Some Assembly Required:
Learning Facts in Isolation Limits Inferences

Benjamin A. Motz (bmotz@iu.edu)
Anna L. Chinni (alchinni@iu.edu)
Audrey G. Barriball (aubarrib@iu.edu)
Department of Psychological and Brain Sciences and Cognitive Science Program, 1101 E. 10th Street
Bloomington, IN 47405 USA

Danielle S. McNamara (danielle.mcnamara@asu.edu)
Department of Psychology and Learning Engineering Institute, Psychology Building
Tempe, AZ 85287 USA

Abstract

When learning with self-testing alone, will a learner make inferences between the tested items? This study examines whether self-testing's benefits extend beyond isolated facts to support broader connections between the facts. Comparing self-testing to self-explanation (a strategy known to facilitate inferential learning), we find that while self-testing participants show superior recall of individual facts, they perform significantly worse at making connections between those facts.

Keywords: comprehension; inference; learning strategies; retrieval practice; self-explanation

Introduction

Remembering facts is not the same as understanding them. One can memorize the sentence, "Colorless green ideas sleep furiously," and then recall it from memory, despite it being complete nonsense (Chomsky, 1957). Similarly, Moravcsik & Kintsch (1993) demonstrated that participants could accurately recall some idea units after listening to a spoken text passage despite having no understanding of what the passage was about, and despite generating inaccurate elaborations about the remembered idea units. To be clear, memory is improved by understanding (e.g., Bransford & Johnson, 1972), but the former does not entail the latter. Given that someone's memory for information can be distinct from comprehension of that information, researchers should be concerned about whether learning strategies have limited benefits or facilitate broader understanding. The goal of the current study is to assess whether testing oneself with feedback facilitates broader understanding.

Cognitive scientists have advocated that self-testing is a quintessentially effective studying strategy (e.g., Dunlosky et al., 2013; Roediger & Butler, 2011). The procedure usually starts by studying the to-be-learned content (an initial study phase) and then testing oneself on the content (a retrieval phase). But recent research has gone even further, suggesting that initial exposure might be unnecessary, and that learners might achieve the same level of learning performance more efficiently with self-testing alone (Asher et al., in press; Carvalho et al., preprint). For example, in three experiments, Asher et al. (in press) compared learning performance between participants who merely took a practice test with feedback (no initial study phase) and participants who watched an online lecture video (an initial study phase) and then took the practice test with feedback. There were no significant differences between these two groups on posttest, but participants who did not watch the lecture took less time to achieve this level of performance.

While it may seem awkward to skip initial exposure and jump straight into self-testing, this is a common practice in education. For decades, pedagogists have advocated that lectures should be replaced with active learning, including self-testing (e.g., Knight & Wood, 2005; Dancy et al., 2024). Further, most college students occasionally miss class (Credé et al., 2010), and do not read assigned texts (Berry et al., 2010), but routinely engage in self-testing (Zung et al., 2022). Under this regime, where students rely on self-testing alone, it is unclear whether learning gains are isolated to the tested items or whether they benefit broader understanding.

This phase, "broader understanding," refers to a specific distinction in the comprehension literature between learning content as propositional facts, and learning a broader, interconnected situation model of those facts (van Dijk & Kintsch, 1983; McNamara & Magliano, 2009). The situation model is characterized by the learner generating inferences that go beyond the presented content, creating coherent connections between the separate facts. In the current study, our goal is to examine whether a learner who is using self-testing alone is building these kinds of inferences.

There is reason to be skeptical that a learner who is answering a series of fact questions will spontaneously draw inferences between them. At the dawn of psychology, William James posited that when encountering "cases" about the world, people would struggle to override "our instinctive perception of them as individual facts" (James, 1890). McKoon & Ratcliff (1992) spent a decade studying inferences during reading and ultimately concluded that a reader will only construct the minimal inferences necessary to create a locally coherent representation, and that it would take additional goal-directed effort to infer connections between propositions. More recently, Tran, Rohrer & Pashler (2014) observed no difference in inference performance between self-testing and rereading conditions, although this null effect may be due to a methodological nuance preventing the learner from making connections between test items (Eglington & Kang, 2018).

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

Table 1. Example materials from a folktale (excerpt from source text, corresponding fact questions and inference question).

Source Text	Fact Questions	Inference Question
…Spider thought Tortoise was a nuisance and decided to be mean to him. So Spider told all the animals that they had to wash their dirty hands before eating. All of the animals went to do as Spider had said. There was a path that went through the forest. Tortoise went to the stream and washed. But he had a problem after that. As he walked back, his hands got dirty again. Tortoise tried over and over, but he couldn't walk any other way. So he never got to eat at Spider's table…	What did Spider require of his guests? (*Answer*: They wash their hands before sitting down and eating) What problem did Tortoise have when going to Spider's feast? (*Answer*: When he would walk back to the feast after washing his hands they would get dirty again)	Why couldn't Tortoise get anything to eat at Spider's feast? (*Answer*: He could not follow Spider's rules)

On the other hand, there is also reason to believe that self-testing may facilitate broader understanding, causing a learner to activate and remember information beyond the tested items. Research on retrieval practice generally finds that testing oneself has benefits for transfer, application, and elaboration (Butler, 2010; Pan & Rickard, 2018). While these past studies of retrieval practice typically include an initial study phase, in Asher et al. (in press), participants who self-tested without an initial study phase performed at least as well on application questions, compared with learners who watched a lecture before self-testing. Furthermore, in the recent ManyClasses study (Motz et al., preprint), self-testing *before* instruction caused learners to activate a relevant mental model, facilitating learning from later instruction.

In summary, the literature offers conflicting predictions of whether learners will infer connections between self-tested items, thus building broader understanding. The current study is designed to help resolve this conflict.

Much of the past research on retrieval practice draws a comparison between self-testing and restudying (e.g., rereading the instructional material). However, restudying is known to be an ineffective learning strategy (Dunlosky et al., 2013) and thus fails to provide a reasonable measure of learners' ability to generate and remember inferences. Specifically, if one observed no difference in inferential learning between self-testing and reading (for example), it would be difficult to determine whether self-testing *could have* benefitted inferential learning, or if it had been prevented by some methodological nuance (Eglington & Kang, 2018). Thus, in the current study, we opt to compare self-testing with self-explanation, a strategy known to facilitate inferential learning (McNamara et al., 2023).

Method

The current study was conducted online using jsPsych (de Leeuw et al., 2023), and was exploratory. All materials are openly available at (Motz et al., 2025).

Participants

We recruited 254 participants, who each received credit for partial fulfillment of a course requirement in Introductory Psychology at a large public university in the midwestern US. There were no exclusions; all participants who completed the study are included in the current analysis. Participants self-reported an average age of 18.8 years (*SD* = 1.4) and gender identities that were 73.6% female, 22.8% male, and 3.6% other or non-disclosed. Participants were 68.9% White, 11.8% Asian, and 7.1% African American or Black, and separately, 9.1% reported Hispanic or Latin ethnicity.

Materials

Nine text passages were used: three folktales, three historical narratives, and three science texts. Folktales were from Waddill (1992), historical narratives were adapted from the LibreTexts open library (libretexts.org) and Khan Academy (khanacademy.org), and science texts were adapted from McNamara et al. (2023) and OpenStax (openstax.org).

For each of these nine text passages, we composed four inference questions and eight fact questions. The answers to inference questions were not directly provided in the texts. For a reader to know the answer to an inference question, they would need to make an elaborative inference between two separate sentences, and the constituent facts from these two sentences were included as fact questions. Thus, importantly, someone who knows the answers to each of the 8 fact questions should have all necessary information to infer the correct answers to the 4 inference questions (see Table 1).

Procedure

There were three phases to the study: learning, memory test, and inference test. The order of the memory test and inference test phases were randomized. All three phases were performed in a single online session (see Figure 1).

Learning Phase. During the learning phase, participants learned material from one randomly sampled folktale, historical narrative, and science text (three sets of material total), in a random order. For each of these, the content was presented on a single page, and participants were exposed to the material in one of two different learning conditions: *self-testing* or *self-explaining* (see screenshots in Figure 2**)**. These learning conditions were individually randomized — participants could be assigned to either condition for one, two, or all three sets of learning material during the learning phase.

Self-testing. In the self-testing condition, participants responded to eight multiple-choice fact questions (each with four response options), and received feedback on their responses, but never saw the original source text. Barring any prior knowledge about the passage, participants in the *self-testing* condition would be guessing. The eight fact questions were always presented in the order that the facts appeared in the text; in effect, the two fact questions corresponding to an inference question were always presented consecutively. Inference questions were never shown in the learning phase. To enforce the order that participants responded to the questions, each participant began with the first question shown, with all remaining questions blurred and disabled. The order of response options was randomized. Participants clicked to select a response, and then clicked a button to submit their answer. If they answered correctly, they received feedback ("Correct"), the answer button became disabled, and the next question became visible with its answer button enabled. If they responded incorrectly, they received feedback ("The correct answer is: [answer]"), and then had to mark the correct answer to proceed. Notably, this extra step extends the procedure described in Asher et al. (in press), which did not require participants to mark the correct answer to proceed. After responding to the eight fact questions, participants were shown a text box with the instruction, "Summarize the main ideas in the questions and answers above," and if participants provided a passable summary (see criteria in the *self-explanation* condition, below), then they were permitted to proceed in the study. By asking participants to summarize the main ideas in the *self-testing* condition, we are actively encouraging participants to make inferences and thus avoiding the possibility that participants would have made inferences had only they been instructed to do so.

Self-explaining. In the self-explaining condition, participants read the original source text of the learning content, and after each paragraph, they typed a response to the prompt, "Explain the text in your own words." The web interface prevented participants from selecting or copying the source text, requiring them to read the source text and manually type their response. Analogous to the *self-testing* condition, each participant began with the first paragraph shown, with all remaining paragraphs blurred. Participants had to type a passable explanation of the paragraph to proceed. Self-explanations were evaluated against three minimal criteria: (1) the length of the explanation had to be more than 4 words, and more than 10 characters total; (2) the explanation had to contain at least one English word; and (3) the cosine similarity between the explanation and the original source text had to be $\geq 80\%$. These minimal criteria were not intended to enforce a *good* explanation, but rather to enforce a *passable* explanation, merely ensuring basic compliance with the task. The second and third criteria were evaluated using word2vec embeddings for 10,000 of the most common English words along 300 dimensions (Mikolov et al., 2013). We calculated the average vector of all words in the participant's response, and the average vector of all words in the corresponding source text paragraph, independent of their order. If the cosine similarity between these average vectors was below 80%, participants received the feedback, "Your summary is missing some key elements. Please be sure to summarize the main ideas in your own words." Upon providing a passable response that met all three criteria, the next paragraph became visible, or (if it was the last paragraph of a passage) a button to proceed became enabled.

Memory Test Phase. During the memory test phase, participants responded to the eight fact questions from the randomly sampled learning materials (one folktale, one historical narrative, and one science text), in the same random order as presented in the learning phase. If participants had been assigned to the *self-testing* condition for one of these sets, the memory test was a repetition of the exact same items for that set. If participants had been assigned to the *self-explaining* condition for one of these sets, the memory test items for this set would have been unfamiliar.

In this phase, all eight multiple-choice questions for each set of learning materials were presented on a single page, identical to the self-testing condition, in the same order, but with three key modifications: no feedback was provided, participants could proceed to the next item no matter their response, and there was no prompt to summarize.

Inference Test Phase. During the inference test phase, participants responded to the four inference questions from

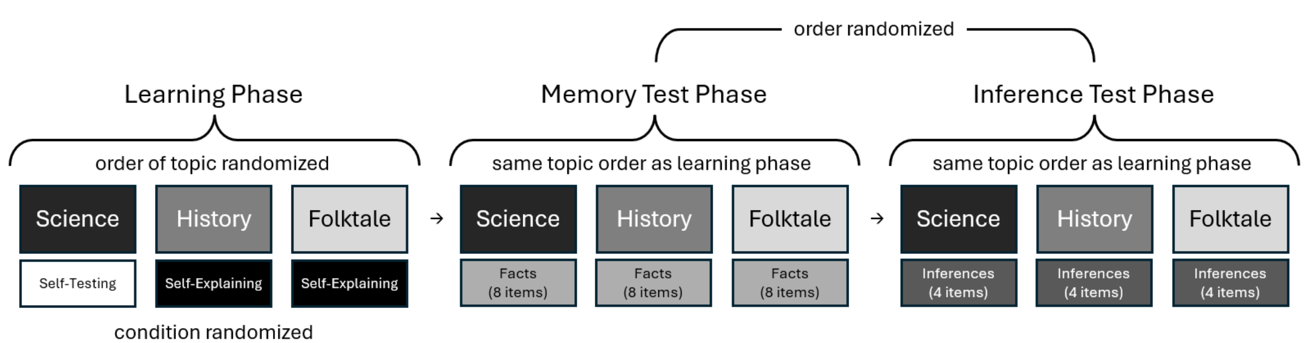

Figure 1. Study procedure diagram.

the randomly sampled learning materials (one folktale, one historical narrative, and one science text), in the same random order as presented in the learning phase. Each inference question involved an elaborative inference between precisely two fact questions, and the inference questions appeared in the same order as their corresponding fact questions would have appeared in the self-testing condition or in the memory test phase.

The inference questions were multiple-choice and were presented on a single page, with no feedback. Participants could proceed to the next item no matter their response. There was no prompt to summarize.

Statistical Analysis

Data from the current study were analyzed using a mixed-effects linear model, with parameters estimated in a Bayesian framework, using brms (Bürkner, 2017), with random intercepts for each participant. Model parameters are estimated with 3 chains and 2,000 iterations. The effective sample sizes (ESS) of all reported estimates are > 1,000 and potential scale reduction factors ($\hat{R}$) are ≤ 1.10.

Results and Discussion

Learning Phase

It took participants a median of 18.6 minutes to complete the learning phase. As expected, for any single set of materials, participants were faster to complete the *self-testing* task (median = 3.5 minutes; where participants marked correct answers) than the *self-explaining* task (median = 6.0 minutes; where participants typed passable summaries).

Self-Testing. During *self-testing*, participants answered eight multiple-choice fact questions with feedback. All fact questions had four response options, only one of which was correct, so the odds of responding correctly by chance were 25%. However, participants' average accuracy on their first response to a fact question was 35.7%, indicating that they possessed some relevant prior knowledge or common sense. The average accuracy on first response was 38.3% for science texts, 36.0% for folktales, and 33.1% for historical narratives. Participants were credibly more accurate in their first response on science questions than historical narratives (estimated accuracy difference in log odds = 0.24; 95% CI: 0.05 to 0.42), which might be expected from the fact that participants had likely been required to take classes covering science topics, but probably had not taken classes covering non-western historical narratives.

If a participant responded incorrectly, they were given the correct answer, which they marked to proceed to the next question. Thus, if a participant responded incorrectly on their first response, their average accuracy on their second response (after being shown the correct answer), was at ceiling, 97.9% on average.

Self-Explaining. During *self-explaining*, participants read a series of paragraphs, and after each paragraph, they explained the text in their own words. Participants' explanations were minimally evaluated for their length, the presence of at least

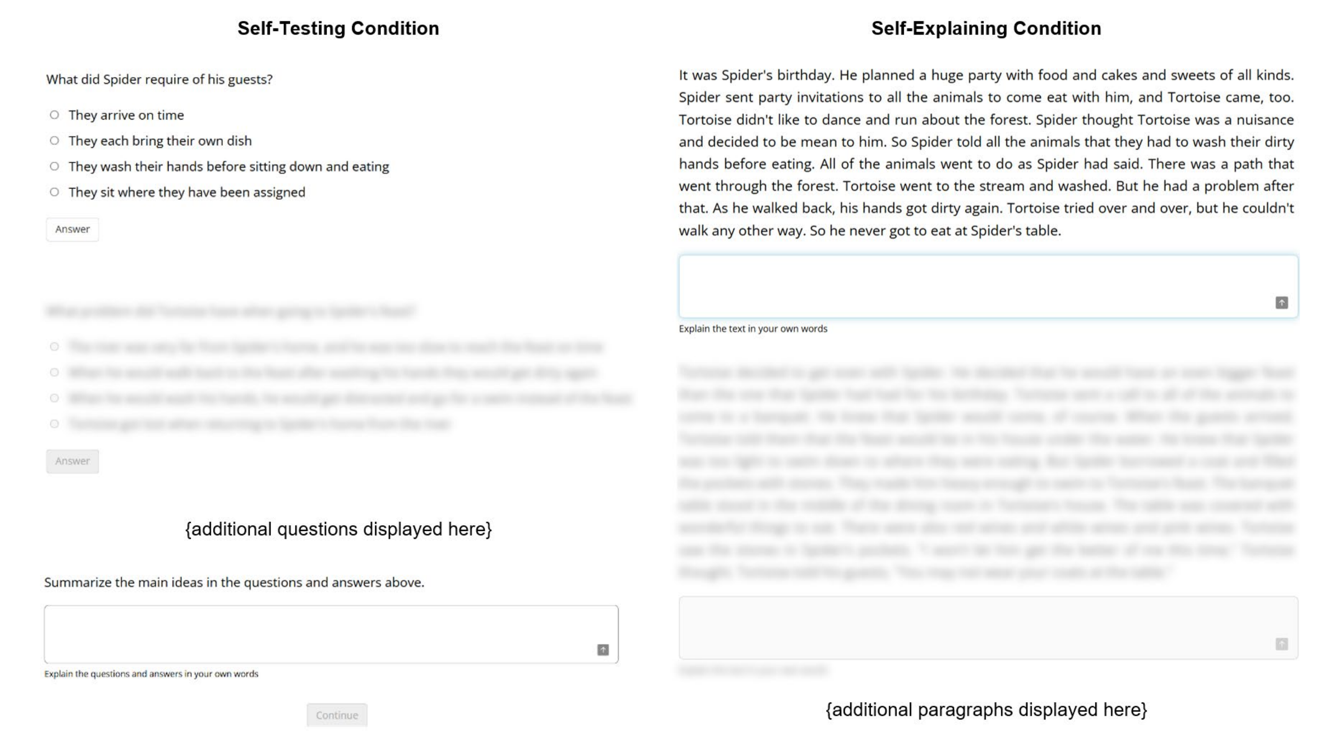

Figure 2. Annotated screenshots of different conditions of the learning phase.

one English word, and the similarity between the explanation and the source paragraph (at least 80% similarity).

Combining all self-explanations submitted in the learning phase, 88.7% passed, 10.6% had low similarity to the source text, and 0.7% were too short (no explanations lacked English words). Of the participants who were assigned to the *self-explaining* task at least once ($n = 216$), 70.4% passed on every opportunity, and never had to revise a submitted explanation.

Passing explanations had an average length of 267.7 characters (*SD* = 140.3; about 54 words) overall; 253.7 for science texts, 276.8 for folktales, and 273.1 for historical narratives. Participants provided shorter explanations for science texts compared with folktales (95% CI: 12.3 to 50.0 fewer characters) and compared with historical narratives (95% CI: 11.8 to 51.3 fewer characters).

On average, passing explanations had 91.6% similarity (*SD* = 4.5%) to the source text. There were no credible differences in similarity between topics.

Comparing Immediate Test Performance Between *Self-Testing* and *Self-Explaining*

Immediately after the learning phase, some participants were randomly assigned to the memory test phase (where they answered fact questions; $n = 108$; 42.5%), while others were randomly assigned to the inference test phase (where they answered inference questions; $n = 146$; 57.5%). Here, we isolate our analysis only to this second block of the study (either memory or inference).

Overall, there was no credible difference in the log odds of answering a test question correctly between fact questions and inference questions (estimate: 0.10; 95% CI: -0.23 to 0.43). There was a main effect of condition, the odds of answering a test question correctly (combining fact and inference questions) were credibly lower when the material was learned in the *self-testing* condition compared with the *self-explaining* condition (estimate: -0.77; 95% CI: -1.03 to -0.51). This is noteworthy and theoretically relevant because, in the self-testing condition (but not the self-explaining condition), participants were explicitly provided the correct answers to all the fact questions. Even so, participants in the self-testing condition were at an overall disadvantage when analyzing test performance overall (combining fact and inference question). However, we observe a strong crossover interaction between question type and learning phase condition (estimate: 1.47; 95% CI: 1.12 to 1.83): participants performed better on fact questions when they had been learned by self-testing, but participants performed much worse on inference questions when the material had been learned by self-testing. This is visualized in Figure 3.

Pairwise contrast confirms that for fact questions, participants performed credibly better when they learned via self-testing compared with self-explaining (estimate: 1.24; 95% CI: 0.69 to 1.77). This is not particularly surprising, because participants were given the correct answers to fact questions in the self-testing condition. Nevertheless, it demonstrates that learning occurred in the self-testing condition: during the learning phase, participants' average first response accuracy on fact question was 35.7%, but accuracy during the immediately subsequent test phase increased to 84.6%. This was across an average retention interval of 8.24 minutes (the median duration between completing the self-testing task in the learning phase with a given set of materials, and then being tested on the same materials in the next testing phase).

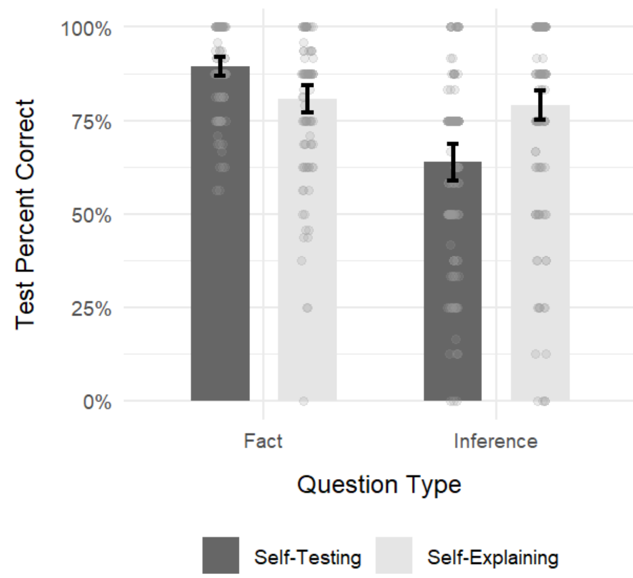

Figure 3. Performance during test immediately after the learning phase. Error bars indicate the upper and lower bounds of the posterior's 95% credible interval of the cell mean. Semi-transparent points are individual participants' scores.

Pairwise contrast also confirms that for inference questions, participants performed credibly worse when they learned via self-testing compared with self-explaining (estimate = -2.11; 95% CI: -2.55 to -1.67). It is noteworthy here that the answers to inference questions were never provided during the learning phase. Further, each inference question was carefully designed as an elaborative inference connecting two fact questions – and despite demonstrating better memory for fact questions, participants who learned via self-testing had worse performance at making the relevant inferences. This was also despite an explicit instruction, in the self-testing condition's learning phase, to summarize the main ideas in the multiple-choice questions and answers.

Does Retrieval Practice Improve Inferences?

Some participants answered inference questions immediately after the learning phase, but others answered fact questions, and *then* answered inference questions in the final block of the session. We might ask whether there were improvements in participants' ability to make accurate inferences in the third block, compared with inference performance in the second block. Such an effect would be expected if retrieval practice, retrieving the items from memory in the second block, improved participants' ability to make inferences.

However, we find no evidence that retrieval practice (answering multiple-choice fact questions during the second block) improved participants' ability to make inferences during the third block, compared with inference performance during the second block (estimate: -0.26; 95% CI: -0.61 to 0.09). Further, there was no difference in inference performance between these two blocks for material that was initially learned by self-testing, nor for material learned by self-explaining. This null finding might be interpreted to suggest that retrieval practice does not operate through elaborative encoding mechanisms, as the act of practicing retrieval of facts from memory did not improve participants' ability to make connections between these facts (Karpicke & Smith, 2012).

Differences Between Learning Materials

To make inferences about the to-be-learned material, a reader needs to integrate the learning material with existing knowledge and construct a coherent representation (Albrecht & O'Brien, 1993). Thus, participants' ability to make inferences depends not only on the learning task they employ, but also on their existing knowledge about the material. In the currents study, we used three different forms of learning material: folktales, historical narratives, and science texts. Here we ask whether the differences between self-testing and self-explaining vary between learning materials.

Overall, combining fact and inference questions and combining learning conditions, we find that learners had lower performance on history topics (estimate: -0.98; 95% CI: -1.51 to 0.46) and science topics (estimate: -1.43; 95% CI: -1.94 to -0.94) compared with folktales. Further, there was a significant two-way interaction, such that self-testing was more beneficial for science texts than folktales (estimate: 1.27; 95% CI: 0.61 to 1.97). However, this two-way interaction is qualified by a significant three-way interaction between question type, condition, and topic: for science texts (compared with folktales), the advantage for self-testing was isolated only to fact questions (estimate: -1.19; 95% CI: -2.15 to -0.22), illustrated in Figure 4.

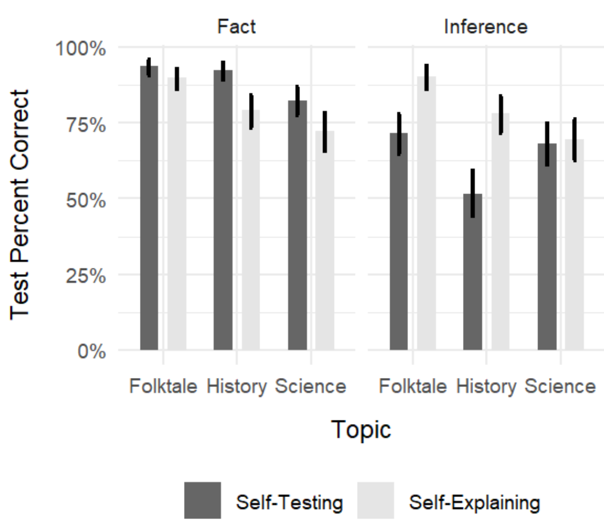

Figure 4. Performance during test immediately after the learning phase, split by question type and topic. Error bars indicate the upper and lower bounds of the posterior's 95% credible interval of the cell mean.

We find that when participants learned via self-testing, they perform credibly worse on inference questions compared with self-explanation (as previously discussed), but this difference does not apply to science topics. There are multiple possible explanations for the lack of a difference for science topics, which are not mutually exclusive. For example, memory test performance was worse for science facts, which likely hindered science inferences. Additionally, participants' self-explanations for science texts were credibly shorter than their self-explanations for folktales and historical narratives, possibly suggesting that participants made fewer inferences about science texts. Our future research will aim to disentangle these possibilities.

Conclusion

The goal of the current study was to assess whether a popular learning strategy, self-testing with feedback, facilitates broader understanding, operationalized as the ability to make inferences between self-tested facts. We find that self-testing yields lower performance on inference questions when compared with self-explaining, an alternative strategy generally known to support inferences.

Researchers have previously advocated that self-testing supports transfer and application of tested facts (e.g., Asher et al., in press; Butler, 2010; Pan & Rickard, 2018). However, being able to generalize factual knowledge to new situations is fundamentally different from making inferences between facts. The idea that self-testing supports the former is not entirely surprising because memory for self-tested content will necessarily carry general semantic meaning (Bransford, Barclay, & Franks, 1972). Here, the current study provides evidence that learning a fact does not necessarily mean that the fact has been integrated into a situation model. Despite strong recall performance of the facts, and despite an explicit instruction for learners to try to integrate the facts, participants who self-tested failed to demonstrate that they were making inferences between those facts at the same level as observed with self-explaining.

In the laboratory, there has been evidence that learning facts alone does not benefit learning higher-order thinking about those facts (Agarwal, 2018). Another recent laboratory study demonstrated that elaborative encoding strategies outperform retrieval practice for making connections between items (Jaeger et al., 2024). Further, in practical learning settings, there is some indication that studying techniques focusing only on memorization can undermine students' long-term success (Ward & Walker, 2008). Taken together, these findings suggest that if broader understanding is the learning objective, some additional assembly will be required for learners who are relying on self-testing alone.

Author Note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305T240035 to Arizona State University, and by a Cox Research Scholarship to A. Barriball. A. Chinni is now affiliated with the Wisconsin Institute for Discovery and the Department of Psychology at University of Wisconsin - Madison.

References

Agarwal, P. (2018). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology, 111*(2), 189–209.

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1061–1070.

Asher, M. W., Sana, F., Koedinger, K. R., & Carvalho, P. F. (in press). Practice with feedback vs. lecture: Consequences for learning, efficiency, and motivation. *Journal of Applied Research on Memory and Cognition.*

Berry, T., Cook, L., Hill, N., & Stevens, K. (2010). An exploratory analysis of textbook usage and study habits: Misperceptions and barriers to success. *College Teaching, 59*(1), 31–39.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*(2), 193–209.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software. 80*(1), 1–28.

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.

Carvalho, P. F., Asher, M. W., Sana, F., & Koedinger, K. (preprint). Skip the reading assignment: Effective and efficient learning with only practice and feedback. *PsyArXiv.*

Chomsky N. (1957). *Syntactic Structures*. Mouton and Company: The Hague.

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research, 80*(2), 272–295.

Dancy, M., Henderson, C., Apkarian, N., Johnson, E., Stains, M., Raker, J. R., & Lau, A. (2024). Physics instructors' knowledge and use of active learning has increased over the last decade but most still lecture too much. *Physical Review Physics Education Research, 20*(1), 010119.

de Leeuw, J.R., Gilbert, R.A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software, 8*(85), 5351.

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.

Eglington, L. G., & Kang, S. H. (2018). Retrieval practice benefits deductive inference. *Educational Psychology Review, 30*, 215–228.

Jaeger, A., Martins, T. H. P. G., Rodrigues, J. P. P., Muniz, B. F. B., da Silveira Fonseca, A. L. S., & de Oliveira Gonçalves, A. (2024). The benefits of elaborative encoding over retrieval practice for associative learning. *Memory & Cognition*.

James, W. (1890). *Principles of Psychology, Vol. 2*. Henry Holt and Company: New York.

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.

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education, 4*(4), 298–310.

Motz, B., (2025). Some Assembly Required: Learning Facts in Isolation Fails to Support Inferences. *OSF Repository.* https://osf.io/ewk82/

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*(3), 440–466.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 51, pp. 297–384). Academic Press: San Diego, CA.

McNamara, D. S., Newton, N., Christhilf, K., McCarthy, K. S., Magliano, J. P., & Allen, L. K. (2023). Anchoring your bridge: The importance of paraphrasing to inference making in self-explanations. *Discourse Processes, 60*(4-5), 337–362.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv,* 1301.3781.

Moravcsik, J. E., & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Canadian Journal of Experimental Psychology, 47*(2), 360–374.

Motz, B., Chinni, A., de Leeuw, J., Jankowski, H., Aggarwal, A., Amato, M., Berlin, K., Britten, K., Brown, A., Cerchiaro, M., Evans, N., Findley, A., Gorman, R., Gregg, K., Hansen, K., Hollender, H., Hullinger, R., Larkin, P., Lion, M., Long, R., Mannarino, A., Mocko, M., Moore, K., Packowski, J., Palmer, C., Ritchie, K., Scott, J., Stanton, M., Talcott, L., Wagner, L., Waite, M., Yeager, R., and Fyfe, E. (preprint). ManyClasses 2: The effects of prequestions on media interactions and learning, *PsyArXiv*.

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

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.

Tran, R., Rohrer, D., & Pashler, H. (2015). Retrieval practice: The lack of transfer to deductive inferences. *Psychonomic Bulletin & Review, 22*, 135–140.

van Dijk, T., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press: New York.

Waddill, P. J. (1992). *The mental representation of narrative: All stories are not created equal*. Doctoral dissertation, Department of Psychology, Purdue University.

Ward, P. J., & Walker, J. J. (2008). The influence of study methods and knowledge processing on academic success and long-term recall of anatomy learning by first-year veterinary students. *Anatomical Sciences Education, 1*(2), 68–74.

Zung, I., Imundo, M. N., & Pan, S. C. (2022). How do college students use digital flashcards during self-regulated learning?. *Memory, 30*(8), 923–941.