Title
Do More Guesses Lead to More Benefits? On the Consequences of Multiple Attempts in the Pretesting Effect

Author
James, Kelsey Katherine

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

DO MORE GUESSES LEAD TO MORE BENEFITS? ON THE CONSEQUENCES OF MULTIPLE ATTEMPTS IN THE PRETESTING EFFECT

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Kelsey K. James

June 2022

The Dissertation of Kelsey K. James is approved:

Professor Benjamin Storm, chair

Professor Nicolas Davidenko

Professor Hannah Hausman

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Tables and Figures

Abstract

Do More Guesses Lead to More Benefits? On the Consequences of Multiple Attempts in the Pretesting Effect

Kelsey James

Pretesting, or answering a question prior to learning the associated information is a powerful tool for learning. By the nature of this type of test, learners are likely to answer a question *wrong.* In this dissertation, I will explore the idea that the incorrect guesses may lead to competition (i.e., source monitoring and/or cue overload/ response competition) between the correct answer and the incorrect guess. If this is the case, any benefit of pretesting is found in spite of this potential competition, and making *more* incorrect guesses should lead to worsened performance compared to making a single guess. However, theories behind the pretesting benefit (i.e., Retrieval Effort Hypothesis, Elaborative Retrieval Hypothesis, Episodic Context Account) would all predict the opposite- that making multiple guesses in response to a pretest question would lead to an *increase* in benefit for later memory. Three experiments were conducted here to determine situations in which generating more guesses as responses to pretest questions will be beneficial or detrimental for learning. Experiment 1 compared the effects of one and three pretest guesses to a read-only control on a final, cued-recall test. Experiment 2 replicated experiment 1 and added a comparison condition in which participants were exposed to prior participants' guesses. Experiment 3 replicated Experiment 1 with the final test changed to multiple-choice and timing being recorded and analyzed across conditions. Across the three experiments, a consistent benefit of making pretesting guesses was found, however, no additional benefit or detriment was found for conditions in which three guesses were made as opposed to one. In short, despite clear reasons derived from the literature to expect that memory might be improved or impaired by the inclusion of extra guesses, no evidence was found to suggest that this is the case. This has important practical implications for teachers and learners looking to implement pretest guessing into their instruction and study strategies as well as implications for the theoretical understanding of the pretesting effect.

Keywords: pretesting, the pretesting effect, the prequestion effect, error correction, competition

Acknowledgements

This dissertation could not have been completed without the help and support of so many wonderful people. I am so incredibly fortunate to have such a strong system of support.

Firstly, I would like to thank Ben Storm, my advisor and a truly kind human being. These six years have not been easy on me but your staunch support and unquestioning assumption that I would succeed helped me to believe in myself as well. You pushed me to improve in ways that I didn't think I was capable of. You gave me space when I needed it. You stepped in when I needed help and stood back when I needed to do things by myself. I cannot thank you enough.

Thanks also to my dissertation committee members, Nick Davidenko and Hannah Hausman. Nick, a person I can always count on, whether it is to pick apart and improve my research ideas or just to be there when I needed someone to listen. I am so grateful for you. Hannah, always an exuberant joy to be around. From conference buddy, to teaching assistant supervisor, to dissertation committee member, you have been such a bright spot of my graduate career.

Thanks also to so many other researchers who have supported me in myriad ways. To my lab mates, past and present. The past members who were my mentors, who supported and inspired me more than I can say. The present members who have listened to me rant and ramble about my research and helped me feel capable even when I was struggling. To everyone at Cog Forum (aka CogMAP) and Cog Fog.

Presenting my research to you all so many times over the years is such a large part of why I am the researcher I am today. To the folks at SOL-ECR (Science of Learning Early Career Researchers Group). I am so grateful for our writing sessions, working groups, and for your friendship. I can't imagine how I would have survived dissertating while on the academic job market without you.

To my family. Your support is everything. From seventh grade when I first told you I would be a research psychologist, to now that I actually am one. You have stood by me in so many ways and I certainly would not be where I am if not for you. Words truly cannot capture my love and gratitude.

To my friends: Nicole, my partner in crime since birth. Nicolle, the best roommate and friend I can even imagine having. Dana-Lis, my lab mate, teaching assistant, mentee, and most of all, friend. Chris, Anie, Justin, and Jordan, my pathfinder crew. There is no group of people with whom I would rather travel across planes in a chicken hut than you all. Mike and Nicole, I wouldn’t give up the pure chaotic joy of friendship with such wonderfully strange people for anything. Kendra and Christina, my dancing crew. I am forever proud and happy to have shared stages with you two. The “council", Emily, Audrey, Shannon, Wenny, and Kristine. I absolutely treasure my friendship with each one of you. Elizabeth, Allison, Yasmin, and Ali. So many of my formative experiences in graduate school happened with you all by my side.

To my strike comrades, Roxy, Julia, Veronica, Allie and so many others. I'm so proud of us for standing up for what is right. I would gladly get fired again with you all by my side.

To Motion Pacific Dance Studio, my second home, my haven. I cannot stress enough how important this space has been during my time in graduate school. I can't imagine how I would have gotten through a pandemic and a dissertation and everything else without you all and without dance. To my teachers, Meredith, Melissa, Rachel, Colette, Shannon, Cookie, and Molly, thank you thank you thank you. To my fellow dancers, I have so much love for all of you. Motion is special, and I am so lucky to have had the opportunities to dance with all of you.

Last but not least, I would like to extend my eternal gratitude to my dog, Indiana, who will never understand the positive effect he has had on my life. The love of an animal is no small thing, and coming home each day to an expression of joy and excitement has kept me moving forward through each of the challenges I have faced during this chapter of my life.

Thank you all for everything. I will never stop being grateful for the support you all have provided me over the years.

Do More Guesses Lead to More Benefits? On the Consequences of Multiple Attempts in the Pretesting Effect

Taking a test after learning, often considered purely as a tool for assessment, is a practice with powerful, long-lasting effects on human memory. This *testing effect* has been demonstrated, time and time again, across learning media, testing format, delay, and a multitude of other factors (e.g., Chan & McDermott, 2007; Little & Bjork, 2015; Roediger & Karpicke 2006; Roediger, Agarwal, McDaniel, & McDermott, 2011; Rowland, 2014; Spitzer, 1939). However, learners can also take a test *prior* to learning the tested information. This practice, called pretesting, is also a powerful tool for learning (e.g., Carpenter & Toftness, 2017; Grimaldi & Karpicke, 2012; James & Storm, 2019; Kornell et al., 2009; Kornell, 2014). Pretesting is used commonly in classrooms, often as a tool for measuring prior knowledge, but can be beneficial even if every question is answered incorrectly (e.g., Richland, Kornell, and Kao, 2009). However, what happens if the teacher encourages students to guess two or three or more times prior to giving a correct answer? It is quite possible that this would increase the benefits of pretesting, but it is also possible that these additional guesses will provide competition. Later on, the learners may be more likely to remember the information that they generated during pretesting, and this heightened accessibility of the incorrect information may interfere with the learner's ability to retrieve the correct information. Throughout this dissertation, I will discuss pretesting and the error generation literature more broadly, theoretical accounts of the pretesting effect, how competition might factor into the benefits of pretesting, how the generation of multiple guesses might lead to competition or benefits, and finally discuss three experiments with the goal of determining if and when generating more guesses might be beneficial or detrimental to learning.

The Pretesting Effect

The benefits of pretesting are well-established in a literature tracing back to the latter half of the twentieth century. Research was conducted on embedded *adjunct questions*, or test questions embedded into a written passage for the purposes of assessing and improving learning in the classroom. Adjunct questions placed before the target information was learned showed early pretesting benefits (e.g., Anderson & Biddle, 1975; Hartley & Davies, 1976; Rothkopf, 1966). Shortly thereafter, the cognitive psychology literature was beginning to study the benefits of guessing a second word in a cue-target pair in a paradigm developed by Slamecka and Feivreski, (1983) who found a notable benefit to making such guesses.

More recently, a newer wave of research on pretesting and related phenomena began with a notable set of experiments conducted by Richland, Kornell, and Kao (2009). Richland and colleagues focused on the benefits of learners attempting to answer questions prior to reading an associated passage as compared to additional time to read said passage. Further, these researchers explored the idea that pretesting might simply lead to attentional benefits by comparing the benefits of attempting to answer pretest questions with the benefits of attentional cues such as italicization and bolding. They consistently found, across five experiments, that attempting to answer pretest questions leads to larger benefits for final test cued-recall performance than learning with additional time and these attentional cues.

The newer wave of pretesting research has found the benefits of pretesting to be quite robust. Such benefits have been found across various learning materials, including trivia questions (Kornell, 2014; Storm, James, & Stone, 2021), multiple choice questions (Little & Bjork, 2016), videos (Carpenter & Toftness, 2017; James & Storm, 2019) word-pairs (Grimaldi & Karpicke, 2012; Kornell, Hays, Bjork, 2009; Vaughn & Rawson, 2012) and text passages (Richland, et al., 2009; St. Hilaire & Carpenter, 2020). Pretesting benefits have been shown to last across a delay of at least a week between feedback and final test (Richland et al., 2009, Storm, James, & Stone, 2021), and a delay of at least 24 hours between pretest and feedback (Kornell, 2014, though see Grimaldi & Karpicke, 2012 and Hays, Kornell, & Bjork, 2013)

It is important to note that the pretesting literature is often conflated with two other literatures: error correction and prequestions. These three terms are often used interchangeably and are difficult to disentangle. The pre*question* literature (e.g., Carpenter, Rahman, & Perkins, 2017) tends to focus more on questions prior to passages, videos, and lectures, while the pretesting literature is more ambivalent to materials. Other than this subtle difference, the terms can be seen as interchangeable. On the other hand, *error correction* (e.g. Slamecka & Feivreski, 1983) tends to focus on word pairs as materials, with a participant guessing the second word in a cue-target combination. Further, and more critically, the term error correction emphasizes that the answer to the question or guess for the target term must be *incorrect*. Pretesting does not necessarily make such assertions. We use the term *pretesting* throughout this paper rather than using the terms interchangeably for the sake of clarity more than due to any of these other distinctions. We do, however, also see the term pretesting as being the most flexible of these three and believe that it captures most closely the situations to which we hope to generalize these results. We define *pretesting* broadly as the act of testing before learning, and *the pretesting effect* as the benefit of taking such a pretest on later test performance.

Theoretical Accounts

Many theories have been posited as to why benefits of pretesting are found. Kornell and Vaughn (2016) review several such theories as they pertain to both testing and pretesting including the Retrieval Effort Hypothesis, the Elaborative Retrieval Hypothesis, and the Episodic Context Account. They argue that it is likely that the same mechanisms can be applied to both taking a test after learning (testing) and before learning (pretesting). Additionally, they argue that these accounts are not necessarily mutually exclusive. Indeed, they argue that it is likely that multiple of these or other mechanisms together lead to the benefits of pretesting and testing.

Retrieval Effort Hypothesis

One theory of testing is called the retrieval effort hypothesis (e.g., Pyc & Rawson, 2009) in which engaging in more effort to retrieve something at initial test leads to better memory at a later test. This theory is derived from the desirable difficulties framework proposed by Bjork and Bjork (2011) in which certain more difficult learning strategies are argued to be better for learning than easier ones. The framework emphasizes that a focus on immediate performance may come at the expense of long-term learning. If learners are engaging in easier study strategies that allow them to perform better during the study session, this may mean that they are *not* receiving the full long-term memory benefits that may arise from more effortful strategies. Retrieval (e.g., Roediger & Karpicke, 2006), spacing instead of massed study (e.g., Baddeley & Longman, 1976), interleaving instead of blocked study (e.g., Kornell & Bjork, 2008), and varying conditions of studying (e.g., Smith, Glenberg, & Bjork, 1978) are some main examples of desirable difficulties. These are all strategies that may lead to difficulties during learning and therefore poorer performance during a study session but have been regularly shown to lead to better retention over time.

The Retrieval Effort Hypothesis argues that it is specifically due to the difficulty or effort itself that testing is beneficial. This benefit may be *cognitive,* meaning that the benefit of additional effort may be derived directly from the increased effort during encoding, or may be *metacognitive*, meaning that the benefit is derived from the experience of the difficulty leading to reflection on levels of understanding which can then lead a learner to choose to expend more effort. The theory does not directly address pretesting, but it stands to reason that one could extend the theory to retrieval practice occurring prior to learning. The act of attempting to retrieve an answer from memory- whether the required information has been learned or not- leads to more effort than simply reading the correct information.

In a notable examination of this theory, Pyc and Rawson (2009) investigated the benefits of retrieval with more or less effort in a testing paradigm. In a first experiment, these researchers manipulated the length of inter-stimulus intervals, or the amount of time between repeated tests of the same set of items. They argue that a longer interval between tests leads to more retrieval effort, and therefore should lead to better memory. Indeed, this is what they found. On a final test, participants performed better after practice testing with longer (~6 min) rather than shorter (~1 min) inter-stimulus intervals.

A second study by Pyc and Rawson (2009) further examined the benefits of retrieval effort as measured by key-press latencies, or the amount of time prior to beginning to type an answer response during an initial test. Supposedly, if retrieval is harder, learners may take more time to search their memories for the correct information and then begin to type their answer. They found that longer latencies, used here as a proxy for more effort, correlated with better memory performance on a final test. These researchers argue that the additional effort in these two experiments- manipulated via short and long interstimulus intervals in experiment 1 and measured via latencies in experiment 2- is the reason for the benefits found in such conditions. More effort prior to a successful retrieval led to better memory.

Note that as these researchers (Pyc & Rawson, 2009) used a testing paradigm and focused only on successful retrieval we cannot directly extend these results to pretesting. However, as argued by Kornell and Vaughn (2016), the retrieval effort hypothesis in general and these results specifically can easily be extended to the amount of effort expended due to testing *prior* to learning. More effortful retrieval attempts during a pretest could reasonably lead to better memory for the pretested information.

Elaborative Retrieval Hypothesis

The Elaborative Retrieval Hypothesis (Carpenter 2009, 2011; Cyr & Anderson, 2018), typically ascribed to *testing,* is perhaps even more pertinent as an explanation for *pre*testing. The theory emphasizes that when testing, learners engage in semantic elaboration, searching through information that is semantically related to the question before settling on a response to a question. In doing so, the mind creates potential mediators to connect the cue and response, thereby adding one or more additional retrieval routes for later access to the retrieved information. This idea stems from the spreading activation theory of memory (e.g., Collins & Quillian, 1972) in which memory activation starts at a single "node" and then spreads out to closely connected nodes. The Elaborative Retrieval Hypothesis argues that these activated nodes can become mediators, and at a final test, these mediators can help the learner trace a pathway back to the correct information.

Despite its original focus on testing after learning, it stands to reason that this Elaborative Retrieval Hypothesis may fit even more squarely with pretesting. The mental act of searching through related, non-target information may be even more likely when learners do not already know the specific answer to the question. When testing information that has already been learned, oftentimes the correct information will be the first thing to come consciously to mind (see Kole & Healy, 2013 for a related discussion). With pretesting, as the information is not yet learned, this is much less likely to be the case. Of course, this does not rule out the possibility that the mind may engage in a search through potential semantic mediators prior to the production of a learned, correct answer, as with testing after learning, but just that it might be even more likely when testing prior to learning.

In one notable examination of this theory, Cyr and Anderson (2018) found that the benefit of error generation was smaller with errors that were "out in left field"- or, very unrelated to the target. To test this, the researchers had participants make guesses for the second word in a cue-target pair with the cue being a homograph with two very different meanings (e.g., band as in elastic band versus music band). Cleverly, they manipulated whether the target matched or mismatched the cue by waiting for the participants to respond to the initial cue and then selecting which word would be the target based on what sense of the word the guess fit most closely. For example, if participants were given band and guessed "drum", the mismatched target would be given as "rubber", while in the same situation but with a *matched* trial, the target may be given as “guitar”. All in all, they found that matched trials were much more beneficial to final test performance than mismatched trials. The researchers argue that this provides evidence for the elaborative retrieval model in that the errors people make were much stronger for final test performance when they were semantically related to the eventual target rather than only related through a different meaning of the homograph cue. The benefit of sematic relatedness fits squarely within the predictions of the elaborative retrieval hypothesis- if the error being made is related to the correct information, it may be more likely to act as a semantic mediator and thereby give a benefit through mental elaboration connecting the guess and target.

Episodic Context Account

Another theory of testing and pretesting is known as the Episodic Context Account (e.g., Karpicke, Lehman, & Aue, 2014; Lehman, Smith, & Karpicke, 2014; Metcalfe & Huelser, 2020). This account argues that an initial test or pretest creates a stronger episodic trace than a control study-only condition. Then, when encountering the same question during a subsequent test, learners may reinstate that episodic context. According to this account, it is the reinstatement of that episodic context that leads to the testing and pretesting effects.

One interesting piece of evidence comes from the literature surrounding amnesia patients. Some studies (Hamann & Squire, 1995; Hayman, MacDonald, & Tulving, 1993) have shown that densely anterograde amnesic patients wind up performing *worse* on a test after guessing an incorrect answer than they do if they simply read the correct information. Metcalfe and Huelser (2020) make the argument that this can be used as evidence that the benefits of pretesting found in non-amnesic patients are due to a benefit to *episodic* rather than *semantic* memory. As anterograde amnesiacs have been shown to have deficits in episodic but not semantic memory, if errors created an increase in semantic memory, presumably the pretesting benefit would still be found in amnesic patients.

Another piece of evidence for the Episodic Context Account comes from Metcalfe and Huelser (2020). These researchers used word-triplets with congruent (e.g., wrist-palm) or incongruent (e.g., tree- palm) cue-pairs for target words (e.g., hand). Participants were asked to either guess a third word relating to the cue-pair (pretest) or study the triplet in-tact (control). On a final test, participants were asked to report both the target item (hand) and their own guess. Critically, the benefit of pretesting as compared to control was only found for items for which their original pretest guess was also explicitly recalled. They argue that this is evidence that the specific episode must be recalled in order for learners to obtain the benefits of generating an error. If learners are recalling the entire episode explicitly in the case of pretesting but not in control conditions, it follows that this explicit recall is leading to the pretesting benefit.

There is no reason that these theories- Elaborative Retrieval, Retrieval Effort, and Episodic Context- should be mutually exclusive, and indeed as argued by Kornell and Vaughn (2016), it is likely that multiple of these mechanisms are together at play to make up what we know of as the testing and pretesting effects. For instance, engaging in more effortful activity like retrieval may be more episodically memorable or may lead to a stronger semantic connection than less effortful activity. Additionally, episodic benefits of pretesting may be most critical early on, while more semantic benefits may arise with repeated retrievals and across a delay.

Pretesting and Competition

By the nature of pretesting, learners are unlikely to generate the correct answers to pretest questions, as they have not yet been exposed to the critical information. When an incorrect guess is made, what happens to the memory of that incorrect information in the mind of the learner? It is quite possible that the incorrect answer may lead to some form of competition. Indeed, it is important to note that, assuming the participant generates any non-target item in response to whatever cue, whenever the pretesting effect is found, it is found *in spite* of there being an additional item (the guess itself) to compete with the answer compared to any control in which participants are only ever exposed to the correct answer. The act of generating the guess during the pretest may lead to the accessibility of that guess being strengthened, and then, during a final test, the guess with heightened accessibility could compete with the memory for the correct answer. If the pretesting effect is found in spite of this, it may be that the field of research surrounding pretesting is underestimating the true benefit of pretesting and that a larger benefit may be observed when conducted in comparison to a condition with a similar level of competition.

The nature of this potential competition may come in several forms. Two possibilities are discussed here, namely source monitoring and cue overload/response competition. These possibilities are not mutually exclusive and it is entirely reasonable for multiple different mechanisms of competition to be at play during pretesting.

Source Monitoring

Errors due to poor *Source Monitoring* is one possible source of competition in pretesting. Source monitoring (for review, see: Johnson, Hashtroudi, & Lindsay, 1993) is the idea that the *source* for a memory includes the context and other information about the initial event in which the memory was acquired. A source monitoring error occurs when someone is unable to accurately determine the source of a memory. Source monitoring errors can occur within many areas of memory research, including errors in eyewitness testimony (Rantzen & Markham, 1992), Cryptomnesia, or plagiarism due to forgetting that information had been learned elsewhere (Brown & Murphy, 1989), and amnesia (Brown & Brown, 1990).

Within the context of pretesting and competition, source monitoring failures may lead learners to be unable to remember whether a recalled potential answer to a test question was their initial incorrect pretest guess or the correct answer. A source monitoring error in this case may lead learners to answer a final test question incorrectly.

Cue Overload/Response Competition

Another possible mechanism for competition in the pretesting effect is that of a *cue overload* and the *response competition* that results from it. Cue overload principle was coined by Watkins and Watkins (1975) as a response to the discussion around the buildup of proactive interference as found in the Brown-Peterson paradigm (Brown, 1958; Peterson & Peterson, 1959). In the Brown-Peterson paradigm, participants learn a list of words, engage in a distractor, then attempt to recall said list. Keppel and Underwood (1962) demonstrated that participants engaging in such a paradigm would perform steadily worse over the course of the study due to a buildup of proactive interference from the learning of prior lists. Watkins and Watkins (1975) argued that this proactive interference effect was actually due to the cue used at test becoming overloaded and that test performance becomes worse over the course of such a study due to there being too many potential responses for any given cue. These researchers conducted a series of experiments, each demonstrating that final test performance got worse as the number of items increased. They argued that this was evidence in favor of a cue-overload interpretation of the buildup of proactive interference effect found in the Brown-Peterson task.

Resulting from the overloading of the cues is the competition mechanism of response competition (e.g., Eriksen & Eriksen, 1974; Eriksen, Coles, Morris, & O'Hara, 1985, Proctor, 1981). Response competition is the idea that when there are multiple possible responses to a cue that are simultaneously activated, they lead to mutual interference as one attempts to produce the desired response. One famous finding surrounding response competition is that of The Stroop Effect (Stroop, 1935, as discussed in Eriksen et al., 1985), in which participants attempt to read the ink color of a word denoting the name of a color. When the ink color differs from the color name, learners must inhibit the prepotent response to read the word written in order to produce the desired response of naming the ink color.

Within the discussion of pretesting, a cue overload/ response competition explanation may argue that making pretest guesses may lead to competition for final test performance due to overloading the cue, which in this case is the pretest question itself. The overloading of the cue then leads to response competition on the final test when the learner is attempting to produce the correct answer to the pretest question instead of the guess.

Number of Guesses

How does making more than one guess affect the pretesting effect? It is quite possible that making more than one pretest guess would lead to an increase in competition for the correct answer. If a participant makes three incorrect guesses instead of one guess, those are three potential competitors which might be remembered instead of the correct answer when it comes time for the final test. A *source monitoring* interpretation of this potential competition would suggest that more guesses could lead to more potential confusion over which item was the correct answer and which items were the guesses. A *cue overload/response competition* interpretation would argue that making additional guesses beyond a single guess would lead to worsened performance on a final test due to an increase in the "load" (potential responses) placed on the "cue" (pretest question). It is also certainly possible that difficulties due to source monitoring could occur simultaneously with those due to cue overload/response competition.

However, on the other hand, the three theories of pretesting described previously would likely all predict that making more guesses would lead to an increase in the benefits of pretesting, though for different reasons. The Retrieval Effort Hypothesis (e.g., Pyc & Rawson, 2009) argues that more effort can lead to better learning, and therefore would argue that the three guesses would be more effective than one guess because of the amount of effort involved. The Episodic Context Account (e.g., Lehman, Smith, Karpicke, 2014) argues that the effects of pretesting are beneficial due to the episodic trace that is created when making a pretest guess, and therefore would argue that the three guesses would be more beneficial than one guess due to the creation of a stronger, more memorable episodic trace with three guesses. Finally, the Elaborative Retrieval Hypothesis (e.g., Carpenter, 2009) argues that pretesting is beneficial because of semantic elaboration that occurs in the mind when attempting to answer a pretest question and therefore three guesses would be more beneficial than one guess due to increased number of potential semantic mediators to create a connection between the question and answer.

Prior work manipulating the number of guesses in response to a pretest has been very limited. In experiment 1 of Vaughn and Rawson (2012), participants guessed a target word in response to a cue word or read an intact cue-target pair one or three times. In this case, they found that performance was actually better in the read trials compared to the pretest trials, in a reversal of the pretesting effect that was not moderated by the number of trials. Notably, overall performance collapsed across read and pretest trials in the three trial conditions outperformed overall performance for the one trial conditions. Subsequent experiments in this paper found benefits of pretesting, but did not return to the one vs. three item manipulation, so it is difficult to conclude much from this study about how number of guesses might moderate the pretesting effect.

Another study on number of generated items can be found in the testing, rather than pretesting literature. Lehman and Karpicke (2016, experiment 5) tested the idea that generating related items might affect memory for target items. In order to test this, participants first learned a list of target words (e.g., earth). Then, participants were asked to generate between two and six items (one, two, three, four, or five non-target associates plus a target from the first phase) in response to a cue word and a two-letter stem (e.g., globe- ea_____, wo_____). Finally, participants took a cued recall test of the original list of target words. They found that the number of non-target items generated in the second phase led to a negative linear trend, such that more items generated led to worse performance on the final test of memory for the target items. The researchers make the argument here that this is evidence that generating more possible answers in response to a target item is *worse* for memory rather than improving it. It is important to note that this research was conducted with a *testing* rather than *pretesting* paradigm. It is quite possible that with a pretesting paradigm, in which it makes more sense for non-target items to be generated due to the fact that the target information is not yet studied, generating more items might lead to a greater benefit than generating fewer items.

The Present Experiments

Across three experiments, the benefits of pretesting with multiple guesses are compared to benefits of making a single guess. In the first experiment, I begin with straightforward comparison of the effects of making one pretest guess, three pretest guesses, and a read-only control on final cued-recall test performance. In the second experiment, half of the participants experience the same procedure as the experiment one participants, while the other half are yoked to experiment 1 participants and, rather than making guesses themselves, are exposed to the guesses made by the prior participants. In a third experiment, the procedure is largely the same as the first experiment, with the exception that the final test is multiple-choice rather than cued-recall. The goal of all three of these experiments is to determine the effects of making one versus multiple guesses on the benefits of pretesting.

While the three theories of pretesting discussed previously (Retrieval Effort, Elaborative Retrieval, Episodic Context) would all predict that additional pretest guesses should lead to a larger learning benefit compared to a single pretest guess, research on the competition in memory (i.e., Source Monitoring, and/or Cue Overload/Response Competition) leads to the prediction that more guesses would interfere with memory for the correct answer and therefore be worse for learning.

Experiment 1

Methods

Participants and Design

A total of 92 participants (Mean Age = 20.1, Range = 18 – 33; 70 Female, 19 Male, 3 Nonbinary, Other, or Declined to State) were recruited from UC Santa Cruz's experiment website, SONA systems for partial credit in an undergraduate-level psychology course. This experiment employed a one by three (Activity: One pretest guess vs. Three pretest guesses vs. read-only control) within-subjects design. A power analysis assuming a small-to-medium effect size ($d = .3$) suggested the need for 90 participants in order to obtain 80% power in order to be able to detect a difference between performance on the 1-guess vs. 3-guess conditions.

Materials

Materials are 48 trivia questions designed to be readily guessable and yet difficult to answer *correctly*. For example, with the question "Which country produces the most vanilla?" (Answer: Madagascar), undergraduates are not particularly likely to answer this question correctly from memory, but should readily be able to come up with three countries as guesses. The goal of these materials is to ensure that participants are reasonably able to make three guesses when instructed to do so. These materials were selected largely from trivia materials used in Kornell (2014) and the pool created by Fastrich, Kerr, Castel, and Murayama (2019). Criteria for selection were the difficulty of the questions and the set size of the target category. With the above example, the target category is "country", which has a set size in the hundreds. It can be reasonably expected that participants would know enough countries to be able to make three guesses. In curating the selection of trivia questions for use in this study, a goal was to select items with as large of a set size as possible. While some target category set sizes are somewhat amorphous (e.g., "thing to be measured"), the minimum set size allowed for inclusion where it was clearer was nine (i.e., "planets"). An additional goal in putting these materials together was to not have too many items with the same target category. Specifically, no more than three questions from any single target category were included. A full set of the 48 trivia questions along with their correct answers and the target categories can be found in Table 1.

Procedure

This study was run entirely over Qualtrics, the survey website. Participants received a link through the SONA website and then were directed to read and digitally sign a consent form. Once this was completed, participants read the following set of instructions:

> “You will now encounter a series of trivia questions. For some, you will need to make one or three guesses. Others will simply be presented with the answer. For the guess items, do not worry if your guesses are correct, but you will need to make the required number of guesses before moving on. We ask that you take these guesses seriously and make sure that they make sense based on the question. For example, if the question asks "Which city....?", your answer(s) should be one or three different cities. Please DO NOT use any external sources to help you answer these questions. If you have fully read and understand these instructions, type "yes" below.”

After typing “yes”, participants moved on and began the initial learning/pretest phase. During this phase, participants were presented with a set of 48 questions. One third of these questions were shown with the correct answer immediately following the question. Another third of the questions were shown with the category name and a single space for one guess as the answer to the pretest question. For example, for the question, “What animal’s milk does not curdle?”, on the screen directly below the question, participants saw “Guess 1 (animal): ____________”. For the final third of the questions in the initial learning phase, participants were expected to make three guesses instead of one and were shown the question with three lines, each restating the category name as with the single guess questions. For these one- and three-guess items, once participants were done making their guess(es) and hit the "next" arrow, they were then shown the question with the correct answer, (for the above example, "camel").

The 48 items were quasi-randomly intermixed and counterbalanced such that each participant would be equally likely to see each question as a read-only control, single guess item, or a three-guess item. The questions themselves were split into three "groups" for counterbalancing purposes and then maintained in a single, quasi-random, fixed order for all participants for both the initial pretest phase and the final test phase. This quasi-random order was such that no more than two items from the same "group" would appear in order. Further, for questions which had multiple items from the same target category (i.e., "country"), these questions were split up within the quasi-random order.

After a ten-minute distractor task (Tetris), participants moved on to the final test. They were told, "You will now take a test on the trivia questions you encountered previously. Do your best to remember and write down the correct answers." The test was entirely cued-recall, with the trivia questions presented in the same fixed quasi-randomized order as the initial test. The questions were presented one question at a time with a single blank below it. The test was self-paced, with participants being able to hit the “next” arrow to move to the next question, but participants were not allowed to go back to previous questions. After the end of this test, participants were given a short survey including demographics questions, a question asking if they had used any external resources (such as the internet) to help them answer the questions, and an open-ended question for any comments or concerns they might have had about the study.

Results

Initial Test Performance

Initial “guess” items were coded for how often the answer was correct for one-guess and three-guess items. In order to compare these conditions, the three guess items were coded as correct if *at least one* of the three guesses were correct, without any additional credit for re-writing the correct answer multiple times. Initial test performance on the three-guess items ($M = 0.21$, $SD = 0.17$) was significantly higher than initial performance on the one-guess items ($M = 0.15$, $SD = 0.17$), $t(91) = 3.95$, $p < .001$, $d = .41$, 95% CI [0.20, 0.62].

An additional analysis of initial test performance was conducted comparing the one and three guess items, however, in this case, the three guess items were coded as correct only if the *first item* was correct as a way to create a clearer equivalency between the two conditions. In this case, initial test performance on the three-item condition ($M = .14$, $SD = .16$) was not significantly different from initial test performance on the one-item condition (*M* = .16, *SD* = .17), $t(91) = 1.87$, $p = .07$, d = .19, 95% CI [-0.01, 0.40].

Final Test Performance

A one-way repeated measures ANOVA was conducted to determine effects of Activity type (one guess, three guesses, and read only) on final test performance. See figure 1 for a graph of final test performance in experiment 1. A significant effect of activity type was found, $F(182, 2) = 45.24$, $MSE = .01$, $p < .001$, $\eta_p^2 = 0.36$. Follow-up pairwise comparisons showed significantly better final test performance on the one-guess items (*M* = 0.81, *SD* = 0.20) compared to the read-only items (*M* = 0.67; *SD* = 0.23), $t(91) = 7.54$, $p < .001$, $d = .79$, 95% CI [.55, 1.02], and on the three-guess items (*M* = 0.80, *SD* = 0.20) compared to the read-only items, $t(91) = 7.09$, $p <$ .001, $d = .74$, 95% CI [.51, .96]. No significant difference was found between final test performance on the one-guess and three guess items $t(91) = .04$, $p = .23$, $d = .13$, 95% CI [-.08, .33].

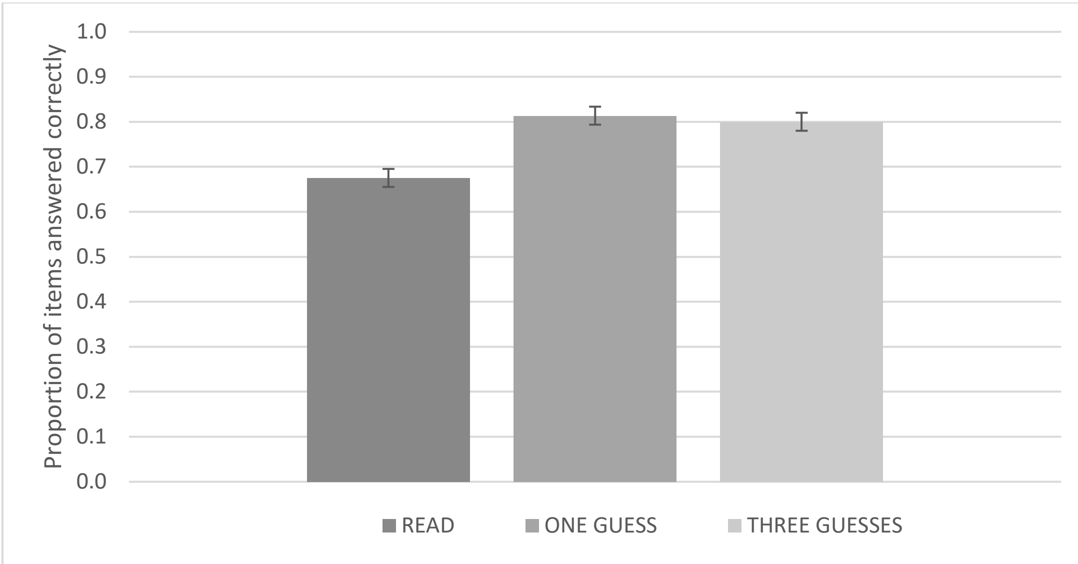

Figure 1. Proportion of items answered correctly on a final cued recall test for experiment 1 based on activity type (read, one guess, or three guess).

An additional set of analyses were conducted here and in subsequent experiments after removing any items that were answered correctly on the initial test. These analyses were conducted twice, once removing the three-guess items if *any* of the three guesses were answered correctly on the initial test, and once only removing the three-guess items from analyses if the *first* guess was answered correctly. Across both sets of analyses and all three experiments, no patterns of data or significance levels were changed from those conducted with all items included. Therefore, analyses reported here and in subsequent experiments are left to include all items.

Discussion

In this first experiment, the effects of making one or three pretest guess answers to trivia questions are compared to a read-only control. While a pretesting benefit was found on a final test of both the one guess and three guess items, no significant difference was detected between the magnitude of this benefit across the two conditions. Making multiple guesses as responses to pretest questions did not lead to any additional benefit or detriment when compared to the benefit of making a single guess.

From this first experiment alone, no conclusions can be made about whether additional guesses lead to additional benefits or an increase in competition. It is possible that there was *simultaneously* a benefit and a detriment that simply cancelled one another out. The benefits of pretesting need to be disentangled from the potentially detrimental effects of competition created by exposure to incorrect information. Directly comparing pretest conditions against matched conditions with the same number of guesses but not involving pretest guesses will give a clearer picture of exactly the magnitude of the pretesting effect found in this experiment.

Experiment 2

The main purposes of experiment 2 are firstly to replicate the results of experiment 1 and secondly to compare the benefits of making one or three pretest guesses with the effects of exposure to similar items that were not themselves generated by the participant. The goal here was to match the levels of competition facing the participants due to the items themselves and to isolate the effects of the generation of the guesses from the effects of competition due to the exposure to/generation of the guesses.

One previous study from the pretesting literature conducted by Grimaldi and Karpicke (2012, exp. 2) included a comparison condition in which participants studied a cue and a lure word (e.g., tide-wave) prior to studying the cue-target combination (tide-beach). This condition led to worse performance than control (though not significantly worse- $p = .11$), likely due to the increased competition from the non-target lure. This lure, despite its similarity in nature to the pretest guess, with the lure being an *externally* generated non-target item, and a pretest guess being an *internally* generated non-target item, led to a detriment in performance at final test. However, the pretesting condition still showed a benefit, despite any competition that may have arisen due to the guessed answer to the pretest question.

The present experiment differs from Grimaldi and Karpicke's in several critical ways. First, rather than the word pairs, the present experiment once again was conducted with trivia questions. It is unclear whether exposure to non-target lures might have the same detrimental effect on later memory performance that was demonstrated with word pairs. Second, in order to match the competition in the pretest condition with the competition in the non-pretest condition as closely as possible, a yoked research design was used here. Each new experiment 2 participant saw the full set of pretest guesses from a previous participant from experiment 1. Third, and perhaps most critically for our purposes, the current experiment also included conditions in which the effects of *multiple* pretest guesses can be compared to that of *multiple* lures.

If worse performance is found in the yoked three-item condition when compared to a one-item condition, then one can reasonably make the argument that there is increased competition (whether from source monitoring errors and/or cue overload/ response competition) from the three items both in the yoked condition and in the pretest condition. Critically, if this experiment once again finds *no additional benefit* in the three-item pretest condition (as in experiment 1) but simultaneously shows a *decrease in performance* in the parallel yoked three-item condition, one can make the argument that there may be a greater benefit of pretesting with multiple items that is being masked by a corresponding increase in competition. This pattern of data would mean that exposure to the multiple possible answers as responses to trivia questions would lead to increased competition and worsened performance, but this is being counteracted in the pretest guess conditions due to a simultaneous increase in benefit due to the generation of multiple items.

Alternatively, if no differences are found between the three item conditions in *both* the pretest and the yoked participants' final test scores, one can say that no evidence is found to support the idea that competition from pretest guesses is affecting final test performance in this experimental paradigm. This would not be to say that no competition whatsoever is occurring in this study, but rather that we do not have any evidence to suggest that competition due to exposure to the three items is affecting final test performance as compared to exposure to one pretest guess item.

Methods

Participants and Design

A total of 192 participants (Mean Age = 20.17, Range: 18-30; 141 Female, 43 Male, 8 Nonbinary, Other, or Decline to State) were recruited from UC Santa Cruz's experiment website, SONA systems for partial credit in an undergraduate-level psychology course. The design for this experiment was a 3 (Within- Activity: One pretest guess, three pretest guess, or read-only control) by 2 (Between- Pretest condition or Yoked condition).

Based on the same power analysis conducted for experiment 1, the goal was to collect 90 participants per between-subjects condition for a total of 180. 92 participants were collected in the Pretest condition and 100 participants were collected in the Yoked condition. Due to an error in the data collection system, 8 of these Yoked participants were duplicates and were yoked to the same experiment 1 data as another experiment 2 participant. All analyses were conducted both with and without these 8 participants, and no patterns of data or significance levels were changed. Thus, analyses are reported with these 8 participants included.

Materials and Procedure

Materials and procedure are identical to experiment 1 with one major exception. Half of the participants received an identical protocol and identical materials to experiment 1. The other half of the participants were each yoked to an experiment 1 participant and exposed to the prior participant's guesses instead of making their own pretest guesses. In the "pretest" phase this yoked group of participants saw the pretest questions alongside the one or three (often incorrect) guesses from the previous participant. Each new participant was matched to a single participant from experiment 1. Despite the fact that some items were left blank on the initial experiment 1 pretests or had odd or irrelevant answers, all items were shown as they were written to their matched yoked participants. Yoked participants were instructed to re-type the one or three guesses made by the previous participant. The full set of instructions given to the yoked participants was as follows:

> “You will now encounter a series of trivia questions and guesses made by a PRIOR PARTICIPANT. For some, you will see one or three guesses made by the previous participant. Others will simply be presented with the answer. For the guess items, simply retype the guesses made by the prior participant. Some of these guesses may be blank or seem strange in some way. Regardless, retype the answer given by the previous participant. For example, you may see an item saying: "Guess 1 (city): [TALLAHASSEE]" Your job is to retype "Tallahassee".

If you have fully read and understand these instructions, type "yes" below.”

After completing the yoked pretest phase, these participants completed the same Tetris distractor task and final tests as was given to the other half of the participants and the experiment 1 participants.

Results

Initial Test Performance

Similar to experiment 1, initial test performance was first analyzed based on a comparison between one-guess and three-guess items for the pretest participants only (yoked participants were simply copying prior participants guesses, and therefore were not analyzed here). Three-guess items were coded as correct if at least one of the three guesses were correct. We found a significant difference such that performance on the three-guess items ($M = 0.16$, $SD = 0.16$) was significantly higher than performance on the one-guess items ($M = .12$, $SD = 0.14$), $t(92) = 3.43$, $p < 0.001$, $d = 0.36$, 95% CI [0.15, 0.57].

Final Test Performance

For final test performance, a three-by-two mixed ANOVA was conducted comparing Activity (Within: 1 vs. 3 vs. read only) with participant Group (Between: pretest vs. yoked). See figure 2 for a graph of final test performance in experiment 2. Significant main effects of both activity, $F(1, 192) = 140.44$, $MSE = 2.62$, $p <.001$, $\eta_p^2 = .42$ and group, $F(1, 192) = 35.50$, $MSE = 4.32$, $p <.001$, $\eta_p^2 = 0.16$ were found. Additionally, a significant interaction between Activity and Group was found, $F(2, 384) = 23.80$, $MSE = .37$, $p < .001$, $\eta_p^2 = .15$, such that there was a greater pretesting benefit of the one or three guess items relative to the read-only items in the Pretest condition as opposed to the Yoked condition.

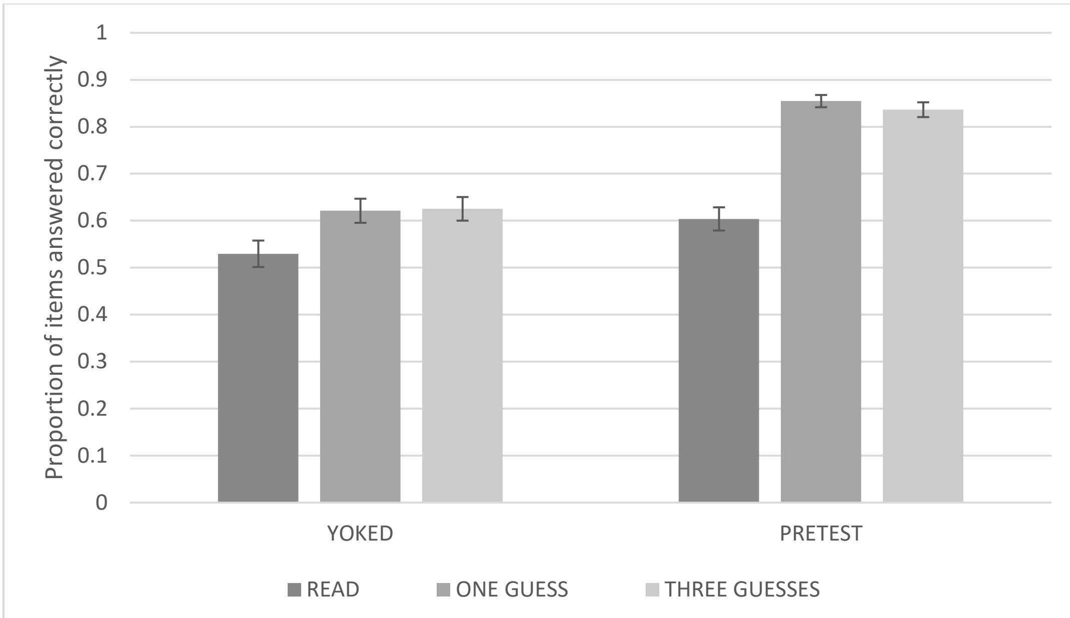

Figure 2. Proportion of items answered correctly on a final cued recall test for experiment 2 based on activity type (read, one guess, or three guess) and group (pretest or yoked).

Follow-up analyses for the Yoked group showed significantly better performance on the one guess condition (M = .62, SD = .26) compared to the read-only condition (M = .53, SD = .28), $t(100)$ = 5.00, p < .001, d = .50, 95% CI[.30, .70] and better performance on the three guess condition (M = .63, SD = .25) compared to the read-only condition, $t(100)$ = 5.39, p < .001, d = .54, 95% CI[.33, .74]. However, no significant difference was found between the one guess and three guess condition, $t(100)$ = .53, p = .60, d = .05, 95% CI[-.14, .25].

A similar pattern was observed for the Pretest group. Significant differences were found between the read-only condition (M = .60, SD = .24) and the one guess condition (M = .85, SD = .12), $t(92)$ = 11.42, p < .001, d = 1.18, 95% CI[.92, 1.45] as well as the read-only condition and the three guess condition (M = .84, SD = .15), $t(92) = 10.83$, $p < .001$, $d = 1.12$, 95% CI[.86, 1.38]. However, no significant difference was found between the one guess and three guess condition, $t(92) = 1.22$, $p = .23$ $d = .13$, 95% CI[-.33, .08].

Discussion

Similar to experiment 1, once again a benefit was found of pretesting compared to control in both the three-guess and one-guess conditions. In the Yoked conditions, we find a benefit of exposure to prior participants' guesses, though this was a much smaller benefit when compared to that of the pretesting conditions (~9 percentage point benefit yoked versus ~25 percentage point benefit pretest). The fact that this exposure to pretest guesses led to any benefit at all is somewhat surprising given that participants did not have to generate anything and given the trending detriment found by Grimaldi and Karpicke (2012). It is possible that the yoked participants may have been engaging in some form of a covert retrieval attempt (Smith, Roediger, & Karpicke, 2013) while reading and retyping the prior participants' guesses, and it was this covert retrieval that led them to benefit from the exposure to the guesses. Alternatively, the small benefits could stem more directly from a benefit due to the exposure to the prior participants' guesses.

Critically for this experiment, we also found no difference between the benefits of three guesses and one guess in either the pretest or yoked conditions. The lack of difference was once again found despite predictions based in theories of pretesting arguing that the benefits of pretesting would be *increased* by multiple guesses and predictions based on competition mechanisms found in other areas of the study of memory to expect that the benefits of pretesting would be *decreased* by multiple guesses.

With the yoked conditions, the fact that exposure to three guesses does not lead to worsened performance relative to exposure to one guess is somewhat surprising. If exposure to the greater number of pretest guess items was leading to more competition, then one would expect worsened performance in this condition, regardless of whether there was worsened performance in the parallel pretest condition. The primary reasoning for the inclusion of this yoked condition was to create a parallel for the pretest condition and to be able to isolate the effects of making one or three guesses from any competition due to the exposure to the guess items themselves. If the three guesses were leading to simultaneously better memory and increased competition, but the two mechanisms were cancelling one another out and leading to no differences in the pretest conditions, one would expect this to be reflected by worsened performance in the parallel one-item yoked condition. Taken all together, without any differences between the three- and one-item conditions for either the pretest or yoked participants, no evidence of competition affecting the benefits of pretesting is found in this experiment.

Experiment 3

In the first two experiments, no evidence was found for the idea that additional pretest guesses might lead to improved or worsened performance. Experiment 3 employs another tactic to determine if competition might affect memory performance following the generation of one or multiple guesses.

In this experiment, the final test type was changed from cued-recall to multiple-choice as a way to reduce the extent to which competition might lead to a reduction in performance. If strengthening incorrect pretest guesses is leading to an increase in competition and is still masking an additional benefit of generating a greater number of items, one can change the test type in order to reduce the degree to which interference can affect the results. By making the final test a multiple-choice recognition test, the likelihood is reduced that the strengthening of the pretest guesses during the pretest will lead to an increase in competition between those guesses and the correct answers on the final test. With a multiple-choice test, if there is competition between the incorrect guess and the correct answer, the incorrect guess is unlikely to be identical to the lure options provided, and therefore is less likely to be reproduced in place of the correct answer.

An additional factor added to experiment 3 was that of timing on the final test. Timing for each type of item (three guess, one guess, or read-only) was recorded and analyzed. The goal here was to (1) provide a more sensitive element of analysis and (2) provide a backup in case of ceiling effects with the easier, recognition final test. If there are differences between one guess and three guess difficulty or competition that is not being captured by final test performance, it is possible that analyzing timing data may show such differences. Specifically, if greater competition occurs due to making three guesses during the initial test, this may lead participants to take longer to answer the multiple-choice item, regardless of whether the participant eventually selects the correct answer.

Methods

Participants and Design

A total of 92 new participants (Mean Age: 19.4; Range: 18-25; 71 Female, 21 Male, 1 Nonbinary, Other, or Declined to State) were recruited from UC Santa Cruz's experiment website, SONA systems for partial credit in an undergraduate-level psychology course. Similar to experiment 1, we employed a one by three (Activity: one pretest guess vs. three pretest guesses vs. read-only control) within-subjects design.

Materials and Procedure

The materials and procedure for experiment 3 were identical to experiment 1 up until the final test. Participants were told they would be timed and to take the final test as quickly and accurately as possible. They then began a multiple-choice test with five total choices including the one target answer and four non-target lures. The four lures were collected from the experiment 1 initial test responses. The goal in the creation of these lures was to select items that were very low frequency responses but still made sense as potential answers. Most items selected were under 1% of total responses from the experiment 1 pretest phase, however this was not possible for certain items such as “What is the only planet in the solar system to rotate clockwise? Answer: Venus”. For this item, as the target category (planet) has a relatively small set size compared to many of the other items, we had to use lures that were more commonly produced as guesses from experiment 1. The full set of lures for each question can be found in Table 1.

During the final test, a clock was displayed to participants as a method of ensuring that they were aware that they were being timed and as a reminder to move quickly through the test. Otherwise, this experiment was identical to experiment 1.

Results

Initial Test Performance

Similar to prior experiments, we found a significant difference in initial test performance such that performance on the three-guess items (M = 0.26, SD = 0.22) was significantly higher than performance on the one-guess items (M = .16, SD = 0.20), $t(92) = 7.50$, $p < 0.001$, $d = 0.78$, 95% CI [0.54, 1.00].

Final Test Performance

A one-way repeated measures ANOVA was conducted to determine effects of Activity type (one guess, three guesses, and read only) on final test performance. See figure 3 for a graph of final test performance in experiment 3. A significant effect of activity type was found, $F(2, 184) = 52.97$, $MSE = .01$, $p < .001$, $\eta_p^2 = .34$. Follow-up pairwise comparisons showed significantly better performance on one-guess items (*M* = .95, *SD* = .10) relative to read-only items (*M* = .83, *SD* = .17) $t(92) = 7.45$, $p < .001$, $d = .77$, 95% CI[.54, 1.00], as well as three-guess items(*M* = .96, *SD* = .09) relative to read-only items, $t(92) = 7.73$, $p < .001$, $d = .80$, 95% CI[.57, 1.03]. However, similar to prior experiments, no significant difference was detected between the one guess and the three guess items, $t(92) = 1.31$, $p = .20$, $d = .14$, 95% CI[-.07, .34].

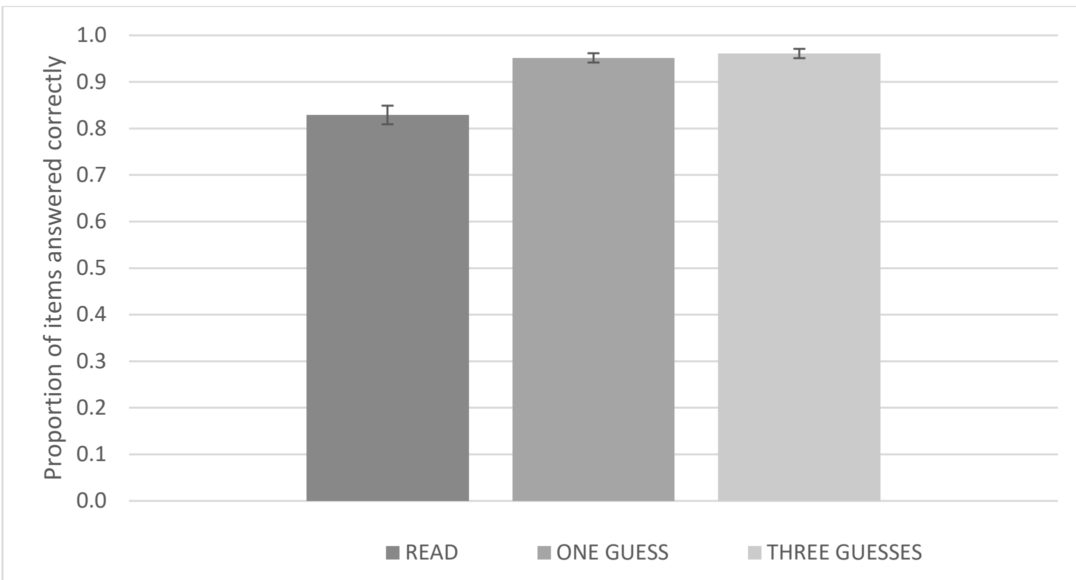

Figure 3. Proportion of items answered correctly on a final multiple-choice test for experiment 3 based on activity type (read, one guess, or three guess).

Final Test Timing

Another one-way repeated measures ANOVA was conducted to determine effects of Activity type (one guess, three guesses, and read only) on final test timing. See figure 4 for a graph of final test timing in experiment 3. Timing here was measured in seconds from the time that the question appeared on screen to the time that the page was submitted. An additional analysis was conducted here with time to submit the page for only the items that were answered correctly. The removal of incorrect items did not change any significance levels and therefore we report analyses without these items removed.

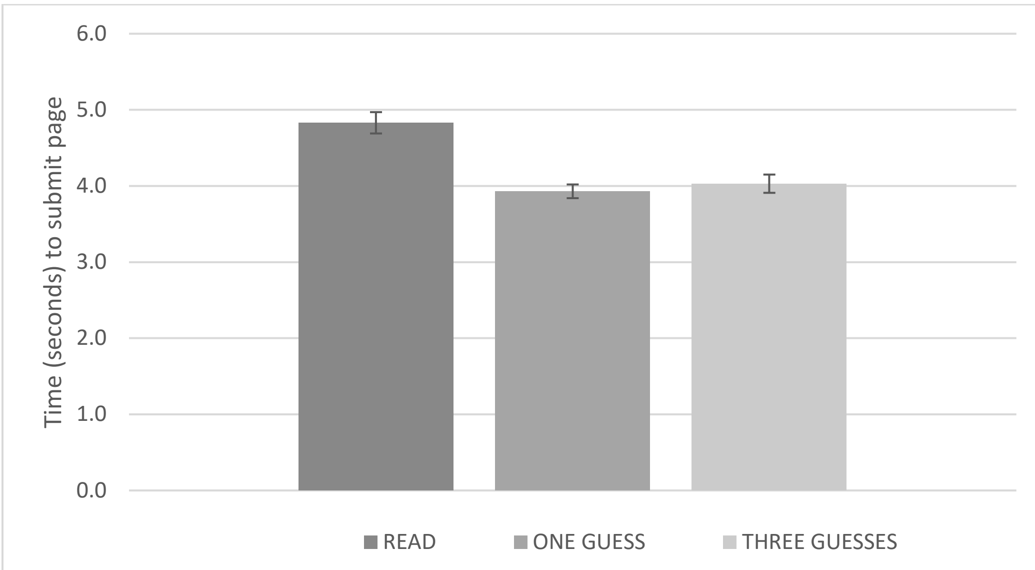

Figure 4. Time in seconds to submit answers by item on a final multiple-choice test for experiment 3 based on activity type (read, one guess, or three guess).

A significant effect of activity type on final test timing was found, $F(2, 184) = 51.17$, $MSE = .43$, $p < .001$, $\eta_p^2 = .36$. Follow-up pairwise comparisons showed a significantly greater amount of time to submit read-only items ($M = 4.79$, $SD = 1.33$) relative to one-guess items ($M = 3.91$, $SD = 0.85$), $t(92) = 9.28$, $p < .001$, $d = .96$, 95% CI[.71, 1.21], as well as read-only items relative to three-guess items ($M = 4.01$, $SD = 1.15$), $t(92) = 7.51$, $p < .001$, $d = .78$, 95% CI[.55, 1.01]. However, no significant difference was detected between timing for the one-guess and the three-guess items, $t(92) = 1.12$, $p < .001$, $d = .12$, 95% CI[-.08, .32].

Discussion

In experiment 3, as with the prior experiments, no evidence was found to suggest any difference in benefits or competition between making multiple guesses as compared to a single guess in response to a trivia pretest. A benefit of pretesting was found in the three-guess and one-guess conditions relative to control. Performance on the final test for the three-guess and one-guess conditions was near ceiling (96% correct for one-guess items and 95% for three-guess), which is unsurprising due to the easier, recognition based final test.

Final test timing for this experiment was also recorded and analyzed in order to provide a more sensitive measure of analysis. The pattern found here is reflective of that found in the performance data. Specifically, participants took significantly longer to answer the read-only questions relative to the three-guess and one-guess questions, but no significant difference was detected in the amount of time taken to answer the three-guess and one-guess items. If the act of making three-guesses in response to a trivia question leads to more competition than making one-guess, but this competition was not leading to worsened performance at final test, this timing measure would have been likely to demonstrate a difference here such that participants would take longer on the three-guess items relative to the one-guess items.

General Discussion

The pretesting effect, or the benefits of pretesting for later test performance, is well established (e.g., Grimaldi & Karpicke, 2012; Kornell, 2014; James & Storm, 2019; Richland et al., 2009; St. Hilaire, Carpenter, & Jennings, 2021). The present experiments expand upon this literature with a consideration of the ways that competition and multiple guesses might factor into the benefits of pretesting. Specifically, the goals of the present research were (1) to determine the effects of making multiple guesses on the pretesting effect, and (2) to determine the extent to which competition from guesses might affect pretesting benefits. Across the three experiments reported here, final test performance is compared for trivia questions to which learners responded with one pretest guess, three guesses, or no guesses in a read-only control. Across all three experiments, significant benefits of pretesting were found with one or three guesses compared to a read-only control. However, no differences were found between the benefits of making three guesses as opposed to one guess.

We find no evidence that making pretest guesses leads to competition between the pretest guess and the correct answer, whether that competition be from source monitoring errors (e.g., Johnson et al., 1993), cue overload/response competition (e.g., Watkins & Watkins, 1975, Proctor, 1981), or any other potential source of competition. This competition was not detected even when pretest guesses were presented to participants rather than being generated (experiment 2) and also was not detected when response times were measured and compared across conditions during final test (experiment 3). Whether this means that making pretest guesses simply does not lead to competition remains to be seen. It is possible that competition occurred in our study but was not detected, or alternatively competition may occur due to pretest guesses under other circumstances, such as in a classroom, with other materials, or perhaps with a greater number of guesses. Future work will be critical to testing the hypothesis that guesses generated by oneself simply do not lead to competition in memory for the correct information.

We also find no evidence of the benefit predicted by theories of pretesting such as the Retrieval Effort Hypothesis (e.g., Bjork & Bjork, 2011; Pyc & Rawson, 2009), the Elaborative Retrieval Hypothesis (e.g., Carpenter 2009, 2011), and the Episodic Context Account (e.g., Lehman et al., 2014). While the present set of experiments was not designed to differentiate between these theories of pretesting, the consistent lack of benefit for the three guesses relative to the one guess conditions poses perhaps the greatest challenge to the Elaborative Retrieval Hypothesis, which makes the strongest predictions surrounding the effects of multiple guesses of the three accounts. The central idea of the Elaborative Retrieval Hypothesis is that pretesting is beneficial due to learners engaging in semantic elaboration in response to a pretest question, searching their minds for the answer, and strengthening a semantic network connecting the question and answer. Then, at a later test, when presented with the question again, this strengthened semantic network can help the learner trace a route back to the correct answer. It follows from this that making multiple guesses should only further strengthen the semantic network and therefore lead to better performance at a later test. The fact that we did not find evidence of any additional benefit of pretesting with multiple guesses poses a problem for this theory. This also fits with the lack of benefit found by Lehman and Karpicke (2016) discussed previously. These researchers found a lack of benefit of studying items related to a target lure in a posttesting paradigm and found that more lures led to worse final test performance. The researchers here argued that this detriment was evidence against the central tenets of the Elaborative Retrieval Hypothesis- that mediators should improve testing benefits.

In addition to advancing understanding of the pretesting effect, this research can benefit learners and teachers from a practical standpoint, when deciding when and how to implement pretesting as a learning strategy. In a classroom setting, students make guesses prior to being given the information necessary to answer such questions correctly. Such a practice is exceedingly effective for improving later memory performance for the pretested information. However, what happens if a student makes *multiple* erroneous guesses? Thankfully for educators, this research did not find any evidence that making multiple guesses in response to a pretest question leads to increased competition or worsened performance. Further, the pretesting benefit, as in previous studies (e.g., Richland et al., 2009) occurred regardless of whether final test analyses were restricted to items that participants answered incorrectly on initial tests.

Additionally, when it comes to classroom benefits, it is also important to understand how incorrect lures might affect learners' memory for the correct information. If a teacher asks a pretest question to the class and one student answers it incorrectly, this may lead to a negative memory effect for the surrounding students. The research in our second experiment which found benefits to reading and retyping prior participants' guesses suggests that exposure to incorrect guesses in response to pretest questions may actually lead to small benefits, even if the guesses are made by other people. More research will certainly need to be done in order to determine any boundary conditions of this specific benefit. For example, the re-typing of the prior guesses might be critical to gaining the benefit of exposure to other participants' guesses.

In other future directions for this research, it will be important to determine situations in which making multiple pretest guesses may lead to beneficial or detrimental effects on memory. For example, it is entirely possible that pretesting with multiple guesses may lead to better or worse later memory performance in an actual classroom setting, with different lengths of delay between the pretest and final test, or when different types of materials are used. It is feasible that making multiple guesses as to the second word in a cue-target pair may be more likely to lead to competition, perhaps due to cue-overload. As we do not manipulate materials here, we will not be able to speak to any effects different materials might have on pretesting and competition. Additionally, the present experiments had learners be *required* to make multiple guesses rather than *choosing* to do so. It may be the case that making multiple guesses is more beneficial or detrimental if those additional guesses were made because the learners actually think those guesses might be correct as opposed to if they are simply presenting guesses as possibilities to fulfill a requirement.

References

Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. In *Psychology of learning and motivation* (Vol. 9, pp. 89-132). Academic Press.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, *21*(8), 627-635.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, *2*(59-68).

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly journal of experimental psychology*, *10*(1), 12-21.

Brown, J., & Brown, M. W. (1990). The effects of repeating a recognition test in lorazepam-induced amnesia: Evidence for impaired contextual memory as a cause of amnesia. *The Quarterly Journal of Experimental Psychology Section A*, *42*(2), 279-290.

Brown, A. S., & Murphy, D. R. (1989). Cryptomnesia: Delineating inadvertent plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 432-442.

Burke, D. M., & Light, L. L. (1981). Memory and aging: the role of retrieval processes. *Psychological bulletin*, *90*(3), 513-546.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563-1569.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547-1552.

Carpenter, S. K., Rahman, S., & Perkins, K. (2018). The effects of prequestions on classroom learning. *Journal of Experimental Psychology: Applied*, *24*(1), 34-42

Carpenter, S. K., & Toftness, A. R. (2017). The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory and Cognition*, *6*(1), 104-109.

Chan, J. C., & McDermott, K. B. (2007). The testing effect in recognition memory: a dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 431-437.

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. Gregg (Ed.), *Cognition and learning* (pp. 117–138). New York, NY: Wiley.

Cyr, A. A., & Anderson, N. D. (2018). Learning from your mistakes: does it matter if you're out in left foot, I mean field?. *Memory*, *26*(9), 1281-1290.

Eriksen, C. W., Coles, M. G., Morris, L. R., & O'hara, W. P. (1985). An electromyographic examination of response competition. *Bulletin of the Psychonomic Society*, *23*(3), 165-168.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, *16*(1), 143-149.

Fastrich, G. M., Kerr, T., Castel, A. D., & Murayama, K. (2018). The role of interest in memory for trivia questions: An investigation with a large-scale database. *Motivation science*, *4*(3), 227-250.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding?. *Memory & Cognition*, *40*(4), 505-513.

Hamann, S. B., & Squire, L. R. (1995). On the acquisition of new declarative knowledge in amnesia. *Behavioral neuroscience*, *109*(6), 1027-1044.

Hartley, J., & Davies, I. K. (1976). Preinstructional strategies: The role of pretests, behavioral objectives, overviews and advance organizers. *Review of educational research*, *46*(2), 239-265.

Hayman, C.A., MacDonald, C. A., & Tulving, E. (1993). The role of repetition and associative interference in new semantic learning in amnesia: A case experiment. *Journal of* 00*Cognitive Neuroscience, 5,* 375–389

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 290-296.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & cognition*, *40*(4), 514-527.

James, K. K., & Storm, B. C. (2019). Beyond the pretesting effect: What happens to the information that is not pretested?. *Journal of Experimental Psychology: Applied*, *25*(4), 576-587.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin, 114*(1), 3-28.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-Based Learning: An Episodic Context Account. *Psychology of Learning and Motivation*, 237-284.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of verbal learning and verbal behavior*, *1*(3), 153-161.

Kole, J. A., & Healy, A. F. (2013). Is retrieval mediated after repeated testing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 462-472.

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 106-114.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"?. *Psychological science*, *19*(6), 585-592.

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.

Kornell, N., & Vaughn, K. E. (2016). How retrieval attempts affect learning: A review and synthesis. *Psychology of learning and motivation*, *65*, 183-215.

Lehman, M., & Karpicke, J. D. (2016). Elaborative retrieval: Do semantic mediators improve memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1573-1591.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787-1794.

Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition*, *43*(1), 14-26.

Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition*, *44*(7), 1085-1101.

Metcalfe, J. (2017). Learning from errors. *Annual review of psychology*, *68*, 465-489.

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia*, *138*, 107296.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of experimental psychology*, *58*(3), 193.

Proctor, R. W. (1981). A unified theory for matching-task phenomena. *Psychological Review*, *88*(4), 291.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory?. *Journal of Memory and Language*, *60*(4), 437-447.

Rantzen, A., & Markham, R. (1992). The reversed eyewitness testimony design: More evidence for source monitoring. *The Journal of general psychology*, *119*(1), 37-43.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning?. *Journal of Experimental Psychology: Applied*, *15*(3), 243-257.

Roediger III, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*(4), 382-395.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249-255.

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, *3*(4), 241-249.

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.

Skinner, B. F. (1958). Teaching machines. *Science*, *128*(3330), 969-977.

Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Memory and Language*, *22*(2), 153-163.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342-353.

Smith, M. A., Roediger III, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1712-1725.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*(9), 641-656.

St Hilaire, K. J., & Carpenter, S. K. (2020). Prequestions enhance learning, but only when they are remembered. *Journal of Experimental Psychology: Applied,* *26*(4), 705-716.

St. Hilaire, K. J., Carpenter, S. K., & Jennings, J. M. (2019). Using prequestions to enhance learning from reading passages: the roles of question type and structure building ability. *Memory*, *27*(9), 1204-1213.

Storm, B. C., James, K. K., & Stone, S. M. (2021). Pretesting can be beneficial even when using the internet to answer questions. *Memory*, 1-8.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643-662.

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and verbal behavior*, *1*(5), 321-334.

Vaughn, K. E., & Rawson, K. A. (2012). When is guessing incorrectly better than studying for enhancing memory?. *Psychonomic Bulletin & Review*, *19*(5), 899-905.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442-452.

Table 1. This chart includes all test questions used in all three experiments as well as the target categories for each item, correct answers, and the non-target lures shown to participants in the final multiple-choice test for experiment 3 only.

	Target category	Correct Answer	Non-Target Lures (Exp. 3 only)	Question
1	Cat	Cheetah	Lynx Jaguar Bobcat Sphynx	What is the only cat in the world that cannot retract its claws completely?
2	Animal	Camel	Bear Horse Pig Kangaroo	Which animal's milk does not curdle?
3	City	Damascus	Istanbul Jericho Pompeii Alexandria	What is the oldest inhabited city in the world?
4	Fish	Shark	Beta fish Piranha Marlin Pufferfish	What is the only fish that can blink with both eyes?
5	Music album	Thriller	The Dark Side of the Moon Lemonade Led Zeppelin IV Nevermind	What is the bestselling music album of all time?
6	Drink	Mocha	Kombucha Fanta Kefir San Pellegrino	Which drink gets its name from a town on the Red Sea coast of Yemen?
7	Sport	Curling	Shuffleboard Badminton Polo Racquetball	Which sport uses the terms "stones" and "brooms"?
8	Mammal	Giraffe	Hippo Hummingbird Squirrel Monkey	Which mammal has the highest blood pressure?

9	Body part	Cornea	Gallbladder Cartilage Eyelashes Retina	What is the only living part of the human body that has no blood supply?
10	U.S. president	Madison	Reagan Jackson Adams Taft	What is the last name of the shortest American president?
11	Vegetable	Onion	Broccoli Raddish Pumpkin Eggplant	On what vegetable did an ancient Egyptian place his right hand when taking an oath?
12	Language	English	Japanese Russian Polish Latin	What language has the largest vocabulary?*
13	Beatles song	Hey Jude	Yesterday Strawberry Fields Forever Come Together Penny Lane	What Beatles song remained the longest on the music charts?*
14	City	Rome	San Francisco Athens Constantinople Mumbai	What was the first city in the world to have a population of more than 1 million?
15	Color	Violet	Grey Yellow Green Blue	What is the color of mourning in Turkey?
16	Plant	Cotton	dandelion Orchid Clover Barley	What plant is attacked by the boll weevil?
17	Country	Monaco	Vatican City Indonesia Korea Malaysia	What is the country with the highest population density?
18	Bird	Ostrich	Turkey	What bird's eye is bigger than its brain?

			Toucan Flamingo Penguin	
19	Planet	Venus	Mercury Neptune Jupiter Saturn	What is the only planet in the solar system to rotate clockwise?*
20	Thing to be measured	Rainfall	Wind Speed Density Moisture Earthquakes	What is measured with an ombrometer?
21	Product	Coffee	Rice Salt Tobacco Plastic	Which product, after oil, is the most frequently traded product around the world?
22	Fish	sunfish	Catfish Seahorse Koi Sardine	Which fish can produce more eggs than any other known vertebrate?
23	Part of tree	Nut	Sap Bean Branch Stem	What part of a cola tree is used to flavor beverages?
24	Trade	Stonecutting	Carpentry Blacksmithing Mathematics Metalwork	What trade was Greek philosopher Socrates trained for?
25	Shopping item	Shoes	Glasses Rings Suit Watch	What are you shopping for if you're measured by a Brannock Device?
26	Body part	Tongue	Eyes Liver Lungs Brain	What is the fastest healing body part on a human?
27	Greek god	Apollo	Dionysus Hermes Pan	Who is the Greek God of music?

			Aphrodite	
28	Fruit	Plum	Persimmon Kiwi Peach Guava	What type of fruit would you pick from a Mirabelle tree?
29	Sport	Table Tennis	Diving Figure Skating Long Jump Gymnastics	Yapping Deng was a world champion in which sport?
30	country	Madagascar	Kenya Argentina South Africa Zimbabwe	What nation produces two thirds of the world's vanilla?
31	Spice	Nutmeg	Saffron Thyme Cardamom Coriander	What spice is extremely poisonous if injected intravenously?
32	U.S. state	Virginia	Montana Maine Kansas Wisconsin	In which state were the first peanuts in the United States grown?
33	National Park	Yellowstone	Glacier Niagara Falls Big Sur Zion	What was the world's first National Park?*
34	Insect	Firefly	Moth Grasshopper Ladybug Mosquito	What insect depends the most on sight, rather than sound, to locate mates?
35	Berry	Mulberry	Gooseberry Elderberry Cherry Boysenberry	Ingesting large amounts of what type of unripe berry can cause moderate hallucinations?
36	City	Hong Kong	Dubai Beverly Hills Miami Las Vegas	What city has the most Rolls-Royce per capita?

37	Flower	Carnation	Lilac Marigold Dahlia Petunia	What is Spain's national flower?
38	Liquid	Alcohol	Blood Vinegar Cyanide Iodine	Which liquid were thermometers filled with in the 17th century, before mercury?
39	Venomous snake	King Cobra	Diamondback Banana Snake Black Mamba Cottonmouth	What is the longest venomous snake?
40	U.S. city	Philadelphia	Detroit Cincinnati Jacksonville St. Louis	In what U.S. city was the first U.S. zoo built?
41	Plant	Ivy	Fern Ficus Mint Grass	What is the more common name of the plant Hedera?
42	Product	Beer	Light Bulb Car Fridge Phone	What was the first trademarked product?
43	Dog breed	Great Dane	Boxer Doberman Rottweiler Greyhound	Scooby Doo is based on what breed of dog?*
44	Company	Xerox	IBM Lenovo Logitech Cisco	What company was the first to offer a mouse on a commercially available computer?
45	Disability	Deafness	Nearsightedness Dementia Polio Diabetes	What disability did Thomas Edison suffer from?
46	Body part	Wrist	Ear Hip	Where in the body would you find the pisiform bone?

			Foot Knee	
47	Flavor	Orange	Caramel Strawberry Rosemary Chestnut	What flavor is the liqueur Cointreau?
48	U.S. state	Alaska	Rhode Island Delaware Oregon Connecticut	What U.S. state has the highest percentage of people who walk or bike to work?