,

the findings establish that no external guidance or training in drawing is necessary to enhance memory, that the effect can be achieved after only 40 s of drawing time, and that the benefits have remarkable staying power, remaining intact after three weeks. While more work is needed, these findings indicate that drawing's benefits are robust in real-world settings and over long periods, two pieces of evidence critical to its utility as a mnemonic. This study provides grounds for further and more extensive classroom-based research on the possible benefits of drawing in educational contexts.

References

Ainsworth, S. E., & Scheiter, K. (2021). Learning by drawing visual representations: Potential, purposes, and practical implications. *Current Directions in Psychological Science, 30*(1), 61–67.

Alesandrini, K. L. (1981). Pictorial–verbal and analytic–holistic learning strategies in science learning. *Journal of Educational Psychology, 73*(3), 358–368.

Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*(4-5), 514–527.

Cheng, K. (2018). Biological basis of behaviour. *Top Hat.* https://tophat.com/marketplace/social-science/psychology/textbooks/oer-biological-basis-of-behaviour-noba-project/1895/

Clark, D. P., & Bruno, D. (2016). Fit to last: Exploring the longevity of the survival processing effect. *Quarterly Journal of Experimental Psychology, 69*(6), 1164–1178.

Dean, R. S., & Kulhavy, R. W. (1981). Influence of spatial organization in prose learning. *Journal of Educational Psychology, 73*(1), 57–64.

Dewhurst, S. A., & Knott, L. M. (2010). Investigating the encoding—Retrieval match in recognition memory: Effects of experimental design, specificity, and retention interval. *Memory & Cognition, 38*(8), 1101–1109.

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science, 27*(5), 302–308.

Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words? *Journal of Educational Psychology, 89*(4), 677–681.

Jonker, T. R., Wammes, J. D., & MacLeod, C. M. (2019). Drawing enhances item information but undermines sequence information in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(4), 689–699.

Kulhavy, R. W., Lee, J. B., & Caterino, L. C. (1985). Conjoint retention of maps and related discourse. *Contemporary Educational Psychology, 10*(1), 28–37.

Larsen, D. P., Butler, A. C., & Roediger, H. L., III. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education, 47*(7), 674–682.

Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction, 22*(1), 16–26.

Lesgold, A. M., De Good, H., & Levin, J. R. (1977). Pictures and young children's prose learning: A supplementary report. *Journal of Reading Behavior, 9*(4), 353–360.

Lesgold, A. M., Levin, J. R., Shimron, J., & Guttmann, J. (1975). Pictures and young children's learning from oral prose. *Journal of Educational Psychology, 67*(5), 636–642.

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior, 25*(2), 284–289.

Mason, L., Lowe, R., & Tornatora, M. C. (2013). Self-generated drawings for supporting comprehension of a complex animation. *Contemporary Educational Psychology, 38*(3), 211–224.

Meade, M. E., Wammes, J. D., & Fernandes, M. A. (2018). Drawing as an encoding tool: Memorial benefits in younger and older adults. *Experimental Aging Research, 44*(5), 369–396.

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory, 20*(7), 717–727.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology, 5*(2), 176–206.

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science, 11*, 137–138.

Ploetzner, R., & Fillisch, B. (2017). Not the silver bullet: Learner-generated drawings make it difficult to understand broader spatiotemporal structures in complex animations. *Learning and Instruction, 47*, 13–24.

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology, 102*(4), 872–879.

Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology, 93*(1), 129–140.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review, 17*(4), 285–325.

Wammes, J. D., Jonker, T. R., & Fernandes, M. A. (2019). Drawing improves memory: The importance of multimodal encoding context. *Cognition, 191*, 103955.

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology, 69*(9), 1752–1776.

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

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2018a). Creating a recollection-based memory through drawing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(5), 734–751.

Wammes, J. D., Roberts, B. R., & Fernandes, M. A. (2018b). Task preparation as a mnemonic: The benefits of drawing (and not drawing). *Psychonomic Bulletin & Review, 25*(6), 2365–2372.

Zhang, Z. H., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations? *Journal of Research in Science Teaching, 48*(10), 1177–1198.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.