2023

The effect of testing on new learning of related and unrelated text sections

Katie Ingram
University of North Florida, katieingram20@gmail.com

THE EFFECT OF TESTING ON NEW LEARNING OF RELATED AND UNRELATED TEXT SECTIONS

By

Katie C. Ingram

A thesis submitted to the Department of Psychology in partial fulfillment of the requirements for the degree of Master of Science in Psychological Science

UNIVERSITY OF NORTH FLORIDA

COLLEGE OF ARTS AND SCIENCES

May, 2023

TABLE OF CONTENTS

LIST OF FIGURES

Abstract

When individuals are presented with a variety of materials, including word lists, face-name pairs, text passages, and more, the presence of a test between sections can enhance future section learning, a phenomenon called the forward testing effect (FTE). In addition to the FTE, studies have suggested that a decrease in the relatedness of the subject matter units can increase learning of the material. The current study examined the interaction between the presence of a test and the relatedness of material using text sections and cued-recall questions. Participants were 119 individuals assigned to a related test, related no-test, unrelated test, or unrelated no-test condition. All participants received a criterial test, or a test on their fourth section. Correct responses on these tests were used to compare performance across conditions. No significant difference between conditions was found, indicating the lack of a FTE. Although the manipulation check indicated that the manipulation of relatedness was successful, future studies should be conducted to examine the impact of relatedness on materials in which the FTE is established and a robust effect. This research will be important for providing recommendations to researchers and educators regarding the efficacy of testing on longer text sections.

The Effect of Testing on New Learning of Related and Unrelated Text Sections

On average, students in the United States take over 112 standardized tests by the time they graduate high school (Strauss, 2015). This number does not account for the tests taken over material learned in classes throughout their education. Both these in-class tests and standardized tests are structured in a way that cause them to not be viewed by students as learning tools, but rather an event for which they must memorize information to be forgotten once it is over (Roediger & Karpicke, 2006). While this is a common view adopted by many students, test taking can be beneficial as a learning tool as students can apply self-testing to their study strategies and it can enhance memory for both tested and untested materials (Pastötter et al., 2011; Wissman & Rawson, 2011). With the prevalence of test-taking in educational settings, it is important to consider factors that affect performance and learning.

The forward testing effect (FTE) refers to the idea that testing individuals on studied information, including native and foreign word lists, recorded lectures, face-name pairs, and text passages can enhance learning for future material (Ahn & Chan, 2023; Szpunar et al., 2008; Szpunar et al., 2013; Pastötter & Frings, 2019; Weinstein et al., 2011; Wissman & Rawson, 2015). That is, the learning of new information is impacted by the presence, or lack thereof, of a test for previously learned information (Szpunar, et al., 2008). Individuals who are tested after each section of learning perform better on each subsequent test as well as on a final free or cued recall test when compared to those who are not tested. Szpunar et al. (2008) examined this phenomenon using lists of words. Participants in their experiments were informed that they would be asked to study five lists of 18 words, and they would be randomly given a free-recall test or a minute of math problems following each list.

In the free-recall tests, participants were asked to recall as many words as possible from the previous list (Szpunar et al., 2008). However, the presence of a free-recall test was not randomized. Instead, half of the participants were given one minute of math problems (a filler task), while the other half were tested following each list. All participants received a test after the final list, known as the criterial list, as performance on that test was evaluated to determine the impact of prior testing, or not testing, on the final list. If words from the past list(s) appeared in the free-recall of the criterial list, this was considered to be the presence of proactive interference. That is, material learned previously negatively impacted learning of the final list. Szpunar et al. (2008) found that individuals not tested after each of the first four lists recalled fewer words from list five on the criterial list than those tested after each list and recalled more words from previous lists, or proactive interference. Several studies have sought to explain why the FTE occurs, which has led to the proposal of various mechanisms, including the context change hypothesis, reduction of proactive interference, and the reset of encoding theory (Chan et al., 2018; Pastötter et al., 2017; Szpunar et al., 2008)

Context Change Hypothesis

One proposed mechanism as to how the FTE occurs is through the creation of a context through the presence of a test, or the context change hypothesis (Szpunar et al., 2008). The context in which memories are both made and stored can be either internal or external (Pastötter et al., 2008). While external contexts include locations, smells, objects, and events that can assist with triggering the target memory, internal contexts refer to moods and thinking patterns. With regard to the FTE, the presence of a test introduces context cues that can become associated with the previously learned information (Jang & Huber, 2008). The act of retrieving material after it has been studied alters the internal context that was associated as the information was learned,

increasing the ability for the individual to discriminate between the sections (Jang & Huber, 2008). With these specific context cues, individuals exhibit enhanced performance on tests for both newly and previously learned material (Pastötter et al., 2011).

Reduction of Proactive Interference

When words from a previous list, or lists, are recalled during free-recall of a criterial list, this is referred to as prior-list intrusions (Szpunar et al., 2008). This occurs more frequently when individuals are not tested compared to those who are tested between each list. This is an example of proactive interference during retrieval, or the idea that previously learned material negatively impacts learning for subsequent material (Wissman et al., 2011). In addition to improved test performance compared to restudy conditions where participants either receive a filler task or continue to the next material, testing is associated with lower levels of proactive interference through prior-list intrusions (Pastötter et al., 2011). This decrease in the prevalence of proactive interference during retrieval can occur without decreasing overall correct recall for the words from the target list (Nunes & Weinstein, 2012; Szpunar et al., 2008). This is because, the presence of a test increases the individual's ability to discriminate between the sections of material, in the previous cases lists of words, to retrieve the information that they are being tested on (Chan, 2009; Szpunar et al., 2008).

Proactive interference can also build up during encoding of material, or the stage in which individuals process the material for storage (Kliegl et a., 2015). The process of proactive interference during the encoding stage has been proposed because encoding of the subsequent lists are impaired without the presence of retrieval between each section. Many theories have been proposed to explain the buildup of proactive interference during encoding, including that of decreased attention on subsequent material and/or increased memory load due to the absence of a

test (Kliegal et al., 2015; Pastötter et al., 2011). The presence of a test between sections of material, may therefore decrease proactive interference that occurs during the stage of encoding subsequent material through decreasing the cognitive load and increasing attention.

Reset of Encoding Theory

Another theory that has been proposed to explain the FTE is that of reset of encoding, or storage of material (Chan et al., 2018; Pastötter et al., 2017). Higher encoding of information has been linked with increased attention to the material and lower levels of alpha power (Pastötter et al., 2008). Alpha power, or the frequency measured by electroencephalogram (EEG) has been linked to memory load and attention (Pastötter et al., 2011). Previous studies examining alpha power have shown that higher levels are associated with greater inattention and cognitive load. A study conducted by Pastötter et al. (2011) examining switching participants between testing and retrieving during studying of material found that those who were tested between sections exhibited lower alpha power and showed enhanced encoding/learning for the newer material. This suggests that the test resets the encoding of material for the new list.

Although this is similar to the context change hypothesis, resetting of encoding does not improve section discrimination through the creation of context cues, but reduces proactive interference through increasing attention to the material and separating the sections of material from each other (Pastötter & Bäuml, 2014). This means that the material is not learned as a large amount of information but as separate 'sessions.' Another study by Pastötter et al. (2018) examined the testing effect on serial position of list items. Participants were either tested following each of the three lists presented or only the third list. The results indicated that the test condition enhanced the primacy effect, or the idea that memory is better for information

presented earlier compared to the middle and more recent information. This supports the reset of encoding theory as encoding was greater when less information was studied between tests.

The Negative Side of Interleaving Tests

Although Szpunar et al. (2008) identified the benefit of testing previously learned material on learning subsequent information, it is important to note that the FTE can cause the opposite effect on learning (Davis & Chan, 2015; Davis et al., 2017; Finn & Roediger, 2013; Karaca et al., 2020). That is, recent studies using basic associations, i.e., face-name (Finn & Roediger, 2013) and flag-country (Karaca et al., 2020), have identified that switching between retrieval and encoding can cause poorer learning of new material when compared to switching between restudy and the presentation of new material. These studies identified that as the frequency of switching between the two tasks, or the interleaving of a test, increased, the detrimental effect on future learning also increased.

Currently, the specific cause for this opposite effect of testing on future learning is unknown, although multiple theories have been proposed (Davis & Chan, 2015). One proposed reason as to why individuals who are tested between sections of presented material show poorer learning of new material compared to those in the restudy condition is that they actively devote more time and attention to relearning the material than after being tested (Davis & Chan, 2015; Metcalfe, 2002). Davis and Chan (2015) propose that this may be due to the test highlighting what the individuals did not recall from the previous information, meaning that they consciously participate in selective attention and focus on reencoding the previously presented information. Although this could be beneficial if the individuals were to be presented with a cumulative test as they would have enhanced memory for the previous information, the detriment acts on the subsequent information.

Another proposed process as to why testing between sections may negatively impact subsequent learning is that switching between the opposing tasks of retrieval and encoding may require more cognitive resources compared to switching between restudy and new encoding as both require encoding material (Karaca et al., 2020). This potential process is considered because the retrieval and encoding are opposing processes and are considered to be incompatible, compared to reencoding and new encoding. Even though this disadvantage of the FTE has been identified with more simple associations including the face-name-profession pairs, it is important to note that this effect can be balanced based on how the tests are placed into the learning process (Davis et al., 2017). That is, in educational settings, instead of consistently switching between retrieval and encoding as that may decrease learning for the newer information, it may be possible to negate the negative impact through testing after larger sections, rather than asking students to retrieve information immediately following its presentation.

The Effect of Relatedness

In educational settings, testing usually occurs after students have studied large units, consisting of vast amounts of related information (Jimenez & Modaffari, 2021). However, when individuals are tested on larger sections of text passages, they recall significantly less information than those tested on smaller sections (Wissman & Rawson, 2015). This means that in order to enhance students' learning, it would be most beneficial to test using smaller sections of materials, as were used in this study. In addition, most course units build upon the previous unit and are highly related. That is, in United States history classes, for example, students may learn about what occurred between 1900 and 1930, take a test, then learn about 1930 to 1960. The FTE suggests that students who are tested between the course units would perform better on a test on the second unit, have fewer intrusions of events from the first time period, and show

overall enhanced learning for the second unit than those untested (Szpunar et al., 2008; Szpunar et al., 2011).

In addition, the increase in consecutive units that are learned is often associated with a decrease in students' ability to discriminate between the concepts (Rohrer, 2012). When units are blocked, or grouped like this, compared to interleaved, in which one unit is followed by a unit on an unrelated topic, students usually have more difficulty discriminating between the material. The context change theory would suggest that changing between topics would improve discrimination between the material because, like the presence of a test, the change in material would create internal context cues that would assist with source monitoring (Weinstein et al., 2014). The proactive interference theory would suggest that changing topics reduces proactive interference possibly due to a change in attention that increases the ability to discriminate between the material (Kliegal et al., 2015; Pastötter et al., 2011). The reset of encoding theory would suggest that the change in material separates each section and provides the opportunity for encoding to be reset (Pastötter et al., 2017). Recent research has led to the most support for the reset of encoding theory, especially as the reduction of proactive interference has been proposed as an effect rather than a cause of the FTE and relatedness of material (Ahn & Chan, 2022; Pastötter et al., 2018; Pastötter & Frings, 2019).

Educational Implications

Previous literature has expanded the FTE to more complex text materials, still demonstrating that tested individuals recall more information than those in the restudy condition (Wissman, et al., 2011). When using unrelated text passages, previous studies have focused on sections that are indirectly related, but within the same subject area. For example, Wissman and Rawson (2015) examined the FTE in one experiment using four sections of text each discussing

government interventions. Although the sections discussed different forms of interventions, they were all based on the same subject area of political history. This study found the presence of a FTE when participants were tested between each section.

This is contrary to educational settings, as students are expected to successfully switch between classes with very different subject material. In one experiment, Nunes and Weinstein (2012) presented participants with five lists of Deese/Roediger/McDermott (DRM) words, with the first four being blocked as words from individual lists and the last list being comprised of a mixture of words from four different lists. Unlike previous research that has indicated that testing decreases the presence of proactive interference and enhances subsequent learning (Szpunar et al., 2008), Nunes and Weinstein (2012) found that there were no significant differences in proactive interference on the fifth list when comparing individuals who were tested between each list and those who were not. This is possibly due to the change in meaning between the lists, and highlights the importance of examining the FTE on related and unrelated material as is seen in educational settings.

It is valuable for teachers and school systems to understand the benefits of the FTE in order for them to develop more efficient and directed tests that not only measure what has been learned, but also promote future learning. Although there are many benefits of frequent and directed testing, there are many variations that testing can take. Students can be tested on larger units or smaller sections, and the formats can include multiple-choice, short-answer, fill in the blank, and others. Developing a better understanding of how testing influences learning, as well as the most efficient test format and information presentation, is imperative for increasing the learning that occurs in educational settings.

Present Study

This study examined the FTE using related and unrelated sections of text sections to match the naturalistic educational setting. Participants were divided into both 'Test' or 'No Test' and 'Related' or 'Unrelated' groups, creating four conditions (related test, related no-test, unrelated test, unrelated no-test). Individuals in the related condition were presented with four text sections within the same subject area, while those in the unrelated condition were asked to study four randomized sections not within the same subject area. Those in the test condition received a test following each section in addition to the criterial test after the fourth section while those in the no-test condition only received the criterial test. As testing on relatively unrelated materials has been shown to improve later learning, I hypothesized that individuals who were tested on the unrelated materials would reflect overall greater learning for each subject compared to those untested for both related and unrelated materials and those tested on related materials.

As discussed before the context change hypothesis suggests that learning of subsequent information is enhanced and proactive interference is decreased through the retrieval cycle when individuals are tested (Jang & Huber, 2008; Szpunar et al., 2008). Previous research has shown that individuals tested between sections of material show higher learning on a criterial test than those untested (Szpunar et al., 2008). This suggests that individuals in the present study who were assigned to the test condition should have exhibited enhanced performance on the criterial test compared to those in the no-test condition. In addition, the context change hypothesis suggests that those in the unrelated condition should perform higher on the criterial test than those in the related condition, regardless of their test condition assignment (Wickens, 1970). This is because, the change from a text in one subject area to an unrelated subject area should create a similar shift in internal context to that of introducing a test between the section, thus increasing

the number of context cues associated with the material and decreasing future prior-list intrusions.

The reset of encoding hypothesis suggests that the presence of a test between sections of material enhances learning of subsequent information and decreases proactive interference as it causes encoding to be reset (Chan et al., 2018; Pastötter et al., 2017). The knowledge that testing between sections of material will enhance memory for subsequent information suggests that in this current study, individuals who were tested between each section would perform better on the criterial test than those untested. In addition, as this hypothesis proposes that the test causes encoding to be reset through the context change, this suggests that for individuals assigned to the unrelated condition, without the presence of a test, encoding would not be reset, so there should be no difference between those in the unrelated condition and those untested in the related condition.

Method

Participants

Participants in this study were 119 individuals recruited through Prolific, an online research platform (www.prolific.co). Data were collected from 120 participants, although one was excluded due to a computer error. Thirty two participants were in the related test condition, thirty six in the related no-test condition, twenty three in the unrelated test condition, and twenty eight in the unrelated no-test condition. The sample size of 26 participants per condition was determined using a .5 effect size, two-tailed, .05 error probability, .8 difference between means power analysis using G*Power (Faul et al., 2009). The sample size per condition was unequal after data collection due to the use of a randomizer in the Qualtrics survey. This is because participants had the opportunity to return the survey if they no longer wished to participate,

without penalty. However, once a participant began the survey, their short participation altered the randomizer count in the Qualtrics survey. That is, although the randomizer counted these as participants, their responses were not collected as they did not complete the entire survey.

Data were collected through two waves. In the first wave, data was collected from 40 participants who received $1.50 for their participation. As the average completion time was just under 15 minutes, above the original 10 minute estimate, compensation was increased to $2.00 in the second wave to meet the Prolific minimum payment of $8.00/hr. Data was collected from 80 participants in this second wave at $2.00 compensation.

Design

The independent variables in this study were the relatedness of the sections and the presence of a test. A 2 (Test x No Test) x 2 (Related x Unrelated) factorial between-subjects design was used. The two independent variables were the presence of a test and the relatedness of the material. The dependent variable in this experiment was performance on the criterial test, or the test following the fourth section, measured by the proportion of correct answers on the test.

Materials

Four passages were identified to develop the text sections used in this study. Each passage required a tenth grade reading level and had to be long enough to be broken into four individual stand-alone sections. After identifying the four passages, four 100-250 word sections were obtained from each, for a total of 16 text sections (see Appendix A). Each section was taken from the source passage with the ability to stand alone or be placed in its origin order to create a continuous passage with the other three. Each participant saw a total of four passages, as discussed in the procedure section below.

Participants were randomly assigned into either a related or unrelated condition regarding the content of the text sections they were presented with. Participants in the related condition saw all four sections from one passage in the original order so that it read as a single unit (i.e., Crocodile Tears 1, Crocodile Tears 2, Crocodile Tears 3, Crocodile Tears 4). Participants in the unrelated condition read one section from each of the four passages. The order in which participants saw a section from each passage was counterbalanced. Each section maintained their position from their original passage (i.e., Section 1 is always presented first, Section 2 second, etc.). For example, one participant would see Crocodile Tears 1, All that Glitters 2, Medieval Hospital 3, Light Therapy 4, another All that Glitters 1, Light Therapy 2, Crocodile Tears 3, Medieval Hospital 4.

For each text section, ten cued-recall questions (five fill-in-the-blank and five short answer), each requiring an answer of one to five words were developed to identify the main ideas of the section. Each question was developed to address a single idea from the section it corresponded to, addressing the Remembering (memorizing and recalling) and Understanding (identifying and explaining ideas) levels of Bloom's Taxonomy (Bloom, 1956). For example, a participant who was tested on Crocodile Tears 1 saw the question "What type of mixture did Johnson apply to the eyes of four species?" with the correct answer being onion and salt.

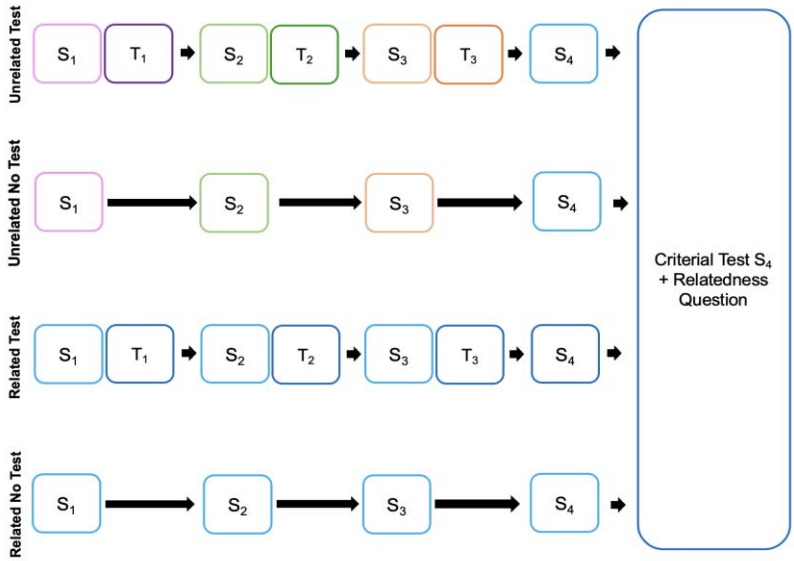
Procedure

The study received approval from the University of North Florida Institutional Review Board (ID #1914785-1). Participants were provided informed consent before being provided with instructions regarding the experiment. Participants were informed that they would read short sections of text and would either be tested through short answer and fill-in-the-blank questions after each section or continue to the next section, with the presence of a test following

the passage being randomized (Figure 1). Although they were informed that they would be randomly assigned a test following the text section, participants were assigned into either the test condition in which they received a test following each section or the no-test condition in which they were only shown a test following the fourth section. All participants, regardless of test condition received a test following their final, or fourth, section, also known as the criterial test.

Figure 1

Outline of Procedure for Each Condition



Participants were presented with one text section at a time and were instructed to read the material at their own pace, but were unable to move on from that section sooner than 45 seconds[1]. In the test condition, participants were then immediately directed to answer 10 questions regarding that section. Participants then studied section two with the same instructions as section one, followed by a 10-question test regarding section two. This process continued for

[1] In a small pilot study, participants read the passages at an average reading time of one minute. This reading time was used to encourage participants to fully read the passage, but dropped to 45 seconds to accommodate faster readers.

participants in the test condition until all four sections had been studied and tested. In the no-test condition, participants were immediately directed to the second section after reading the first section, then the third and fourth, with no test between the sections. After the fourth section, all participants, regardless of test condition, received a test of 10 questions based on section four, known as the criterial test. Responses to each question were coded using the following scoring; correct, no response/indicated did not know, intralist intrusion, extralist intrusion, incorrect/other. Correct responses were considered those that covered the main idea of the question. Questions that required an answer of two separate ideas, i.e., onion and salt or a specific year required the exact answer, while others required partial answers. For example, an answer of nuns when the full correct answer was third-order nuns was accepted. Following the criterial test, participants in each condition were asked to complete a relatedness judgement using a 5-point Likert Scale. This was completed through the use of the question 'On a scale of 1 (not at all related) to 5 (highly related), how related were the sections?' This question determined if the participants were aware of the relatedness of the sections as they were reading them, and if the manipulation was successful.

Results

Manipulation Check

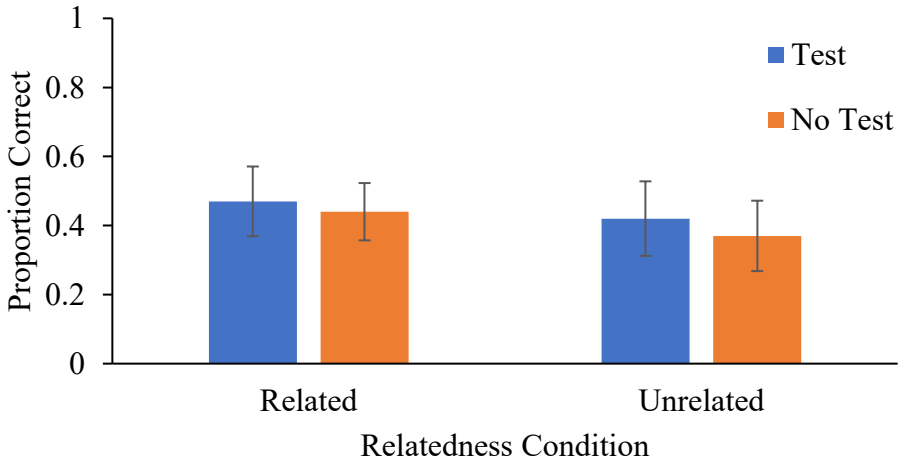
At the end of the survey, participants were asked to rate the relatedness of the material on a scale of 1 (not at all related) to 5 (highly related) as a manipulation check. This manipulation was successful in that participants found the related text passages to be significantly more related ($M = 4.43$, $SD = 0.89$) and unrelated sections as relatively not related ($M = 1.94$, $SD = 1.30$), $t(117) = 12.38$, $p < .001$. This indicates that the manipulation was successful, and participants were aware as to whether or not their sections were related to each other.

Forward Testing Effect and Relatedness

A 2 (Test vs No Test) x 2 (Related vs Unrelated) between-subjects factorial ANOVA was performed to examine the proposed main effects of test condition and relatedness condition on criterial test performance as well as the interaction between test and relatedness conditions (Figure 2). Contrary to the proposed hypothesis, no significant main effect of test condition was found, $F(1, 115) = 0.75$, $p = .39$. This indicates that overall, there was no forward testing effect in this study. In addition, contrary to the context change hypothesis and reduction of proactive interference mechanisms previously discussed, no significant main effect of relatedness condition was found, $F(1,115) = 1.35$, $p = .25$. This indicates that the relatedness of the material did not impact performance on the criterial test, although this supports the reset of encoding hypothesis. Finally, no significant interaction between relatedness and test condition on criterial test performance was found, $F(1, 115) = 0.06$, $p = .80$.

Figure 2

Criterial Test Performance



Although the between-subjects factorial ANOVA did not reveal any significant effects, either a main effect of test condition or relatedness, or an interaction between the conditions, pre-registration of this study included examination of the presence of a test based on relatedness condition[2]. This was due to the possibility that the study would not have enough power to detect a statistically significant difference between the means, as analyses were planned using .8 power. If this were true, post-hoc tests would be able to identify any existing effect. Contrary to the proposed hypothesis, in the related conditions, performance on the criterial test in the test condition (M = .47, SD = .28) was not significantly different compared to those in the no-test condition (M = .44, SD = .24), $t(66) = 0.46$, p = .64. In the unrelated condition, performance on the criterial test also did not differ between the test condition (M = .42, SD = .25) and no-test condition (M = .37, SD = .26), $t(49) = 0.75$, p = .46, indicating the lack of a forward testing effect, regardless of relatedness condition.

Discussion

The current study aimed to examine the forward testing effect (FTE) as it relates to text sections and the potential for an interaction with the relatedness of the material. Three effects were examined, with the proposal of a test and relatedness main effect as well as a test and relatedness interaction. Regarding the interaction between test and relatedness condition, two hypotheses were proposed. According to the context change hypothesis, the presence of a test and unrelated material should create new context cues for each section, which should lead to a smaller FTE for the unrelated text sections (Jang & Huber, 2008; Szpunar et al., 2008; Pastötter et al., 2011). The proposed mechanism of the reduction of proactive interference predicts the same effect, of a smaller FTE for the unrelated material. This is because, this mechanism

[2] Preregistration can be found here; https://osf.io/8pc2b/?view_only=de357b99c01b4f9da29f87bafc667ccd

suggests that both the presence of a test between sections and the unrelated material should increase attention and overall recall of these materials, thus increasing discrimination of material, or reducing prior-list intrusions (Nunes & Weinstein, 2012; Pastötter et al., 2011; Szpunar et al., 2008).

Contrary to these mechanisms, the reset of encoding theory would predict no difference in the FTE between the related and unrelated materials. This is because, the reset of encoding theory states alpha power increases with the study of material and is associated with greater inattention as well as cognitive load (Pastötter et al., 2011; Pastötter & Bäuml, 2014). However, the presence of a test between materials should reset this alpha power, thus resetting encoding, and enhance the learning of the new sections. Based on this proposed mechanism, the relatedness of the material should not have an impact on the reset of encoding. This means that the test should reset encoding for both the unrelated and related conditions equally.

Contrary to the proposed hypotheses and discussed mechanisms, no FTE, or main effect of test condition, was found. This means that there was no significant difference in performance on the criterial test between those in the test and no-test conditions. This null effect is important to note and surprising because the materials used in this study were designed based on previous studies that have shown a well-established FTE. This finding suggests that the FTE may not be universal across all types of materials as has been previously thought. Although no FTE was found, it is important to note that the results obtained indicated the hypothesized direction of the FTE. That is, although the difference was small, participants in the test condition performed marginally, but not significantly, better (M = .47, SD = .28) than participants in the no-test condition (M = .44, SD = .24) when comparing criterial test performance. This means that the

test enhanced, although not significantly, the correct recall on the criterial test, as indicated by the FTE.

As stated before, the null effect in this study of no significant FTE was surprising. Although the effect occurred in the expected direction, in that participants in the test condition gave a slightly higher number of correct answers on the criterial test than participants in the no-test condition, this difference was not significant. However, research using a variety of material types, including word lists, face-name pairs, and text passages have shown a significant FTE (Szpunar et al., 2008; Weinstein et al., 2011; Wissman & Rawson, 2015; Wissman et al., 2011).

The FTE must be identified as a significant effect in order to examine how additional factors may impact the phenomenon. As there was no FTE in this study, the impact of the relatedness of the material could not be examined, and thus cannot be considered as influential or excluded as a possible confounding variable. Therefore, it is important to discuss what would have been expected had a significant FTE been observed. According to both the context change hypothesis and the reduction of proactive interference proposed mechanisms, it was expected that participants in the unrelated condition would perform higher on the criterial test than those in the related condition (Jang & Huber, 2008; Nunes & Weinstein, 2012; Pastötter et al., 2011; Szpunar et al., 2008). According to the reset of encoding theory, the relatedness of the material should not have an impact on the FTE (Chan et al., 2018; Pastötter et al., 2017; Pastötter & Bauml, 2014). Although these proposed effects are important to identify, the lack of a significant FTE in this study does not allow for the testing of this interaction.

Why did this study not identify the FTE?

The FTE has been an established and robust phenomenon across a variety of materials, including word lists, face-name pairs, and short text passages (Szpunar et al., 2008; Weinstein et

al., 2011; Wissman & Rawson, 2015). That is, throughout various studies, participants who are tested between presentation of material perform better on the criterial test compared to those assigned to the restudy condition. As this study followed the general procedure of most studies examining the FTE, it is important to discuss possible explanations as to the lack of the FTE. One important distinction between most of the previous research on the FTE and this study was the use of text passages compared to word lists or face name pairs, as few studies have been conducted using the type of text materials from the current study.

One study using similar word length and reading level passages which found the FTE was that of Wissman et al. (2011). Throughout multiple experiments, this study used a variety of types of text passages with different lengths and topic focus. The experiments used a 779-word passage regarding government intervention that was broken into three sections with related headers, a 1,062-word passage on greenhouse gases divided into three sections, or a 1,318-word passage focusing on Hollywood's description of history that was broken into four sections. In addition, each section was read at the 11th to 12th grade reading level. As the sections in the current study were developed at the 10th grade reading level and between 100-250 words, they are similar to those developed by Wissman et al. (2011).

A similar study in which the FTE was identified using text passages was that of Davis and Chan (2023). This study used approximately 750-word sections regarding lasers at the 11th grade reading level. The similarity of material reading level between these previous studies and the current one indicates that the lack of a FTE in this study should not be due to the reading difficulty. In addition, as both the Wissman et al. (2011) and Davis and Chan (2023) studies developed their text materials to be integrated, or read together, this study should have found a

FTE in the related condition as the four sections were developed from one source passage and similar word length to the Wissman and colleagues (2011) materials.

In addition to the type of text sections used in these studies, it is important to examine the form of testing used. The current study used cued-recall questions in which participants were given retrieval cues to identify the main ideas in the sections. These were formed as both fill-in-the-blank and open-ended questions. In the Wissman et al. (2011) study, participants were given a free-recall test in which they were asked to recall as much information from the section as possible. This is different from a cued-recall test in that they were not given any retrieval cues for the material. Wissman et al. (2011) found the FTE using this type of recall test. In the Davis and Chan (2023) study, participants were given cued-recall or multiple choice tests. This study found the FTE for both forms of tests, with the effect being stronger for the cued-recall. This suggests that the FTE is created by both cued and free recall questions. This indicates that free recall is not required to identify a significant FTE for text passages, and it should have been observed in the current study.

Overall, these previous studies identified the FTE using similar materials, both the text passages and tests, as the current study (Davis & Chan, 2023; Wissman et al., 2011). So why was no FTE found in the current study? The main difference between the Davis and Chan (2023) study and this study was the length of the material. As the Davis and Chan (2023) study used longer text sections, it is possible that participants benefited from the short length of the material from this study, and the presence of a test was not influential. However, as the length of these sections were similar to that of the Wissman et al. (2011) study, this is unlikely. This suggests that the lack of the FTE may be due to the combination of the section length and test type, although it is difficult to say which combination would be more effective in producing the FTE

as one study did not report effect sizes. It is possible that this study would have found the FTE using free-recall with the current section length, or if the sections were longer with a cued-recall.

Another difference between the current and previous studies is the lack of a filler task between conditions for those in the no-test condition. Rather than have a specific task designated to break up the sections, participants in the current study moved immediately into each new section. In addition, participants in the test condition did not experience a delay between the test and the next text section. Many previous studies examining the FTE have used a filler task, including math problems or video games, for one to five minutes (Davis & Chan, 2015; Szpunar et al., 2008; Szpunar et al., 2013; Weinstein et al., 2011; Weinstein et al., 2014). These factors may be related to the lack of the FTE in the current study, meaning that it is important for future research to consider the use of a filler task to match these previous studies. However, it is possible that the presence of a filler may lead to a smaller mean difference in test performance, and erase any FTE as this would provide a break from study as a test would.

An important element of the procedure used to collect data in this study is that of the use of the online platform Prolific. Currently, there are a variety of platforms that are used for online research, including Amazon Mechanical Turk, Prolific, and others (Tang et al., 2022). This study used Prolific, which has been shown to provide better participant generalizability through the use of a representative sample, higher data quality (more consistent and accurate responses), and better comprehension of instructions compared to Amazon Mechanical Turk (Douglas et al., 2023; Tang et al., 2022). In addition, there are many strengths associated with collecting data online. These include the ability to collect data quickly, provide quality data with lower costs and time, and most participants perform higher on attention checks than in-person participants (Cobanoglu et al., 2021; Newman et al., 2021).

Although the use of online data collection has become increasingly popular, it is important to note that there are many limitations that may influence the generalizability and validity of the research (Newman et al., 2021). A major factor is that there is a lack of external control on participants (Cobanoglu et al., 2021). That is, the researcher is not able to monitor the participants in real time as they complete the study. This means that participants may not attend to each question completely, provide incorrect responses intentionally, and/or cheat (Cobanoglu et al., 2021; Newman et al., 2021). For the current study, this may have presented itself through participants screenshotting the material for reference on potential tests or choosing to not read or pay full attention to the text section. This may have led to the null forward testing effect as cheating and not reading the material may have created response variability. That is, if some participants performed very well on the test while others performed poorly, the scores on the criterial test would be widely distributed throughout the range of 0-10, creating a lack of the FTE.

Another important consideration when completing online research is that of bots, or software systems that have been created to complete online tasks (Lebeuf et al., 2017). Bots are able to complete many online surveys successfully if they consist of multiple choice or logic answers. For example, a bot may be able to complete an illogical attention check (Cobanglu et al., 2021). Although bots are able to select answers that would require clicking buttons, they are currently not able to answer questions with phrases or words that make sense through a cue. Due to the nature of this study, it is unlikely that any included data was sourced from bots, although it is important to consider the use of a difficult attention check or a Captcha verification process for future research as one was not used in the present study (Cobanoglu et al., 2021; Tang et al., 2022).

Although the reason as to why no FTE appeared in the current study, given the reliability of collecting data online and the similarities to previous studies, is unclear, this null finding is still important. Although the research question regarding the impact of the relatedness of the material was unable to be examined, the lack of the FTE indicates that it may not be a universal effect. That is, it may not be applicable to all studied materials, or it may be limited by the combination of material and test type. This means that it is important to consider conducting future research on the materials in which the FTE has been a robust effect in order to examine the effect of the relatedness of the material. For example, the FTE has been well-established using word lists, meaning that the current research question may be applicable to these materials. Overall, this indicates that the forward testing effect may not be a universal effect for all types of learning materials and should be examined further to develop recommendations for educators and researchers.

References

Ahn, D., & Chan, J. C. K. (2022). Does testing enhance new learning because it insulates against proactive interference? *Memory & Cognition, 50*, 1664-1682. https://doi.org/10.3758/s13421-022-01273-7

Ahn, D., & Chan, J. (2023). Does testing potentiate new learning because it enables learners to use better strategies? *Journal of Experimental Learning Memory and Cognition*. https://doi.org/10.1037/xlm0001233

Bloom, B. S. (Eds.). (1956). *Taxonomy of educational objectives, handbook1: Cognitive domain (2nd ed)*. New York David McKay Co Inc.

Buklijaš, T. (2008). Medicine and society in the medieval hospital. *Croatian Medical Journal, 49*(2), 151-154. https://doi.org/10.3325/cmj.2008.2.151

Chan, J. C. K. (2009). Long-term effects of testing on the recall of nontested materials. *Memory, 18*(1), 49-57. https://doi.org/10.1080/09658210903405737

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin, 144*(11), 1111-1146. https:doi.org/10.1037/bul0000166

Cobanoglu, C., Cavusoglu, M., & Turktarhan, G. (2021). A beginner's guide and best practices for using crowdsourcing platforms for survey research: The case of Amazon Mechanical Turk (MTurk). *Journal of Global Business Insights, 6*(1), 92-97. https://www.doi.org/10.5038/2640-6489.6.1.1177

Davis, S. D., & Chan, J. K. (2015) Studying on borrowed time: How does testing impair new learning? *Journal of Experimental Psychology, 41*(6), 1741-1754. https://doi.org/10.1037/xlm0000126

Davis, S.D., & Chan, J.C.K. (Submitted February 2023). Effortful tests and repeated metacognitive reflection enhance future learning. Educational Psychology Review.

Davis, S. D., Chan, J. C. K., & Wilford, M. M. (2017). The dark side of interpolated testing: Frequent switching between retrieval and encoding impairs new learning. *Journal of Applied Research in Memory and Cognition, 6*(4), 434-441. https://doi.org/10.1016.j.jarmac.2017.07.002

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE,* https://doi.org/10.1371/journal.pone.0279720

Faul, F., Erdfelder, E., Buchner, A., * Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Finn, B., & Roediger, H. L. III. (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1665-1681. https://doi.org/10.1037/a0032377

Gilens, J. F., Hoss, M., Lyon, C., & DeSanto, K. (2021). Does light therapy decrease in older adults? *American Family Physician, 104*(4), 417-418.

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 112-127. https://doi.org/10.1037/0278-7393.34.1.112

Jimenez, L., & Modaffari, J. (2021). *Future of testing in education: Effective and equitable assessment systems.* Center for American Progress.

https://www.americanprogress.org/article/future-testing-education-effective-equitable-assessment-systems/

Karaca, M., Karpas, N., Wilford, M. M., & Davis, S. D. (2020). Too much of a good thing: frequent retrieval can impair immediate new learning. *28*(10), 1181-1190. https://doi.org/10.1080/09658211.2020.1826526

Kliegal, O., Pastötter, B., & Bäuml, K.-H. (2015). The contribution of encoding and retrieval processes to proactive interference. *Journal of Experimental Psychology*, *41*(6), 1778-1789. https://doi.org/10.1037/xlm0000096

Lebeuf, C., Storey, M., & Zagalsky, A. (2017). Software Bots, *IEEE Software, 35*(1), 18-23. https://doi.org/10.1109/MS.2017.4541027

Neff, E. P. (2017). All that glitters is gold-frozen fish. *Lab Animal, 46*(10). 355, https://doi.org/10.1038/laban.1359

Newman, A., Bavik, Y. L., Mount, M., & Shao, B. (2021). Data collection via online platforms: Challenges and recommendations for future research. *Applied Psychology: An International Review, 70*(3), 1380-1402. https://doi.org/10.1111/apps.12302

Nunes, L., & Weinstein, Y. (2012). Testing improves true recall and protects against the build-up of proactive interference without increasing false recall. *Memory, 20*(2). 138-154. https://doi.org/ 10.1080/09658211.2011.648198

Pastötter, B., & Bäuml, K.-H. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology, 5*(286). https://doi.org/10.3389/fpsyg.2014.00286

Pastötter, B., Bäuml, K.-H., & Hanslmayr, S. (2008). Oscillatory brain activity before and after an internal context change – Evidence for a reset of encoding process. *Neuroimage, 43*(1), 173-181. https://doi.org/10.1016/j.neuroimage.2008.07.005

Pastötter, B., Engel, M., & Frings, C. (2018). The forward effect of testing: Behavioral evidence for the reset-of-encoding hypothesis using serial position analysis. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01197

Pastötter, B., & Frings, C. (2019). The forward testing effect is reliable and independent of learners' working memory capacity. *Journal of Cognition, 2*(1), 37, 1-15. http://doi.org/10.5334/joc.82

Pastötter, B., Schicker, S., Niedernhuber, J., Bäuml, K.H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, (37)*2, 287-297, https://doi.org/10.1037/a0021801

Pastötter, B., Tempel, T., & Bäuml, K-H. (2017). Long-term memory updating: The reset-of-encoding hypothesis in list-method directed forgetting. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.02076

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Shaner, D. M., & Vilet, K. A. (2007). Crocodile tears: And thei eten hem wepynge. *BioScience*, *57*(7), 615-617. https://doi.org/10.1641/b570711

Strauss, V. (2015, October 24). *Confirmed: Standardized testing has taken over our schools. But who's to blame?* Washington Post. https://www.washingtonpost.com/news/answer-

sheet/wp/2015/10/24/confirmed-standardized-testing-has-taken-over-our-schools-but-whos-to-blame/

Szpunar, K. K., Khan, N., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America, 110*(16), 6313-6317. https://doi.org/10.1073/pnas.1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology; Learning, Memory, and Cognition, (34)*6, 1392-1399. https://doi.org/10.1037/a0013082

Tang, J., Birrell, E., & Lerner, A. (2022). How well do my results generalize now? The external validity of online privacy and security systems. https://doi.org/10.48550/arXiv.2202.14036

Weinstein, Y., Gilmore, A., Szpunar, K., & McDermott, K. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 40*(4), 1039-1048. https://doi.org/10.1037/a0036164

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review, (18)*518. https://doi.org/ 0.3758/s13423-011-0085-x

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77*(1), 1–15. https://doi.org/10.1037/h0028569

Wissman, K. T., & Rawson, K. A. (2015). Grain size recall practice for lengthy text material: Fragile and mysterious effects on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 439-455. https://doi.org/10.1037/xlm0000047

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18*(6), 1140-1147, https://doi.org/10.3758/s13423-011-0140-7

Appendix A

Sections and Test Questions

Shaner & Vilet (2007)

Crocodile Tears Section 1

In mythology, tears of the crocodile are attributed to feigned sorrow. While the exact origins of this fancy are unknown, references date to 340 CE. This crocodilian remorse is considered insincere because the crocodile continues to gobble its prey. The Voyage and Travel of Sir John Mandeville, a book distributed widely around 1400, popularized for the general public the picture of the crocodile tearfully consuming humans: "In that country be a great plenty of crocodiles.... These serpents slay men, and they eat them weeping.” That crocodiles shed tears while eating was accepted widely until the early 1700s, when John Scheuchzer, a physician and naturalist in Zurich, wrote: "The foundations and substance of this famous old tale are so feeble that today we would be well advised to do without it." George Johnson examined these foundations most recently. After applying a mixture of onion and salt directly to the persistently dry eyes of four species, he concluded that "the popular notion of Crocodiles shedding tears is entirely a myth." At approximately the time of Johnson's investigation, F. A. Bogorad, in Russia, named a human neurologic syndrome "the symptom of crocodile tears." Upon eating, profuse, involuntary tears pour from the affected eye of some individuals afflicted with facial palsy. Bogorad invoked a theory of "parareflexes" to explain the medical phenomenon.

1. What type of mixture did Johnson apply to the eyes of four species?

   Onion and salt

2. Who found the same tear-shedding syndrome as crocodiles in humans?

   F. A. Bogorad

3. What have scientists previously attributed crocodiles shedding tears to?

   Feigned sorrow

4. Where was John Scheuczer from?

   Zurich

5. When did John Scheuchzer contradict the claim that crocodiles shed tears while eating?

   The 1700s

6. Some individuals afflicted with _____ shed involuntary tears while eating.

   Facial palsy

7. The theory of _____ was used to explain humans shedding tears while eating.

   Parareflexes

8. The book _____ of Sir John Mandeville was distributed around 1400.

   The Voyage and Travels

9. The idea that crocodiles shed tears out of sorrow dates back to _____

   340 CE

10. _____ determined that crocodiles shedding tears was a myth.

    George Johnson

Crocodile Tears Section 2

He claimed, after Andre Thomas, that defects in a newer phylogenetic system might lead to the appearance of older phylogenetic reflexes. In other words, the human facial weakness allowed older, possibly crocodilian neurological pathways to emerge. This hypothesis presumes that crocodilians lacrimate while eating. Our investigation of this human ailment prompted us to pursue the biological basis of the crocodilian metaphor and its underlying scientific theory. On 22 March 2006, we digitally filmed seven crocodilians--two common caimans, two Yacare

caimans, and three American alligators--as they were fed. Other species within the order Crocodilia, such as crocodiles and gharials, were not examined. Five of the seven crocodilians developed moisture in their eyes, bubbles, or overflow bubbles within minutes before, during, or after eating. One Yacare caiman and one common caiman did not lacrimate. All animals used in the study had been trained to move to feeding stations, where they were out of the water and relatively dry. We fed them dried foods (chicks, quail, and a pelletized alligator food biscuit made by Mazuri Feeds). We perceived no environmental effects that might account for moisture in the eyes from any source other than ocular secretions.

1. How many crocodilians did not lacrimate?

   Two

2. Where were the crocodilians trained to move to?

   Feeding stations

3. What occurred in human faces that allowed the neurological pathways to emerge?

   Facial weakness

4. How many crocodilians were filmed?

   Seven

5. What year did they conduct this study?

   2006

6. Andre _____ claimed that defects in the phylogenetic system could lead to the appearance of older reflexes.

   Thomas

7. Two common caimans, two Yacare caimans, and three _____ alligators were filmed.

   American

8. The crocodilians were fed _____, including chicks, quail, and alligator food.

   Dried food

9. No _____ could have accounted for the moisture in the crocodilian's eyes.

   Environmental effects

10. The purpose of the study was to examine the _____ basis for the metaphor.

    Biological

Crocodile Tears Section 3

One of the common caimans developed a small patch of foam in the rostral canthus of its left eye 40 seconds after it began to feed; the foam dissipated over the next 20 minutes. A Yacare caiman developed small bubbles in the right eye within 15 seconds of receiving its first dry food. This animal was agitated, hissing, and snapping before it began to feed. With continued feeding, no additional lacrimation or bubbles were seen. All three American alligators produced tears. A 3.4-meter male developed small bubbles in the left eye 64 seconds after receiving the first quail. The eye was also moist around the lids. The right eye demonstrated bubbling 2 minutes and 18 seconds after the alligator began to feed. A 3.5-meter male American alligator witnessed the feeding of the abovementioned alligator and may have anticipated being fed. This alligator already had small bubbles in its right eye and obvious glistening in its left eye before it began feeding. It emitted a low growl as we approached with the food. Two minutes and 5 seconds after feeding on five biscuits, fluid appeared in its left eye. Three minutes and 25 seconds after the animal initiated feeding, extensive bubbling began in its right eye.

1. 40 seconds after feeding, a _____ developed a patch of foam.

   Common caiman

2. What type of crocodilian developed small bubbles in the right eye?

Yacare caiman

3. How many of the American alligators produced tears?

   All/3

4. How long was the male that developed the small bubbles in the left eye after the first quail?

   3.4 meters

5. In a male, the right eye demonstrated bubbling _____ the alligator began to feed

   After

6. Approximately how long after the final alligator ate did the extensive bubbling in the right eye begin?

   Three and a half minutes

7. What type of crocodilian developed moisture before beginning to eat?

   American alligator

8. The first caiman showed aggression _____ feeding.

   Before

9. The final alligator _____ as they approached with the food.

   Growled

10. It took about _____ for the first crocodilians tears to dissipate.

    20 minutes

Crocodile Tears Section 4

The left eye of a 4.1-meter male also began bubbling, though to a milder degree, immediately upon our approach, after an aggressive gape but before any audible hiss. This alligator was fed 13 times with biscuits and quail without the bubbles changing in appearance or new bubbles

forming. When present in either alligator or caiman, bubbling predominated in one eye, with the jaws closed and without any particular associated jaw movement. The nasolacrimal duct of crocodilians opens directly into the nasopharynx, near the posterior limit of the preconcha, where it is greatly expanded as a lacrimal sinus. The duct is conveyed in a lacrimal bone into the lower eyelid in Crocodylusporosus, splitting into two small but distinct ducts in the embryo. We speculate that the ebullition of tears results when air forced from the lungs through the throat and into the nasopharynx drives secretions up the lacrimal duct into the eye. Surfactants or proteins may cause the tears to bubble. One of the authors of this article has noted extensive bubbling of tears during aggressive exchanges between adult males and following agonistic social display in the absence of eating. Crocodiles also shed overflow tears unassociated with eating when they have been out of the water for some time. This latter phenomenon is separate from the behavior and physiology described here.

1. How many times was the first alligator fed?

   Thirteen

2. Where does the nasolacrimal duct of crocodilians open into?

   The nasopharynx

3. The nasolacrimal duct is expanded as a _____.

   Lacrimal sinus

4. In addition to eating, bubbling in adult males has been identified during _____ exchanges.

   Aggressive

5. The crocodilians _____ their jaws when bubbling occurred.

   Closed

6. Crocodilians tend to shed _____ after having been out of water for a period of time.

   Overflow tears

7. The duct is conveyed into the _____ into the lower eyelid.

   Lacrimal bone

8. What was forced from the lungs that caused secretion from the eye?

   Air

9. What are two possibilities as to what causes the tears?

   Proteins or surfactants

10. What type of gape did they perceive from the first male as they approached?

    Aggressive

Neff (2017)

All that Glitters Section 1

For Mary Hagedorn, it was a moment over twenty years in the making: stirring before her was a zebrafish embryo, alive and seemingly well after taking a dip to 196 C. Hagedorn is a conservation biologist at the Smithsonian Conservation Biology Institute and Hawaii Institute of Marine Biology with a long interest in cryopreservation. "I've cryopreserved easily over a million, maybe even more, zebrafish embryos in my life and every single one of them turned to mush," she recalls. "To see one even intact for fifteen minutes was a huge, huge thing. And then to see them move at 24 hours was mind-blowing." The trick was a laser and a little gold. Cryopreservation of living cells first left the realm of science fiction in the 1950s. The basic approach involves freezing and thawing a sample quickly and uniformly to prevent the formation of ice crystals. Adding cryoprotectants helps prevent ice damage and smooths the "vitrification"

process by which the sample essentially turns to glass. Over the decades, cryopreservationists became pretty adept at handling mammalian cells and embryos, including those of humans. But fish, and other nonmammalian organisms like reptiles, amphibians, and birds, remained out of reach.

1. What was the temperature that the zebrafish embryo encountered?

   196 C

2. Mary Hagedorn has _____ over a million zebrafish embryos

   Cryopreserved

3. How long after being frozen was it surprising to see the embryos alive?

   Fifteen minutes

4. When did cryopreservation become a real-world technique?

   The 1950s

5. What does freezing and thawing a sample quickly and uniformly help avoid the formation of?

   Ice crystals

6. What type of organisms have been successfully cryopreserved?

   Mammalian cells/humans

7. Adding cryoprotectants helps smooth the _____.

   Vitrification process

8. _____ organisms had yet to be cryopreserved.

   Nonmammalian

9. According to Mary Hagedorn, all cryopreserved embryos turned to _____.

   Mush

10. The trick to cryopreservation was _____.

    A laser and gold

All that Glitters Section 2

With biophysicist Fritz Kleinhans, she developed a biophysical model to better understand the cryopreservation challenges inherent to zebrafish, which was then a rapidly emerging model organism. The underlying issues were two-fold. Zebrafish embryos are physically large--they need to be self-sustaining, so they include a large yolk to supply nutrients--and their internal membranes are quite impermeable to water to prevent bursting from changes in osmotic pressure. Ice formation is problematic with larger volumes, and cryoprotectants do not diffuse into the embryo's different compartments as easily as in other species. To overcome those issues, Hagedorn and her collaborators took advantage of microinjection technology to improve cryoprotectant delivery. This advance enabled them to deep-freeze countless embryos. But they couldn’t bring them back. Technology at the time was just not capable of warming them fast enough to prevent ice damage, Hagedorn says, so in 2004 she stopped...until a cryopreservation conference nearly a decade later. John Bischof, a mechanical engineer at the University of Minnesota, was presenting his lab's work with an emerging approach called laser warming. Unlike convective methods, laser warming relies on an absorber that indirectly transfers heat to surrounding tissue. Kleinhans and Peter Mazur at IUPUI had successfully laser-warmed mouse oocytes surrounded by India ink, but zebrafish embryos were still too large (and internally complex) for an external absorber to be sufficient.

1. The problem with zebrafish embryos is that they are too _____ and their internal membranes are impermeable to water.

Large

2. Why do zebrafish have a large yolk?

   To supply nutrients

3. What type of technology was used by Hagedorn to improve cryoprotectant delivery?

   Microinjection

4. Hagedorn stopped cryopreservation in _____ because technology was not capable enough.

   2004

5. What type of approach was John Bischof using?

   Laser warming

6. Laser warming is different compared to convective methods because it _____ transfers heat.

   Indirectly

7. What type of animal did they successfully laser-warm at IUPUI?

   A mouse

8. Ice formation is problematic with larger _____.

   Volumes

9. Where was John Bischof a mechanical engineer?

   University of Minnesota

10. Zebra fish are impermeable to water to avoid bursting from changes in _____.

    Osmotic pressure

All that Glitters Section 3

Gold, Kholsa explains, is inert and has long been used in biomedical applications. It also absorbs light in the wavelength of the Bischof lab's laser: 1,064 nm. Kholsa devised a mixture of the cryoprotectant polyethylene glycol and gold nanoparticles, microinjected it into 223 embryos, and froze them with liquid nitrogen. Then he hit a frozen embryo with the laser, a step lasting just one one-thousandth of a second. Though 223 fish didn't bounce back to life and normal development, a handful did, a feat that had eluded researchers for decades. From this proof-of-concept, Kholsa is ready to refine. "I would say this first paper, it was a lot of scaling, a lot of calculations and modeling and lot of trials to figure out that this can be done," he says. "Right now, we are just showing that there's structure and some functionality after the re-warming." In the future, he hopes to determine what exactly is happening to the embryo in the incredibly short warming burst, and to improve the nanoparticles and their distribution throughout a sample.

1. Gold has long been used in _____.

   Biomedical applications
2. The mixture created by Kholsa was microinjected into _____.

   223 embryos
3. Kholsa says they need to improve the distribution of the _____ throughout the sample

   Nanoparticles
4. The embryos were hit with a _____ after being frozen

   Laser
5. The mixture consisted of _____ nanoparticles and cryoprotectant polyethylene glycol.

   Gold
6. What were the embryos frozen with?

Liquid nitrogen

7. What did Kholsa say they were able to show after the re-warming?

   Structure and functionality

8. What is one reason gold has been used in these types of experiments?

   Absorbs light

9. In addition to trials, what needs to be done to determine the effect?

   Calculations and modeling

10. What does the laser do to the embryos?

    Short warming burst

All that Glitters Section 4

In addition to zebrafish, the team thinks that laser warming could be extended to many other organisms that have been beyond the reach of cryopreservation in the past. That has important value to Hagedorn, who currently studies coral conservation. "It's transformational in what it will allow us to do to conserve species," she says, "Because really all you have to do is freeze them. The laser function can happen later." Theoretically, says Kholsa, the technique could be applied to any 1-2 mm embryo or tissue sample; the nanoparticles used and temperature targets can be adjusted as needed. Hagedorn looks forward to deploying the technique at her lab in Hawaii, where Kholsa will join her for a few weeks to think through how to increase throughput and mechanize the thawing process. "Engineers can do amazing things," she says, "and we're just going to keep plucking away at it."

1. In addition to zebrafish, Hagedorn also studies _____.

   Coral conservation

2. Kholsa and Hagedorn need to increase throughput and mechanize the _____.

   Thawing process

3. Cryopreservation involves freezing _____ lasering.

   Before

4. Cryopreservation requires biologists and _____.

   Engineers

5. Currently, the laser function for other species is _____, not practical.

   Theoretical

6. What size embryos did Kholsa say the technique could be applied to?

   1-2 mm

7. Why is cryopreservation an important technique for species?

   Conservation

8. What needs to happen to the temperature targets before applying the technique to another species?

   Adjustment

9. In addition to embryos, what type of samples could the technique be applied to?

   Tissue

10. In what state are Hagedorn and Kholsa going to be working?

    Hawaii

Gilens et al. (2021)

Light Therapy Section 1

Light therapy appears to be mildly effective in treating depression in older adults, but ideal wavelength, intensity, and length of treatment are unknown. (Strength of Recommendation [SOR]: B, based on a systematic review of randomized controlled trials [RCTs] with heterogeneity and extrapolated from systematic reviews of RCTs of adults of all ages.) In adults, bright white light exposure in the mornings for less than 60 minutes may be most effective. (SOR: B, based on a systematic review of RCTs.) A 2018 systematic review and meta-analysis of six RCTs (N = 359) examined the effectiveness of light therapy among patients older than 60 years with nonseasonal depression. The trials compared light therapy (of varying wavelength, intensity, and duration) to either dim red or white light, nothing, or usual therapy.

1. Ideal wavelength, intensity, and _____ of light therapy are unknown.

   Length

2. Patients older than _____ were examined.

   Sixty

3. The trials were comparing light therapy of varying _____, intensity, and duration.

   Wavelength

4. The systematic review was based on _____ trials.

   Randomized control

5. The systematic review was conducted in _____.

   2018

6. How long is it suggested that adults expose themselves to bright light in the mornings?

   Sixty minutes

7. What mental health condition were the older adults experiencing?

   Nonseasonal depression

8. What color lights were the trials comparing?

   Red and white

9. In addition to light therapy and no treatment, what was the third comparison?

   Usual therapy

10. What color light is considered to be the most effect for adults?

    White

Light Therapy Section 2

Trials were conducted for four weeks (three trials), three weeks (one trial), or no more than two weeks (two trials). The pooled results of all trials found that geriatric depression symptoms improved more with a small to moderate effect in the light therapy groups compared with the control groups. Subgroup analysis by length of intervention revealed no statistically significant differences between treatment and control groups at two weeks, three weeks, or four weeks. In all six trials, there were no significant adverse reactions in the treatment group. The treatment and control groups did not differ in the rates of adverse reactions reported. Because of limited available evidence, the authors were not able to make conclusions regarding the ideal wavelength, intensity, or duration of light therapy in the study population. The findings are limited by moderate heterogeneity.

1. The treatment and control groups did not differ in the rates of _____ reactions.

   Adverse

2. The findings were _____ by moderate heterogeneity.

   Moderate

3. The trials were conducted between two and _____ weeks.

Four

4. There was a small to _____ effect in light therapy.

   Moderate

5. There were a total of _____ trials.

   Six

6. The trials found that depression symptoms improved more with what therapy?

   Light

7. At which length of intervention did they find statistically significant differences between treatment and control?

   None

8. Why did the authors not make a conclusion of the three factors related to light therapy?

   Limited evidence

9. What type of depression was being examined?

   Geriatric

10. How many trials were conducted in the three week condition?

    One

Light Therapy Section 3

A 2020 systematic review and meta-analysis of 23 RCTs (N = 1,120) examined the effectiveness of light therapy in adults of all ages with non-seasonal depression. The studies compared light therapy (of varying wavelength, intensity, and duration) with placebo or control. In one trial, patients received light therapy as monotherapy or as adjunctive therapy to fluoxetine (Prozac). The pooled results of the 23 trials found that light therapy had a mild to moderate effect in improving depression symptoms compared with placebo or control. Subgroup analysis found that

the bright white light subgroup had significant improvement with a small to moderate effect and significant heterogeneity.

1. The studies compared light therapy to a _____.

   Placebo

2. The light therapy had a _____ to moderate effect.

   Mild

3. The bright light subgroup had a small to moderate effect with _____ heterogeneity.

   Significant

4. Patients either received light therapy as _____ or in conjunction with medication.

   Monotherapy

5. _____ analysis found the significant improvement in the light therapy group.

   Subgroup

6. In what year was the systematic review conducted?

   2020

7. What medication was used in addition to the light therapy?

   Fluoxetine

8. How many RCTs were analyzed?

   Twenty three

9. What color light allowed for significant improvement?

   White

10. What was the sample for this review?

    Adults with nonseasonal depression

Light Therapy Section 4

The effects of the other types of light were no different from those in the control groups. Subgroup analysis of the timing of therapy found that morning delivery had a moderately significant effect compared with the control group, whereas delivery at any other time was no different than in the control group. Further subgroup analysis found that compared with the control group, treatment for less than 60 minutes per day had a moderate effect and treatment for 60 minutes or more per day had a small effect. The overall findings of the study are limited by moderate heterogeneity.

1. Morning delivery had a _____ significant effect.

   Moderately
2. Treatment for less than _____ per day ad a moderate effect.

   Sixty minutes
3. The study is limited by _____ heterogeneity.

   Moderate
4. The control group and other light groups were _____ different.

   Not
5. _____ analysis was used to determine the effect of timing of therapy.

   Subgroup
6. What light duration had a small effect?

   Sixty minutes or more
7. What type of analysis was used to compare the length of treatment?

   Subgroup
8. The light duration effect was determined by comparing the treatment group which group?

   Control

9. Delivery at times other than the morning had what effect?

   None

10. Overall, light therapy had what type of effect?

    Moderately significant

Buklijaš (2008)

Medieval Hospitals Section 1

Hospitals today are places where medical treatment is provided, but also places where major life events, such as birth and death, occur. Yet, their history is relatively short; they were born, together with modern medicine, some two hundred years ago in the revolutionary Paris. Around 1790, large hospitals and pioneering research blossomed throughout Europe, replacing the Hippocratic model of disease with the localizationist paradigm. The rise of the modern hospital began in Paris when the social change brought about by the French Revolution provided the momentum for the transformation. For the first time in history, cure of the body and care for the soul were separated, and physicians, rather than the church and rich lay patrons, took charge of medical institutions. Medical treatment was no longer a privilege of the rich (at home) or charity for the poor (in hospital), but an indispensable human right.

1. Around _____, hospitals and research took hold in Europe.

   1790

2. The change in the modern hospital was brought upon by the _____.

   French Revolution

3. The change in medicine began the separation of the body and the _____.

   Soul

4. Medical treatment was originally a privilege for _____ or charity.

   The rich

5. _____ change was brought about that lead to the transformation momentum.

   Social

6. How long ago were hospitals ‘born’?

   Two hundred years

7. What type of model of disease did hospitals replace?

   Hippocratic

8. What type of right did medical treatment become?

   Indispensable human

9. Before physicians, who was in charge of medical institutions?

   The church and the rich

10. Where did the rise of the modern hospital begin?

    Paris

Medieval Hospitals Section 2

While institutions providing some form of medical treatment existed in ancient Greece and Rome, neither of these cultures organized community care for the sick, poor, and needy. A radical change occurred in the late Antiquity, with the rise of Christianity, which embraced charity as one of its basic doctrines. The first hospitals were founded when Christianity became the state religion of the Roman Empire. Hospital tradition in Byzantium continued into the Middle Ages, but the West experienced a centuries-long break. At the end of the early Middle Ages, the Benedictine monks revived the hospital institution. Hospitals flourished in the crusades, with the rise of orders specialized for that service, such as Hospitaller Knights. But, by

the thirteenth century, growing urban communities had taken over the leading cultural role from monasteries. While monastic hospitals and hospital orders, such as the energetic Sisters of Mercy, continued to develop, hospitals physically and administratively moved to the cities.

1. Medical treatment was most prominent in _____.

   Greece and Rome

2. When _____ became the state religion of the Roman Empire, the first hospitals were founded.

   Christianity

3. Hospital tradition continued into the _____ before the West experienced a break.

   Middle Ages

4. The Benedictine _____ revived the hospital institution.

   Monks

5. A change occurred in the late _____ to encourage the embrace of charity.

   Antiquity

6. What type of care was lacking in ancient medical treatment?

   Community

7. By which century did urban communities take over from monasteries?

   Thirteenth

8. Where did hospitals begin to move to, both physically and administratively?

   Cities

9. When were hospitals flourishing?

   The crusades

10. Before hospitals were founded, what were the places providing care called?

Institutions

Medieval Hospitals Section 3

Italian merchant urban communes, such as Florence, Padua, and Venice, spearheaded urbanization and partial secularization of hospitals, which were being increasingly established by local governments, confraternities, and rich individuals. Hospitals guarded the social order and enabled uninterrupted running of commerce and manufacture in cities. Considered as institutions of social prevention, they simultaneously protected marginal social strata from homelessness and hunger, and the society from the marginal social strata. They brought under the same roof all those who could not afford better accommodation--abandoned children, travelers, the sick, and the poor. In contrast to monastic institutions, they employed university-educated medical practitioners. This was the period when early-medieval type of religiousness, marked by asceticism, withdrawal from the worldly life, and contemplation, was replaced by the late-medieval "secular" type, which emphasized the need to act socially and charitably. Thus, the number of hospitals was often higher than what the population size required. The representatives of the secular type of religiousness were confraternities.

1. Urban communes, such as Florence, Padua, and _____, began urbanization.

   Venice
2. Hospitals were increasingly established by _____, confraternities, and rich individuals.

   Local governments
3. The number of hospitals was often _____ than necessary.

   Higher

4. The period discussed here was the _____ type.

   Early medieval

5. Hospitals protected _____ strata from homelessness.

   Marginal social

6. What were the hospitals considered institutions of?

   Social prevention

7. What were the representative of the secular type of religiousness?

   Confraternities

8. In addition to withdrawal from worldly life and asceticism, what was this period of religiousness marked by?

   Contemplation

9. This religiosity was replaced by the late medieval _____ type.

   Secular

10. In addition to children and travelers, who did hospitals take in?

    Sick and poor

Medieval Hospitals Section 4

Zadar, the capital of Venetian Dalmatia, had many hospitals. They were financed with lay money and managed either by priests or by laymen under the bishop's supervision. In 1295, the rich nobleman Cosa Saladin founded a hospital as part of a monastery for eight Franciscans with a chapel, garden, and pharmacy. Similarly, the hospital founded by Teodor de Prandino accommodated Franciscan third-order nuns whose duty was to take care of the poor. The list of founders comprised many well-known families, but the hospital founded by the rich merchant Grgur Mrganic in the mid-fifteenth century in the immediate proximity of the church of St.

Anastasia was especially famous. In his will, Mrganic specified the number and characteristics of the poor that should be admitted to the institution: 13 poor patients, Zadar citizens or foreigners, but no patients with plague. Although Zadar with its central administrative position and greater economic power had more and larger hospitals than either Split or Trogir, it was still subject to the vagaries of Venetian politics. As in Split, the naval wars of the seventeenth century caused the conversion of the city hospital of St. Mark into a military hospital.

1. Hospitals were either managed by _____ or laymen.

   Priests
2. Cosa Saladin founded a hospital for _____ Franciscans.

   Eight
3. The hospital founded by Teodor de Prandino accommodated _____ who took care of the poor.

   Third-order nuns
4. Grgur Mrganic specified that the hospital should only admit _____ patients.

   Thirteen poor
5. Zadar's hospitals were still subject to the _____ politics.

   Venetian
6. What is the capital of Venetian Dalmatia?

   Zadar
7. When did Cosa Saladin found a hospital?

   1295
8. The hospital founded by Cosa Saladin included what in addition to a chapel?

   A garden and pharmacy

9. When was the hospital founded by Grgur Mrganic?

   Mid-fifteenth century

10. What occurred in the seventeenth century that caused St. Mark to become a military hospital?

    Naval wars