UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE IMPACT OF INTERVIEW MODALITY AND TIMING EFFECTS ON MEMORY REPORTS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

MACKENZIE R. RIGGENBACH
Norman, Oklahoma
2021

THE IMPACT OF INTERVIEW MODALITY AND TIMING EFFECTS ON MEMORY REPORTS

A THESIS APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Scott Gronlund, Chair

Dr. Hairong Song

Dr. Jenel Cavazos

Abstract

Following a witnessed event, eyewitnesses are typically asked to give their report of the event. However, depending on the crime, witnesses may be interviewed differently. Previous research suggests that the way in which an individual is interviewed can impact the quantity and quality of the details reported. Research investigating different types of interviews suggests that having witnesses talk or write about an event can lead to different report qualities. Thus, one goal of the present study is to investigate how different types of interviews impact memory reports. Additionally, being tested on previously learned information has been shown to improve memory for that information compared to re-studying that information. Therefore, another goal is to examine how questioning participants at Test 1, compared to a second exposure to the information, impacts the quality of memory reports. Participants watched a short video clip and then were interviewed about its contents immediately, one week and one month later. Participants either wrote or spoke about what happened in the video. In general, writing leads to better quality memory reports compared to speaking and seems to carry over one week later. Therefore, having individuals provide a written compared to spoken report may be more effective in producing more accurate and detailed memory reports.

Table of Contents

The Impact of Interview Modality and Timing Effects on Memory Reports

Following a witnessed event, eyewitnesses are typically asked to give their account of the event. However, depending on the crime, a witness may be interviewed differently. For example, for serious crimes, spoken reports are typically obtained, but for civil procedures, a written report is usually collected. Prior research has investigated the numerous ways in which investigators interview witnesses and how that can impact the number and types of details reported. Consequently, when examining interview modality, the limited research reveals two competing ideas: Evidence supports both a written superiority effect and a spoken superiority effect. Thus, it is of key interest to investigate how the different modalities in which an individual is questioned can impact both the quantity and quality of the details reported.

Research that supports a written superiority effect suggests that writing is better because it allows for self-pacing and the ability to monitor what information has previously been produced. Sauerland et al. (2014) found that, in general, written free recalls led to better memory performance compared to speaking. Kraus et al. (2017) conducted several types of interviews after participants watched a video of a criminal event. Self-administered interviews (SAI), police officer questioning (POQ), and written free recall (FR) techniques were used for questioning. The SAI, a structured questionnaire that witnesses fill out, led to reports of more correct victim and setting details compared to the participants in the POQ or FR conditions. The SAI group also reported more correct offender and action details compared to the FR group. However, the POQ group did report more offender details compared to both the SAI and FR groups. This study suggests that writing (SAI) in general leads to better memory performance compared to speaking (POQ), although it is possible that the way in which different written interviews occur can impact the quality and quantity of eyewitness' reports. In contrast, other work suggests that writing places higher demands on working memory because writing is slower than talking, less practiced, and requires activation of grapheme representations for spelling words (Kellogg, 2007). It is this increase in cognitive load when writing that is thought to negatively impact performance on long-term memory retrieval (Moscovitch, 1994).

On the other hand, research that supports a spoken superiority effect suggests that speaking about an event leads to better memory performance because it demands fewer cognitive resources (Kellogg, 2007). Sauerland and Sporer (2011) found that having participants talk about a video event led to more detailed and accurate crime descriptions and more accurate central perpetrator details, but writing was better for reporting peripheral perpetrator details. However, it is important to note that, although speaking is considered more productive, it is not necessarily more efficient. For example, when speaking, individuals may repeat what they have previously stated. Mechanisms thought to induce a spoken superiority effect include that speaking requires less muscular energy, is acquired earlier in life, and therefore is easier and more practiced (Sauerland & Sporer, 2011). As a result, speaking is thought to lead to a lower level of cognitive demand. Consequently, if cognitive load is low, as it is thought to be when speaking, it is possible that individuals have more working memory capacity available to report and describe details that require more effortful retrieval.

Previous research also has shown that the act of being tested on previously encoded information provides another instance of learning that material. This testing effect, the ability to learn from being tested compared to the restudying of the information, provides an avenue to examine how the testing effect is impacted by the modality of testing. Roediger and Karpicke (2006) showed that immediately testing individuals after reading a passage led to better long-term retention rates compared to merely restudying the passage. This effect remained even after a retention interval of a week. Additionally, research suggests that rates of forgetting are exacerbated following an increase in retention interval as a function of restudying versus repeated testing (Wheeler, Ewers, & Buonanno, 2003). More specifically, the forgetting of 40 items over an interval of seven days occurred much faster in the study-only condition compared to the repeatedly tested condition.

Does the modality of testing mediate this testing effect? Given that testing typically improves memory accuracy compared to repeatedly studying the information, parsing out potential differences between a written test compared to a spoken test is important, especially as it relates to the types of details correctly recalled. For example, some research suggest that testing can increase the rate of semantically related false memories when there is a theme within a set of stimuli (McDermott, 2006). Thus, investigating the impact of repeated testing and test modality on memory accuracy for both true and false information is critical. Additionally, the type of correct details reported is important to consider. In a study involving children, Bouwmeester and Verkoeijen (2011) found that the testing effect was related to differences in gist processing. Children who were able to process stronger gist traces during retrieval practice benefited the most from testing. Bouwmeester and Verkoeijen (2011) suggest that the semantic overlap of the learning material is fundamental to the presence of a testing effect. It is important to note that this study implemented Deese-Roediger-McDermott (DRM) lists, which may not directly transfer to other sorts of stimuli. These DRM lists are comprised of a set of words that are semantically related to a given theme but the theme's classifying word is not actually presented (Deese, 1959; Roediger & McDermott, 1995). For example, participants may be presented with a list of words that all fall into the category of sleep (i.e., dream, bed) but the word "sleep" is never actually presented. However, participants are susceptible to incorrectly reporting the word "sleep" as previously being studied. Given the differential mechanism of gist and verbatim recall, investigating differences in report qualities for varying detail type is crucial.

Research suggests that different types of information can be recalled on the basis of how that information was previously encoded. For example, Fuzzy-trace theory posits that individuals encode details of an event as a function of gist and verbatim information (Reyna & Brainerd 1995; Brainerd & Reyna, 2005). The theory assumes that individuals encode both the general idea of an event along with verbatim details about it. According to fuzzy-trace theory (Reyna, V. F. & Brainerd, 1995), studying is thought to promote verbatim processing while testing is thought to promote gist processing (Bouwmeester & Verkoeijen, 2011). Consequently, semantic relatedness can differentially impact the presence of a testing effect and some research suggests that when a theme is present within a set of stimuli, the enhancement of gist processing associated with testing may serve as a helpful retrieval cue. Contrarily, restudying may be more effective at promoting retention through the enhancement of verbatim processing (Delaney, Verkoeijen, & Spirgel, 2010).

In general, prior research supports the notion that writing versus speaking has differential effects on an individual's ability to accurately report information. However, studies that have investigated the differences in interview modality have minimally investigated how the passage of time can impact these types of memory reports and research that does so typically uses a retention interval of around a week or shorter. It is possible that a witness may be interviewed many times following a witnessed event and it is likely that their memory reports change over time as a function of subsequent interviews. Therefore, exploring the timing of interviews is another key component to consider. Kraus et al. (2017) found that those who completed the SAI immediately after observing the crime reported more correct details without a loss of accuracy one week later and had higher accuracy in the Cognitive Interview (CI, Geiselman, 1984) compared to participants in the FR and no-initial interview group. Additionally, Warren and Lane (1995) manipulated the type of initial test (no test, neutral, or misleading) as well as the type of second test that occurred one week later (no test, neutral, or misleading). They found that immediate neutral testing led to an enhancement in inoculating against forgetting and suggestibility. Pansky and Tenenboim (2011) found results consistent with Warren and Lane (1995) but implemented a delay period of 48 hours. Nonetheless, it is crucial to investigate how memory reports change over longer periods of time even while examining a testing effect because some research has suggested that being immediately tested can improve final recall even one month after initial encoding (Butler & Roediger, 2007).

Thus, given the limited and contradictory evidence of the effect of interview modality on memory reports, in conjunction with the differential effects of interview timing, the goal of the present study is to gain a greater understanding of how interview modality and timing interacts to impact memory reports. That is, does writing or speaking differentially impact memory reports for an event immediately, 1-week, and 1-month following initial encoding? We expect the participants in the written condition to report more correct details in general compared to participants in the spoken condition. We also expect that a written superiority effect is likely to be present when tested immediately. Additionally, we anticipate that participants who write or speak during the Test 1 phase of the experiment will show a testing benefit overall.

Experiment 1

Method

Participants

A total of 125 introductory psychology students (29 males, 96 females; $M_{\text{Age}} = 19.04$ years, $SD_{\text{Age}} = 2.21$) from the University of Oklahoma ($N = 95$) and Ohio Northern University[1] ($N = 30$) participated in this study in exchange for partial course credit. All students were enrolled in an introductory psychology course and were recruited via a university recruitment portal (SONA study flier). The flier informed potential participants that they would watch a video and then be asked questions about the video at three different timepoints. Participants received a maximum of 2.5 research credits for their psychology course. They received credit following the completion of two laboratory sessions and one email response. To participate, students were at least 18 years of age and able to provide consent or had received parental consent if they were less than 18 years of age. In addition, participants indicated that they were proficient in English.

Participants were randomly assigned to one of four conditions. The four conditions included Control-Written ($n = 30$), Control-Spoken ($n = 33$), Written ($n = 31$), and Spoken ($n = 31$). All participants' data were anonymous and kept separate from identifying information. No significant risks were encountered by the participants, and they were treated in accordance with APA (American Psychological Association) ethical standards. The study was approved by both the University of Oklahoma IRB (Institutional Review Board) and Ohio Northern University IRB.

[1] Dr. Phillip R. Zoladz, along with three trained undergraduate research assistants from the Department of Psychological Sciences at Ohio Northern University (Ada, OH), helped with data collection for Experiment 1.

Materials

Participants completed a demographic survey that involved them self-reporting their gender and age. They then viewed an eight-minute excerpt from the Disney movie *Looking for Miracles* (Grant & Sullivan, 1989), which depicts the adventures of two brothers at summer camp. This video served as the witness event. This video was chosen as the witnessed scenario because it is an older film and is unlikely to have been previously seen by the participants. After watching the video, all participants were asked whether they had seen the video (n = 0). This video also was used because the dynamics of each scene allowed participants to have opportunities to report a multitude of different details. Following the video, participants were asked questions about the video at different timepoints. Both the video and question materials are like those used in previous studies (Zaragoza, Payment, Ackil, Drivdahl, & Beck, 2001; Zoladz et al., 2017). Depending on the condition, participants either wrote about what they saw in the video or spoke about it into an audio recorder. A pilot study (n = 11) was conducted to determine which reported details would be classified as either central or peripheral to the video.[2]

The interviews consisted of open-ended questions, pointed questions, and a combination of the two. The open-ended questions were general questions asking about each of the three main scenes in the video. For example, "The first scene took place in the dining hall. Please talk about what events occurred, who was in the scene, describe the people who were there and any other details that you can remember, such as, did any important conversations happen?" This same format was repeated for the second and third scenes. The pointed questions were questions that asked about something specific instead of allowing the participants to report whatever they could remember. Some of these questions were true and some were false. A true question asked about an event or detail that appeared in the video, whereas a false question asked about an event or

[2] During this pilot study, participants watched the video and were asked to report everything that they could remember. Based on this, details reported by more than six of the participants were classified as central details and details reported by five or fewer of the participants were classified as peripheral details. On average, a central detail was reported by 7.5 participants and a peripheral detail was reported by 3.7 participants.

detail that was plausible but did not occur in the video. There was a total of eight true questions and five false questions asked. An example of a true question is “The cook brought out a cake because it was one of the boy’s birthdays. What did the cake say?” An example of a false question is “After Delaney fell, where did he say that he injured himself?” The pointed questions served to introduce false information to the participants to determine how the susceptibility of false information might change depending on interview modality. Participants were not forced to answer; they could indicate that an event did not happen or that they could not remember an answer. Interviews that occurred during the first lab session only consisted of the open-ended questions. Interviews that occurred during the second lab session and via email consisted of the same open-ended questions plus the addition of the pointed questions.

Design and Procedure

The present study is a 4 (Interview Modality: Control-Written, Control-Spoken, Written, or Spoken) x 3 (Interview Timing: Immediate, 1-week delay, and 1-month delay) mixed design. Interview modality is the between-subjects factor and interview timing is the within-subjects factor.

After obtaining informed consent from all participants, participants were asked if they would provide their cell phone number to the researcher to receive session reminders throughout their one-month sequence of sessions. Participants were not required to provide their cell phone number. Next, all participants completed the demographic survey. Following completion of the survey, all participants watched the video. The previously mentioned procedural steps were identical for all participants. It is at this point that the procedure changes depending on condition[3].

Control Conditions. Following the conclusion of the video, participants in the two control conditions (Control-Written and Control-Spoken) watched the video again. Control participants were not interviewed during their first lab session. After the control participants

[3] *Figure 9* in the Appendix illustrates the differences between all four conditions.

finished watching the video for the second time, they were asked if they had ever seen the video. The conclusion of the control participants' first lab session included being reminded of their next session and then being dismissed from the lab. The purpose of watching the video twice, and not being interviewed immediately, was to investigate how repeated exposure to an event, without being immediately tested, compared to the memory reports of individuals who were exposed to an event only once but were immediately tested. In other words, studying the video twice provides an opportunity to investigate the differential effects of interview modality on the testing effect.

One week later, control participants returned to the lab and were randomly assigned to either the Written (Control-Written) or Spoken (Control-Spoken) condition. Before beginning the actual interview, participants were asked two warm-up questions: What were the two main character's names (Delaney and Sullivan)? If participants answered incorrectly, the researcher informed them of the correct answers. If participants answered correctly, the researcher informed the participant that he or she was correct and then the researcher continued with the actual interview. The purpose of having the participants identify the characters, and be corrected if they answered incorrectly, was to make sure that they could correctly reference the two main characters in the video. During the interview, participants were first queried with the open-ended interview questions and then with the pointed questions. However, depending on the condition, participants either wrote out their responses on lined sheets of paper (Control-Written) or spoke into an audio recorder (Control-Spoken). Following the interview, participants were debriefed. They also were reminded that they would receive an email in three weeks to complete the final phase of the experiment. One month following participants' first lab session, participants received an email containing the pointed and open-ended interview questions along with instructions for how to complete the interview. Participants were given one week to return their responses to the researcher. The email interview should have taken participants 15 minutes to complete.

Experimental Conditions**.** Participants randomly assigned to the experimental conditions (Written or Spoken) watched the video only once and then were immediately interviewed. Participants were first asked three warm-up questions. The first asked whether they had seen the video before. The next two warm-up questions asked about the two names of the main characters. If participants answered incorrectly the researcher informed them of the correct answers. If participants answered correctly, the researcher confirmed the participants' answers and then continued with the actual interview. The interviews began with open-ended questions. Those in the written condition (Written) wrote their responses to the questions on lined sheets of paper whereas those in the Spoken condition (Spoken) spoke their responses into the audio recorder. Following the interview, participants were reminded of their next session and were dismissed.

One week later, participants returned to the lab and were queried using the open-ended questions and then the pointed questions. Interview modality was not mixed. That is, participants in the Written condition during their first session were also asked to write their responses during their second session. Following the interview, participants were debriefed and then reminded that they would receive an email to complete the final phase of the experiment in three weeks. Participants received the email interview one month after their first session. The email contained the pointed questions and the open-ended questions along with instructions for how to complete and return their responses. All participants were given one week to respond to the questions and return them to the researcher.

It is important to note that a researcher was present throughout the entirety of the first and second sessions, regardless of condition type. The experimenter set up the video to be played and then sat across from the participant who was seated at a desk in the laboratory. The researchers sat across from the participants during the entire session to conduct the interview and record the Spoken responses. Researchers also recorded the time it took to complete the interview, though participants were free to take as much time as they needed. Also, all lab sessions had to be completed sequentially. That is, to participate in the email session, the first and second lab sessions must have been completed prior to and within the allotted timeframe.

For all types of interviews, the research assistants were trained on the proper protocol for interviewing participants. The researcher assistants responded to each answer with a transitory comment such as "Okay, the next question is...". This was meant to reduce possible chances of confirmatory feedback or other cues that may have indicated to the participant the verity of their responses (Zaragoza et al., 2001).

Results

A total of 125 participants completed the first lab session. Of those, 118 (94.4% return rate) participants completed their second lab session, and 96 (76.8% response rate) participants completed the email response. Only participants who completed at least the first and second lab sessions were included in the subsequent analyses. However, before conducting any analyses, the data were cleaned, which resulted in an additional 10 participants being removed because of technical error or incomplete data. Therefore, 115 participants' data were used for the Test 1 and Test 2 data analyses, and 91 participants' data were used for the Test 3 analyses.

All audio-recordings were transcribed before coding. Interviews were coded by three individuals. Interrater reliability scores for the coding of all interviews ranged between a Kappa value of 0.77 and 1.00. All coding disagreements were discussed amongst the coders until a mutual decision could be made. Open-ended responses were coded for central and peripheral details as well as intrusions and any other detail that was reported correctly but not deemed as either central or peripheral to the video (according to the pilot study).

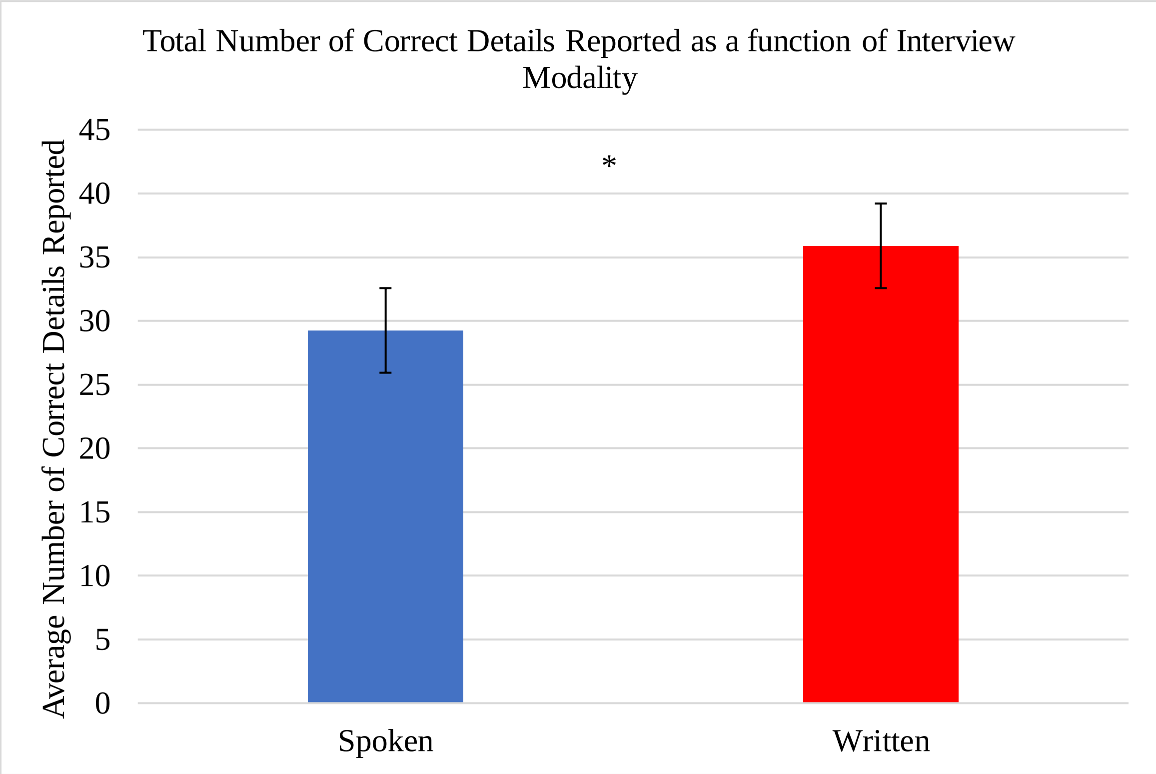
Test 1

The Test 1 data were first checked for outliers resulting in the removal of two participants' correct other details reported and one participant's total number of intrusions reported. Next, tests of normality indicated that for Test 1, the central, peripheral, and correct other details reported were normally distributed. However, the number of intrusions reported for participants in the Spoken and Written conditions were not (Shapiro-Wilk normality test, W = .926, df = 53, p = .003); these data were log transformed before conducting any analyses. Additionally, the number of words either spoken or written during the Test 1 interview and the time in seconds of those interviews were not normally distributed and were log transformed. These transformations fixed the non-normalities (W = .981, df = 57, p = .486 and W = .959, df = 52, p = .069 for the word counts and interview time, respectively). To reiterate, participants in the two control conditions (Control-Written and Control-Spoken) were not tested during the Test 1 phase of the study.

The total number of central, peripheral, and correct other details reported during Test 1 were combined to reflect the overall correct number of details reported for each participant. A one-way ANOVA was conducted to compare the total correct number of details reported by the participants in the Spoken and Written conditions; $F(1, 53) = 6.247$, $p = .016$, $\eta^2 = .105$. Participants in the Written condition (M = 35.89, SD = 7.40) reported significantly more correct details overall compared to participants in the Spoken condition (M = 29.25, SD = 11.73). These findings suggest that, in general, writing immediately following encoding improves memory reports compared to speaking about what transpired. These findings support the hypothesis regarding the written superiority effect. This effect is illustrated in Figure 1[4].

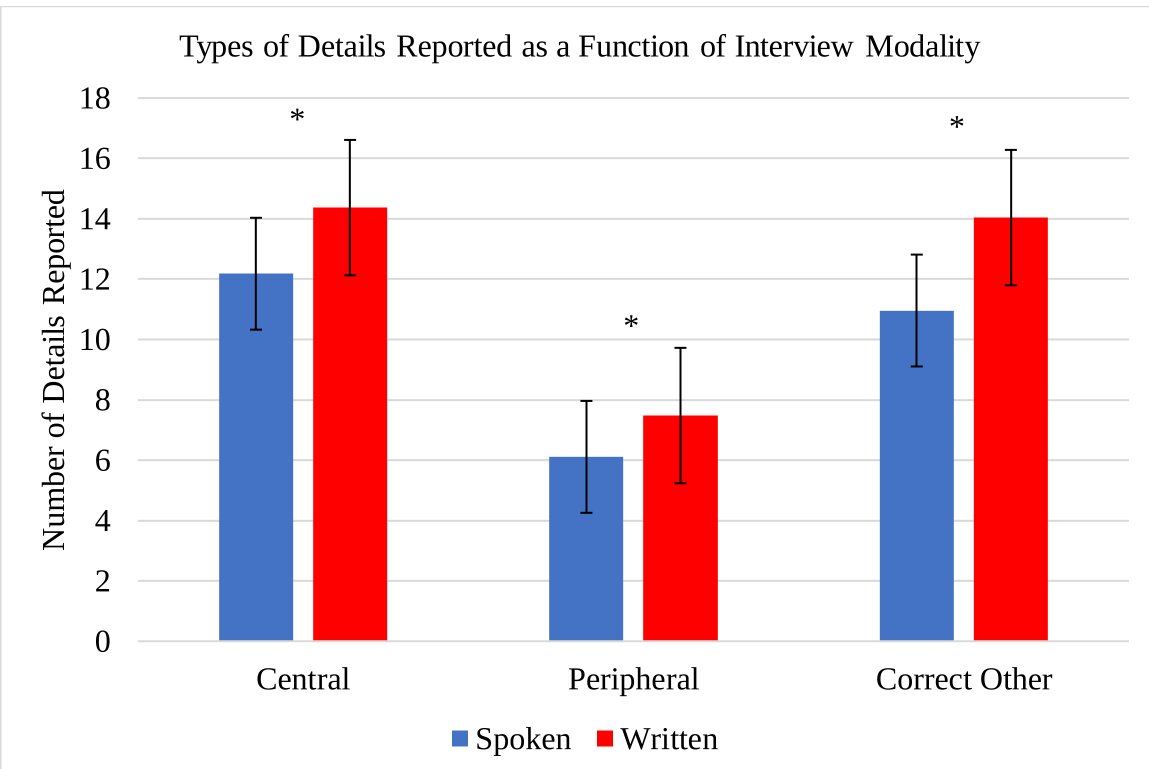
[4] All error bars presented represent standard error.

Figure 1



A MANOVA was conducted to compare the number of central, peripheral, and correct other details in the Spoken and Written conditions. There is a significant difference in the number of central details reported between the two conditions: $F(1, 53) = 5.659$, $p = .021$, $\eta^2 = .096$, such that during Test 1, participants in the Spoken condition ($n = 28$) reported significantly fewer ($M = 12.18$, $SD = 3.87$) central details than participants in the Written ($n = 27$) condition ($M = 14.37$, $SD = 2.87$). Additionally, there is a significant difference in the number of peripheral details reported: $F(1, 53) = 3.961$, $p = .052$, $\eta^2 = .070$, with participants in the Spoken condition ($M = 6.11$, $SD = 2.59$) reporting fewer peripheral details compared to participants in the Written condition ($M = 7.48$, $SD = 2.53$). Lastly, there is a significant difference in the number of details that were reported correctly but not deemed central or peripheral: $F(1, 53) = 4.617$, $p = .036$, $\eta^2 = .080$. Again, participants in the Spoken condition ($M = 10.96$, $SD = 6.62$) reported fewer correct other details compared to the Written condition (M = 14.04, SD = 3.42). The top panel of Figure 2 illustrates these findings and shows the number of each detail type reported as a function of interview modality. It is important to note that the maximum number of central and peripheral details that could have been reported are 21 and 18 details (associated proportions are reported in the bottom panel of Figure 2), respectively. These findings provide further support for the hypothesis that participants in the Written condition perform better than participants in the Spoken condition across all types of details examined.

Figure 2



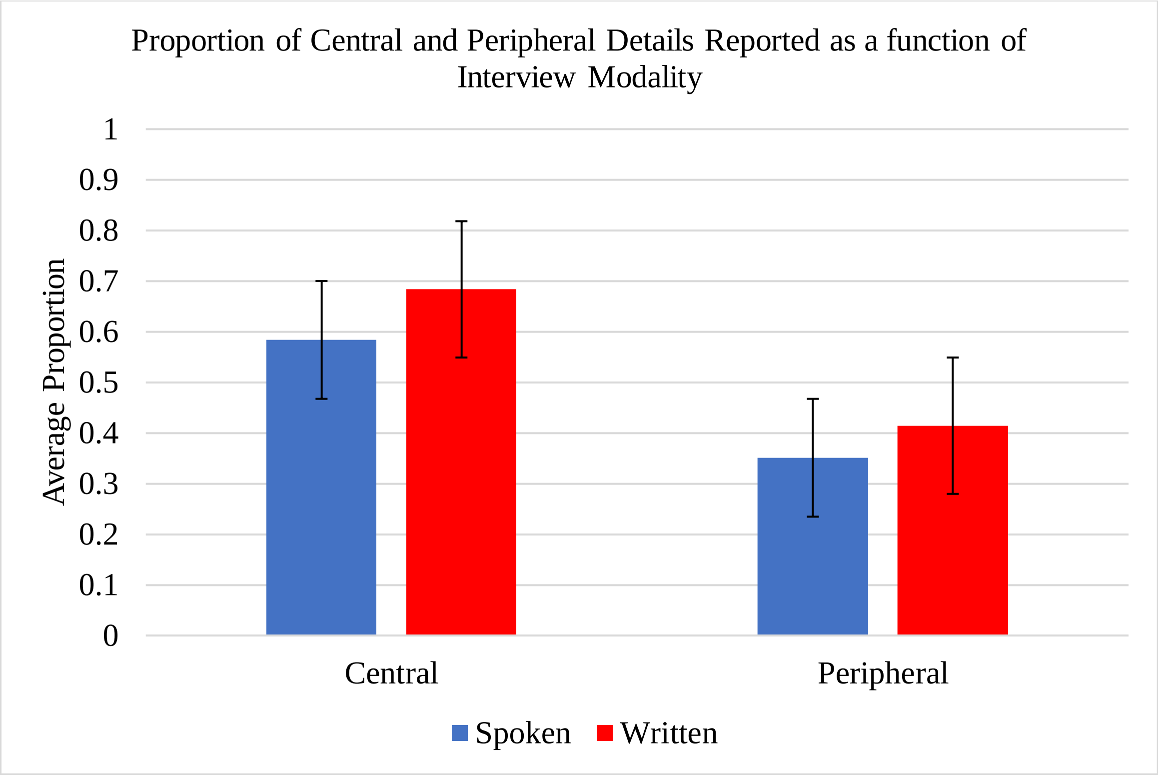
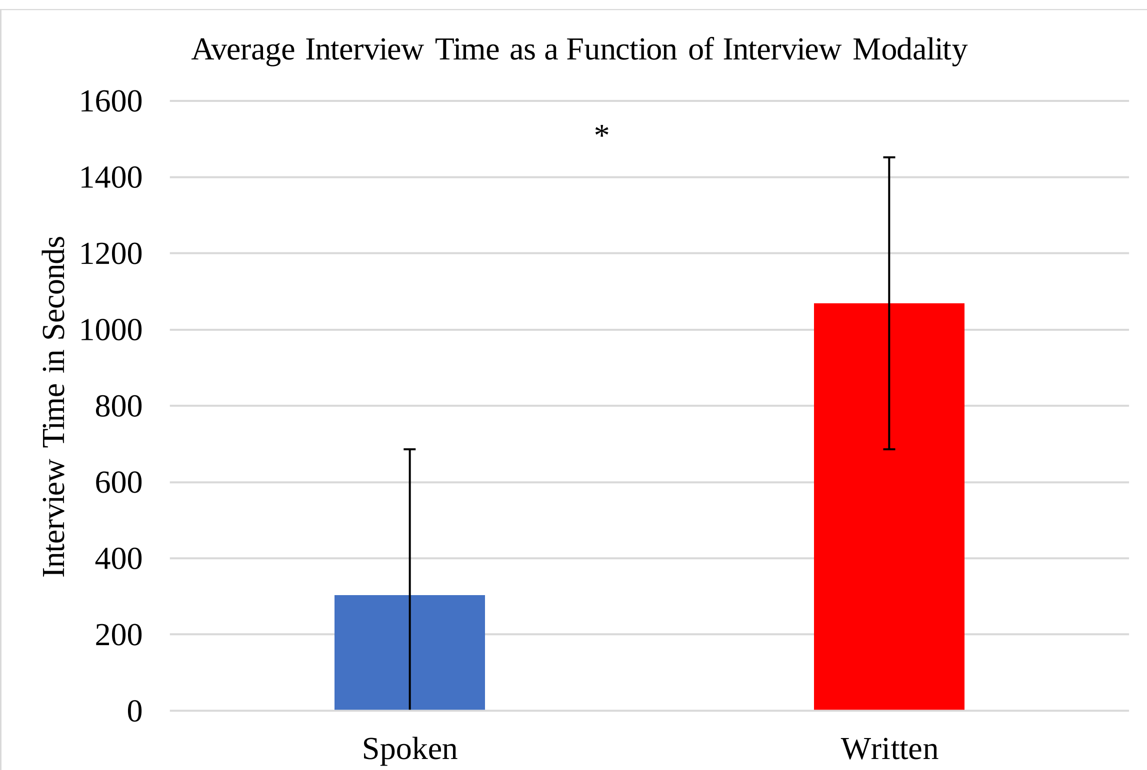


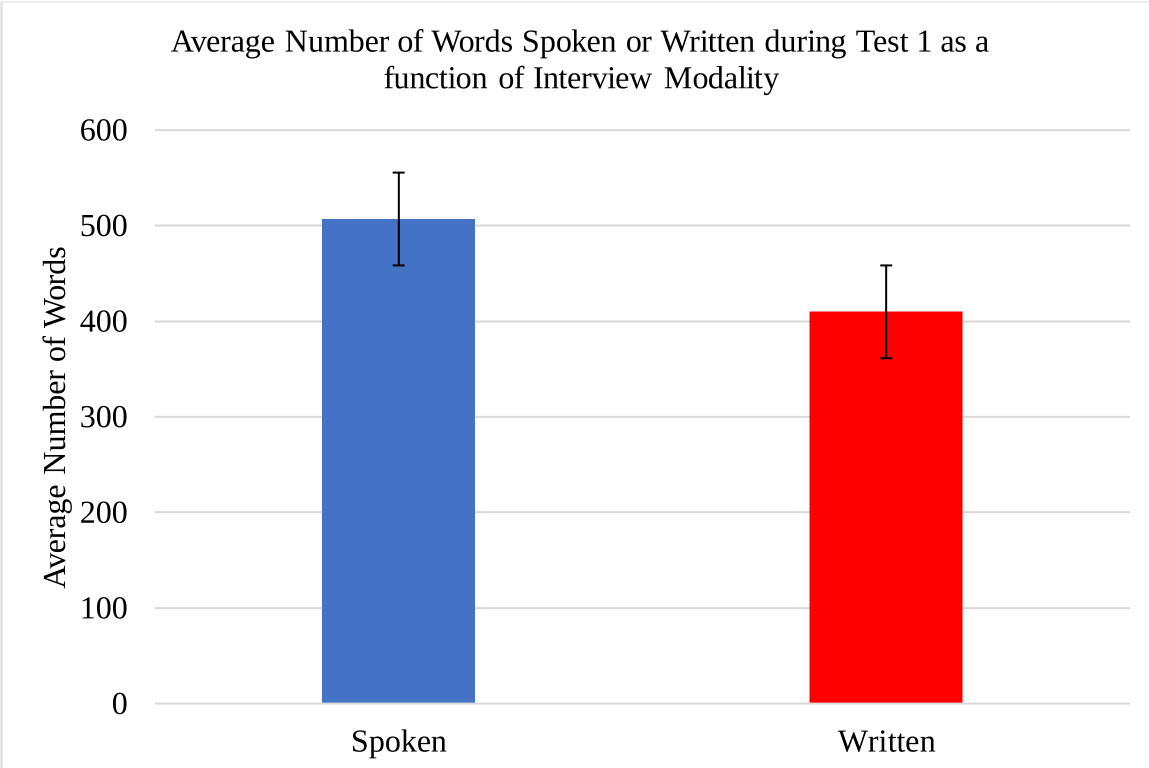
Figure 2. The maximum number of details that could have been reported as either central or peripheral are 21 and 18 details, respectively.

The next set of analyses were conducted to determine if participants in the Written condition performed better because they worked more productively and proficiently than participants in the Spoken condition. An ANOVA examining the differences between interview completion times for the Written and Spoken conditions was conducted. There is a significant difference in the average number of seconds that the interviews took to complete: $F(1, 50) = 175.130$, $p = .001$, $\eta^2 = .778$, such that, participants in the Spoken condition ($n = 29$) took significantly less time ($M = 303.00$, $SD = 118.00$) to complete the interview compared to the participants in the Written ($n = 23$) condition ($M = 1069.91$, $SD = 347.71$). These findings can be seen in the top panel of Figure 3. This finding is not surprising given that speaking is faster than writing. However, another ANOVA compared the average number of words spoken or written during the interviews; there is no significant difference in the overall word counts between the two conditions: $F(1, 55) = 1.063$, $p = .307$, $\eta^2 = .019$. Therefore, given that approximately the same amount of information is being reported in the two conditions, this suggests that participants are working equally hard in each condition, but participants in the Written condition are working more efficiently (see bottom panel of Figure 3).

Lastly, another ANOVA compared the number of intrusions reported by the participants in the Spoken and Written conditions. The average number of intrusions reported was not significantly different between the two conditions: $F(1, 51) = .191$, $p = .664$, $\eta^2 = .004$. This suggests that during the Test 1 phase, participants in the Written condition are not more resilient against reporting incorrect information even if they are more likely to report more correct information compared to the participants in the Spoken condition.

Figure 3



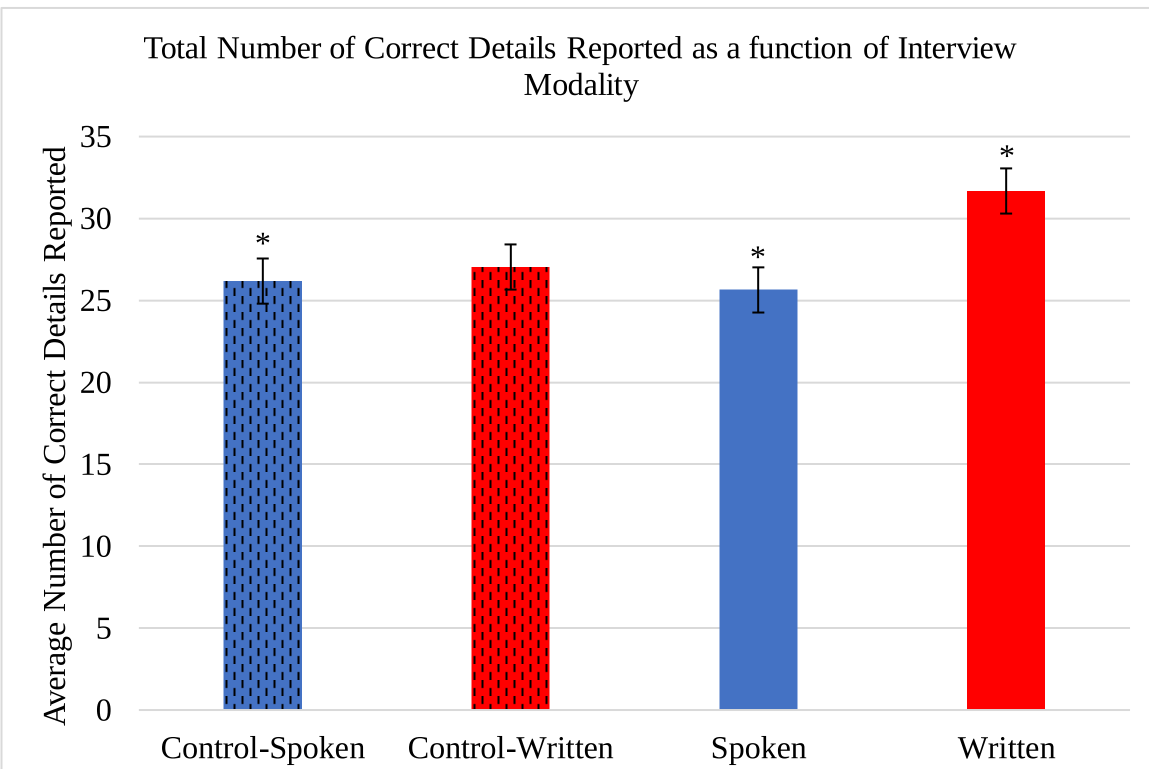


Test 2

The Test 2 data were checked for outliers and transformed when necessary. The number of words spoken or written during the interviews, the time to complete the interview, along with the number of peripheral, intrusions, and other correctly reported details needed to be transformed. Additionally, the proportion of true interview questions answered correctly, the proportion that participants incorrectly answered a false question, and the proportion participants correctly rejected a false question was transformed. Following all transformations, nine outlying data points were removed before conducting any analyses. Only individual data points were removed instead of removing the entire participant from the analyses.

The total number of central, peripheral, and correct other details reported during Test 2 were aggregated to reflect the overall correct number of details reported for each participant. A one-way ANOVA was conducted to compare the total correct number of details reported as a function of interview modality; $F(3, 108) = 3.393$, $p = .021$, $\eta^2 = .086$[5]. Participants in the Written condition ($M = 31.68$, $SD = 7.52$) reported significantly more correct details compared to participants in the Spoken condition ($M = 25.64$, $SD = 9.00$). Participants in the Written condition ($M = 31.68$, $SD = 7.52$) also reported (marginally significantly: $p = .064$) more correct details overall compared to participants in the Control Spoken condition ($M = 26.18$, $SD = 9.02$), which is indicative of a testing effect. These findings suggest that initial written recall improves memory reports, even after a one-week delay, compared to initial spoken recall, and to some degree, initial written recall improves memory reports compared to re-studying the original event a second time. The top panel of Figure 4 illustrates these findings.

Figure 4



[5] Only Bonferroni-adjusted pairwise comparisons with a *p*-value less than .05 are reported.

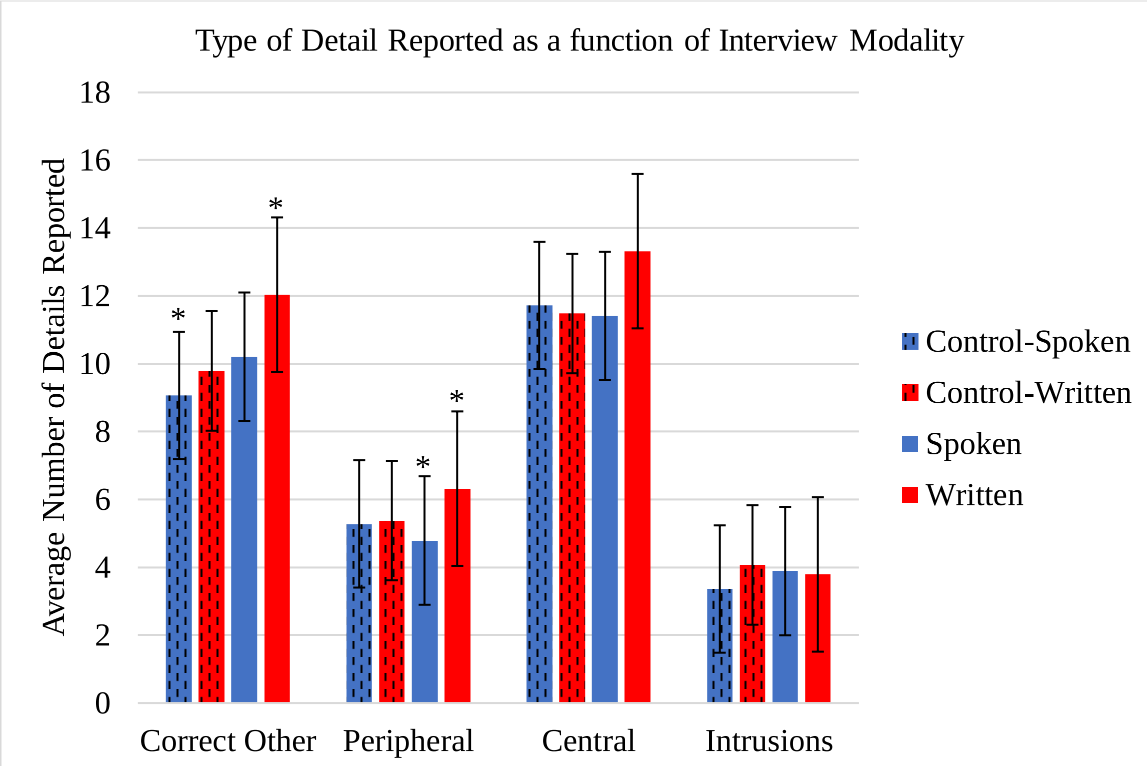


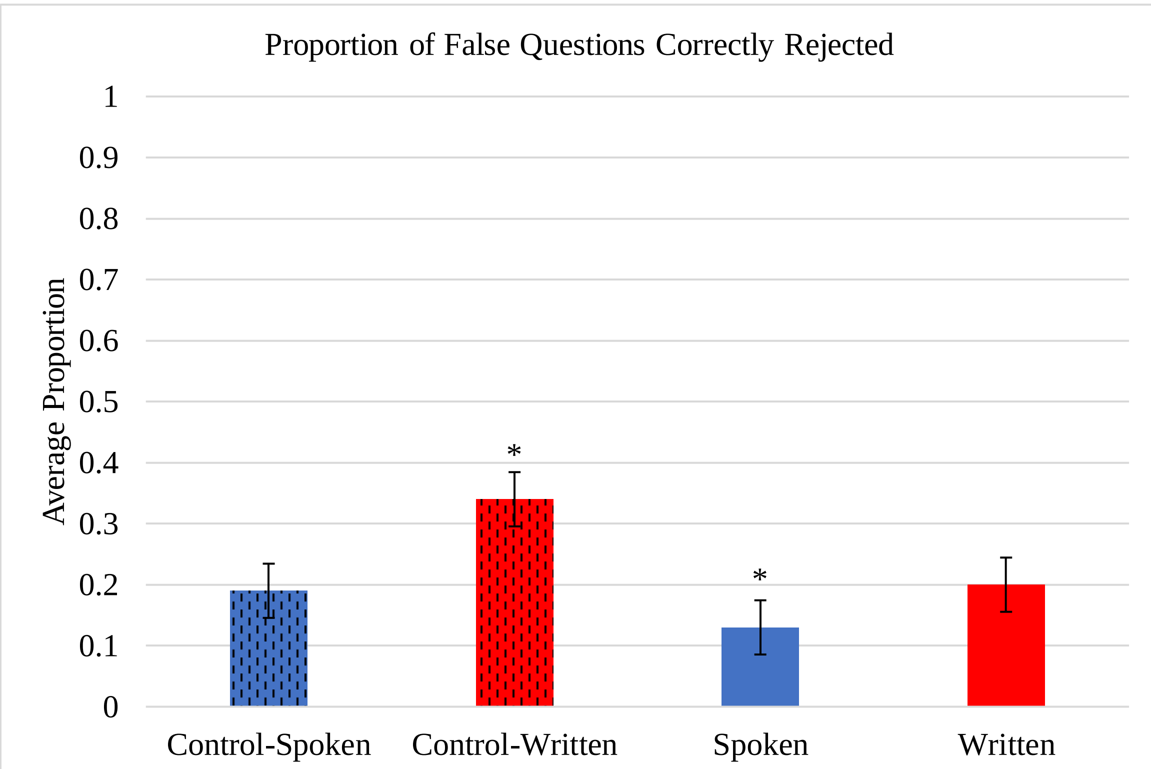
Figure 4. The max number of details that could have been reported as either central or peripheral are 21 and 18 details, respectively.

Multiple one-way ANOVAs were conducted on the disaggregated Test 2 data to better understand the effects that interview modality had on the testing effect. This was preferred to conducting a repeated-measure ANOVAs using the Test 1-Test 2 independent variable because the two control conditions (Control-Written and Control-Spoken) were not tested during the Test 1 phase. There is a significant difference between interview modality on the number of correct other details reported; $F(3, 108) = 2.806$, $p = .043$, $\eta^2 = .072$. More specifically, there is a marginally significant difference ($p = .06$) between the number of extra correct details reported, such that participants in the Written condition ($M = 12.04$, $SD = 3.73$) reported more extra correct details than participants in the Control-Spoken condition ($M = 9.07$, $SD = 4.37$). This is indicative of a cross-modality (Written to Control-Spoken) testing effect, but not a within modality (Written to Control-Written) testing effect. However, given the numerical difference (not statistically significant) between the Control-Written and Control-Spoken conditions, it suggests that writing improved performance overall, even when not tested initially. In other words, the benefit of recalling by writing may mitigate the anticipated enhancement due to testing.

Additionally, there is a marginally significant main effect of interview modality on the number of peripheral details reported; $F(3, 108) = 2.468$, $p = .066$, $\eta^2 = .064$. Participants in the Written condition ($M = 6.32$, $SD = 2.28$) reported more peripheral details compared to participants in the Spoken condition ($M = 4.79$, $SD = 2.38$). This also suggests that writing leads to better memory reports than speaking and that this enhancement may remain after a one-week delay. There is not a significant main effect of interview modality on the number of central details correctly reported ($F(3, 111) = 2.058$, $p = .110$) or the number of intrusions reported ($F(3, 102) = .656$, $p = .581$). The bottom panel of Figure 4 shows the average number of details reported as a function of interview modality. Given that there is no significant difference in the number of central details reported, it is possible that it becomes easier to report more gist pieces of information compared to verbatim as retention interval increases. That is, it was easier for participants to report the gist of the event, compared to verbatim information, one week following encoding.

When examining the impact of interview modality on the testing effect, multiple one-way ANOVAs were conducted on the proportions recorded for the pointed questions. There is a significant difference of interview modality on the correct rejection of false questions; $F(3, 59) = 3.329$, $p = .026$, $\eta^2 = .145$. Participants in the Control-Written condition ($M = .34$, $SD = .25$) had a greater correct rejection rate than participants in the Spoken condition ($M = .13$, $SD = .18$). In other words, participants in the Control-Written condition identified the false questions as asking about a false detail at a greater rate than participants in the Spoken condition. This is another indication that the benefit due to writing may override the anticipated enhancement of a testing effect. There is no significant difference of interview modality for the proportion of true questions answered ($F(3, 81) = 2.354$, $p = .078$) or the proportion of false questions endorsed ($F(3, 76) = 1.851$, $p = .145$). These findings are illustrated in Figure 5.

Figure 5



A series of one-way ANOVAs were conducted to examine if the greater efficiency of the Written condition observed in Test 1 carried over to the Test 2 phase. The interview times (converted to seconds) of Test 2 were analyzed using a one-way ANOVA. Unsurprisingly, participants in the writing conditions took significantly more time compared to participants in the speaking conditions, replicating the results from Test 1.

Of greater interest is the differences in the interview word counts. The word counts were divided between the number of words reported during the open-ended portion and in the pointed questioning portion of the interview. There is a significant difference in the number of words reported during the open-ended portion of questioning; $F(3, 110) = 5.475$, $p = .002$, $\eta^2 = .130$. Participants in the Control-Written condition ($M = 314.59$, $SD = 93.34$) had lower word counts compared to both the Control-Spoken ($M = 478.21$, $SD = 232.34$) and Spoken ($M = 461.43$, $SD = 205.62$) conditions. Additionally, participants in the Written condition ($M = 339.36$, $SD = 99.75$) had marginally ($p = .086$) lower word counts compared to participants in the Control-Spoken condition. For the pointed questioning word counts, there is also a significant difference in the number of words reported; $F(3, 111) = 4.760$, $p = .004$, $\eta^2 = .114$. Participants in the Written condition ($M = 76.46$, $SD = 29.60$) had significantly lower word counts compared to participants in both the Control-Spoken ($M = 104.76$, $SD = 37.02$) and Spoken ($M = 100.14$, $SD = 33.59$) conditions. These findings, taken together, show that a written superiority effect remains even one week after encoding. Moreover, despite the lower word counts, participants who write perform more efficiently than participants who speak.

Test 3

The Test 3 data were checked for outliers and transformed when necessary[6]. Following all transformations, one outlying data point was removed before conducting any analyses. The time to complete the interview was not recorded for Test 3 because this test occurred via email. The next set of analyses all produced results similar to Test 1 and Test 2, however, most were

[6] The number of words reported during the interviews, along with the number of peripheral, intrusions, and extra correctly reported details were all log transformed. Additionally, the proportion of true interview questions answered correctly, the proportion that participants incorrectly answered a false question, and the proportion participants correctly rejected a false question, were logit transformed.

only marginally significant. Even one month following encoding, there seems to be a numerically relevant difference between those who write compared to those who speak. However, the advantages associated with writing are only marginally significant one month following initial encoding.

The following set of analyses were conducted to examine if the writing superiority effect seen in Test 1 and Test 2 carries over to Test 3. The total number of central, peripheral, and correct other details reported during Test 3 were aggregated to reflect the overall correct number of details reported for each participant. A one-way ANOVA was conducted to compare the total correct number of details reported as a function of interview modality; $F(3, 84) = 1.421$, $p = .242$, $\eta^2 = .048$. There is not a significant difference in the total number of correct details reported across interview modality. These findings suggest that any enhancement writing induced immediately and at a one-week delay, does not hold up following a one-month retention interval. Multiple one-way ANOVAs conducted on the disaggregated Test 3 data revealed a marginally significant difference between interview modalities on the number of peripheral details reported; $F(3, 84) = 2.526$, $p = .063$, $\eta^2 = .083$. This is driven by a marginally significant ($p = .074$), pairwise comparison with participants in the Written condition ($M = 5.30$, $SD = 3.01$) reporting more peripheral details than participants in the Spoken condition ($M = 3.18$, $SD = 2.02$). Again, this provides some marginal support that writing improves memory reports compared to speaking, even one month following encoding. There is not a significant difference between interview modalities on the number of central details reported ($F(3, 87) = .843$, $p = .474$, $\eta^2 = .028$), intrusions reported ($F(3, 81) = .408$, $p = .747$, $\eta^2 = .015$), or correct other details reported ($F(3, 84) = .355$, $p = .785$, $\eta^2 = .013$).

The previous set of analyses suggests that the writing superiority effect only marginally extends to one month after encoding, however, these analyses are representative of the performance for the open-ended questions. When examining the impact of interview modality on the testing effect over time, multiple one-way ANOVAs were also conducted on the proportions recorded for the pointed questions. There are no significant differences between interview modalities on the proportion of true questions answered, the proportion of false questions endorsed, or the proportion of false questions correctly rejected as occurring (all p-values greater than .15).

The word counts were analyzed separately for the number of words reported for the open-ended portion and the pointed portion of questioning. There is no significant difference in the number of words reported in the open-ended portion of the question ($F(3, 86) = .592$, $p = .622$) or the pointed portion of the interview ($F(3, 87) = 1.010$, $p = .392$).

Experiment 2

Experiment 2 was conducted to further evaluate the findings of Experiment 1. The COVID-19 pandemic created a natural experiment examining the impact of removing some of the social factors that may impact interviewee performance. Experiment 2 was identical to Experiment 1 with the exception that the lab sessions occurred via Zoom rather than in-person. In addition to the various cognitive factors that may impact memory retrieval, there are social factors that may impact performance, like the presence of an interviewer. Bergmann, Jacobs, Hoffmann, and Boeing (2004) had patients complete a written questionnaire and personal interview related to their medical history. They found that when the interviewer was absent, the reporting of serious diseases was less likely. In a review of relevant research, Rosenthal (2002) suggested that the presence of an interviewer may inadvertently introduce cues to the witnesses to report more central rather than peripheral details, though the interviewer was not instructed to do so. This might occur because the interviewer's presence may increase the witness's overall motivation to perform. Additionally, Sauerland et al. (2014) found that conditions in which the interviewer was absent while writing led to better recall performance. Therefore, Experiment 2 had participants complete this study via Zoom, with both the interviewer and participant having their cameras turned off.

Method

Participants

A total of 84 introductory psychology students (32 males, 52 females; M_{Age} = 20.04 years, SD_{Age} = 4.54) from the University of Oklahoma participated in this study in exchange for partial course credit. All students were enrolled in an introductory psychology course and were recruited via a university recruitment portal (SONA study flier). The flier informed potential participants that they would watch a video and be asked various questions about the video at three different timepoints. Participants received a maximum of 2.5 research credits for their psychology course. They received credit following the completion of two Zoom sessions and one email response. To participate, students must have been at least 18 years of age and able to provide consent or receive parental consent if they were younger than 18 years of age. In addition, participants must have considered themselves proficient in English.

Participants were randomly assigned to one of four conditions: Control-Written (n = 19), Control-Spoken (n = 20), Written (n = 25), and Spoken (n = 20). All participants' data were kept anonymous and separate from all possible identifying information. No significant risks were encountered by the participants, and they were treated in accordance with APA ethical standards. This study was approved by the University of Oklahoma IRB.

Materials

The materials used in this experiment were identical to those used in Experiment 1.

Design and Procedure

The design of Experiment 2 is the same as Experiment 1. A 4 (Interview Modality: Control-Written, Control-Spoken, Written, Spoken) x 3 (Interview Timing: Immediate, 1-week delay, 1-month delay) mixed design with interview modality as the between-subjects factor and interview timing as the within-subjects factor.

Experiment 2's procedure is identical to Experiment 1 apart from the sessions occurring via Zoom. Also, before beginning the sessions, participants were instructed to turn their computer cameras off. The researcher also kept their computer camera off for the duration of the session. The researcher shared their screen to show the participant the video. Also, instead of recording participants' interviews with an audio recorder as in Experiment 1, the Zoom meeting was recorded and uploaded to MyMedia for transcription, after which a research assistant edited the transcription and fixed any transcription errors. Only participants who were in the Spoken conditions (Control-Spoken and Spoken) had their interviews recorded and only the interview was recorded; the parts of the session that occurred before the interview were not recorded. Participants in the written conditions (Control-Written and Written) typed their responses in a Word document instead of writing on lined sheets of paper. These participants emailed their responses immediately to the researcher when the session ended. The responses were de-identified and saved.

Results

A total of 84 participants completed Test 1. Of those, 75 (89.3% return rate) participants completed Test 2, and 44 (52.4% response rate) participants completed the email response. Only participants who completed Tests 1 and 2 were included in the subsequent analyses. Nine participants' data were removed before analyses due to incomplete participation. Therefore, 75 participants' data were used for the Test 1 and Test 2 data analyses and 44 (58.7% response rate) participant email responses were used for the Test 3 analyses.

All audio-recordings were transcribed before coding. Interviews were coded by two different individuals. Open-ended responses were coded for the number of central and peripheral details reported (based on the same pilot study discussed in Experiment 1). The interviews were also coded for the number of intrusions and any other correctly reported details that were not considered either central or peripheral to the video. An intrusion was anything reported that did not actually occur in the video.

The time that the interviews took to complete also was recorded for the first and second interviews. Interview time was not recorded for the third interview as that occurred via email like in Experiment 1. Additionally, the frequency of words either written or spoken for all three interviews was reported.

Test 1

The Test 1 data were checked for outliers and no outliers were present. Tests of normality indicated that the number of central details reported and the number of intrusions reported were normally distributed, however the number of peripheral details reported and extra correct details reported were not. These data were log transformed before conducting any analyses (Shapiro-Wilk normality test, $W = .943$, $df = 39$, $p = .05$ and $W = .968$, $df = 39$, $p = .335$, respectively). Additionally, the number of words either spoken or written during the Test 1 interview and the time in seconds of those interviews were not normally distributed. Log transformations fixed the non-normalities for the word counts ($W = .978$, $df = 39$, $p = .645$) and improved the non-normalities for interview time. Again, participants in the two control conditions (Control-Written and Control-Spoken) were not tested during the Test 1 phase of the study.

The total number of central, peripheral, and correct other details reported during Test 1 were aggregated to reflect the overall correct number of details reported. A one-way ANOVA compared the total correct number of details reported in the Spoken and Written conditions; $F(1, 37) = 2.761$, $p = .105$, $\eta^2 = .069$. Participants in the Written condition ($M = 31.84$, $SD = 8.83$) did not differ significantly in the number of correct details reported compared to the Spoken condition ($M = 26.70$, $SD = 10.39$). These findings are not consistent with Experiment 1, although the direction of the effect and the magnitude of the mean difference are similar.

A MANOVA was conducted to compare the number of central details, peripheral details, correct other details, and intrusions reported separately in the Spoken and Written conditions. This analysis examines the immediate effect that interview modality has on memory report accuracy. There is no significant difference in the number of central details reported between the two conditions: $F(1, 37) = .683$, $p = .414$, contrary to Experiment 1. This finding is interesting given that central details should be considered more important to recall compared to secondary details. However, there is a marginally significant difference in the number of peripheral details reported as a function of interview modality; $F(1, 37) = 3.892$, $p = .056$, $\eta^2 = .095$; participants in the Written condition ($M = 5.53$, $SD = 2.74$) reported more peripheral details compared to the Spoken condition ($M = 4.10$, $SD = 2.79$). These findings are consistent with Experiment 1. There is also a significant difference in the number of details that were reported correctly but not deemed central or peripheral: $F(1, 37) = 4.583$, $p = .039$, $\eta^2 = .110$. Again, participants in the Written condition ($M = 14.05$, $SD = 4.33$) reported more correct other details compared to the Spoken condition ($M = 11.35$, $SD = 4.55$), consistent with Experiment 1. Additionally, the average number of intrusions reported was not significantly different between the Spoken and Written conditions: $F(1, 37) = .305$, $p = .584$, replicating Experiment 1. These findings support the hypothesis that participants in the Written condition perform better than participants in the Spoken condition. More specifically, participants in the Written condition reported more peripheral and extra correct details during Test 1 compared to participants in the Spoken condition, but did not report more intrusions. This suggests that during the Test 1 phase, even though participants in the Written condition report more correct types of information, they are not more likely to report less incorrect information as well. More specifically, there is not a liberal criterion shift of reporting details at work here. Participants in the Written condition are not reporting more of each detail type because if they were, these participants would have also reported more intrusions. These findings can be seen in Figure 6.

Figure 6

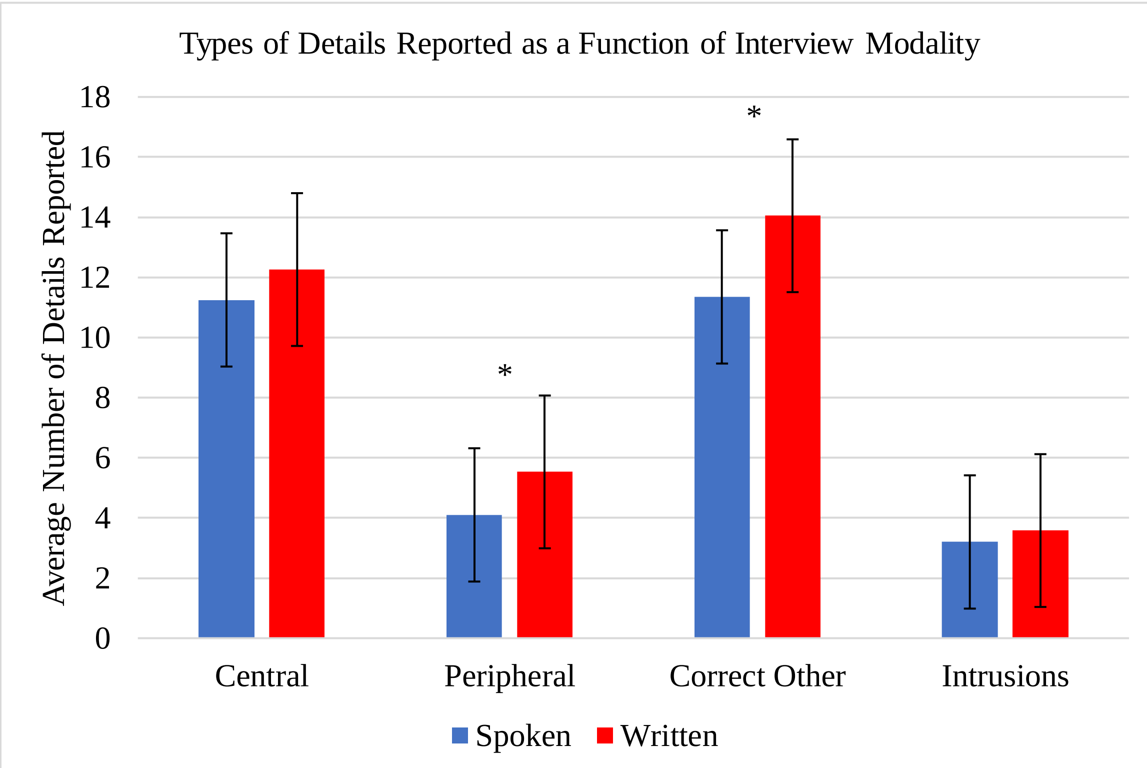
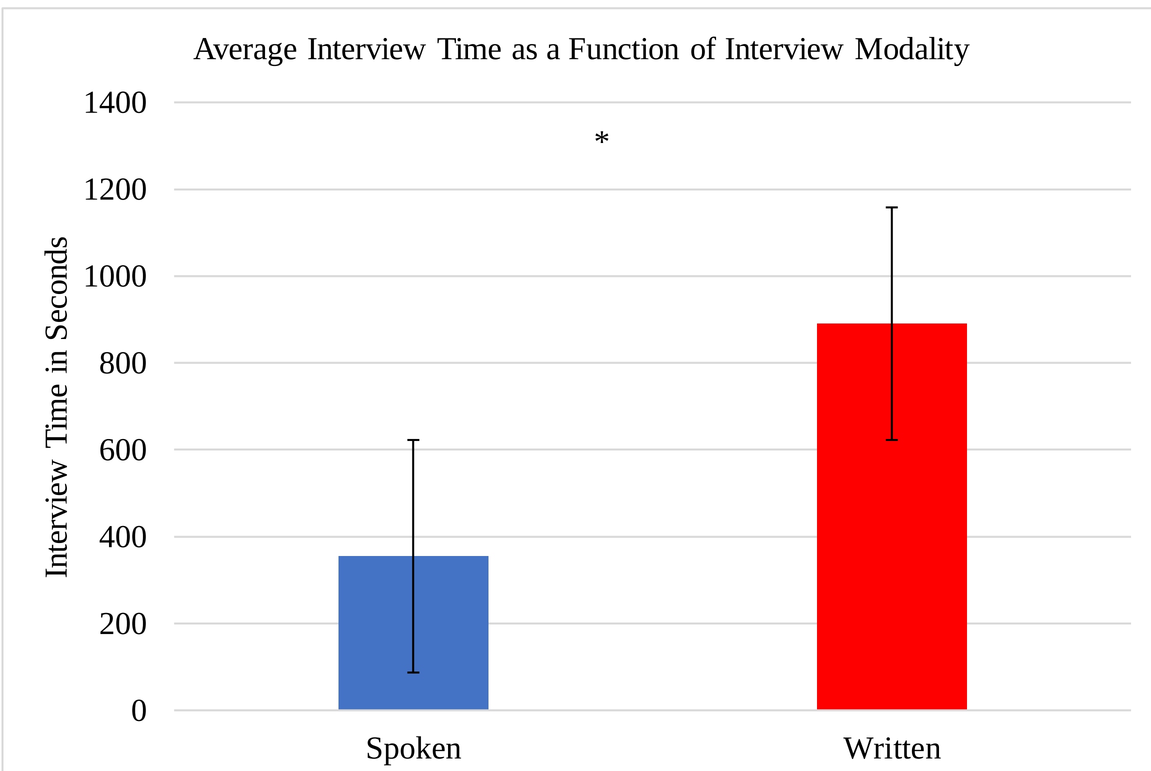
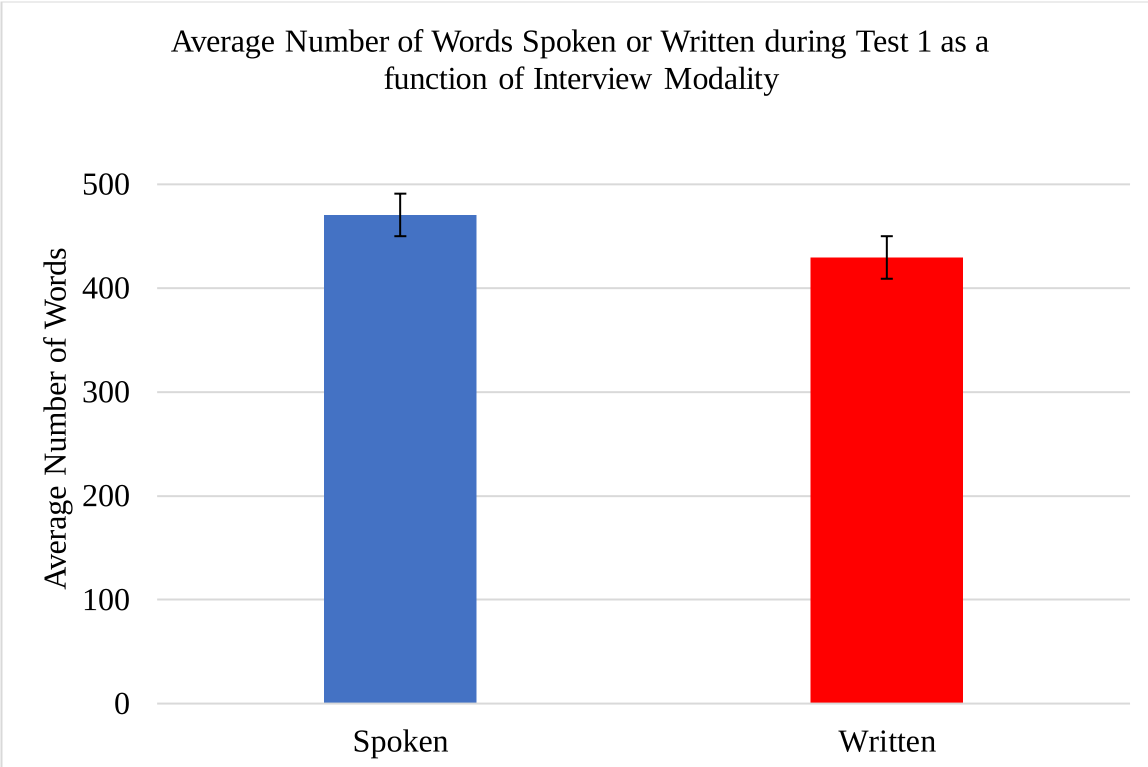


Figure 6. The max number of details that could have been reported as either central or peripheral are 21 and 18 details, respectively.

The next set of analyses were conducted because it is possible that participants in the Written condition had better memory reports for extra correct and peripheral details because they worked more productively and proficiently than participants in the Spoken condition. More specifically, the interview completion times for each condition were examined and there is a significant difference in the average number of seconds that the interviews took to complete: $F(1, 37) = 68.267$, $p < .001$, $\eta^2 = .649$. Not surprisingly, participants in the Spoken condition (n = 20) took significantly less time ($M = 354.80$, $SD = 125.29$) to complete the interview compared to the participants in the Written (n = 19) condition (M = 890.37, SD = 328.97) (see top panel of Figure 7). However, the average number of words spoken or written during the interviews were compared and there is no significant difference in the overall word counts between the two conditions: $F(1, 37) = .190$, $p = .665$. Thus, like Experiment 1, approximately the same amount of information is being reported by the participants regardless of interview modality. This suggests that participants are working equally hard in both conditions, but participants in the Written condition are working more efficiently (see bottom panel of Figure 7).

Figure 7





Test 2

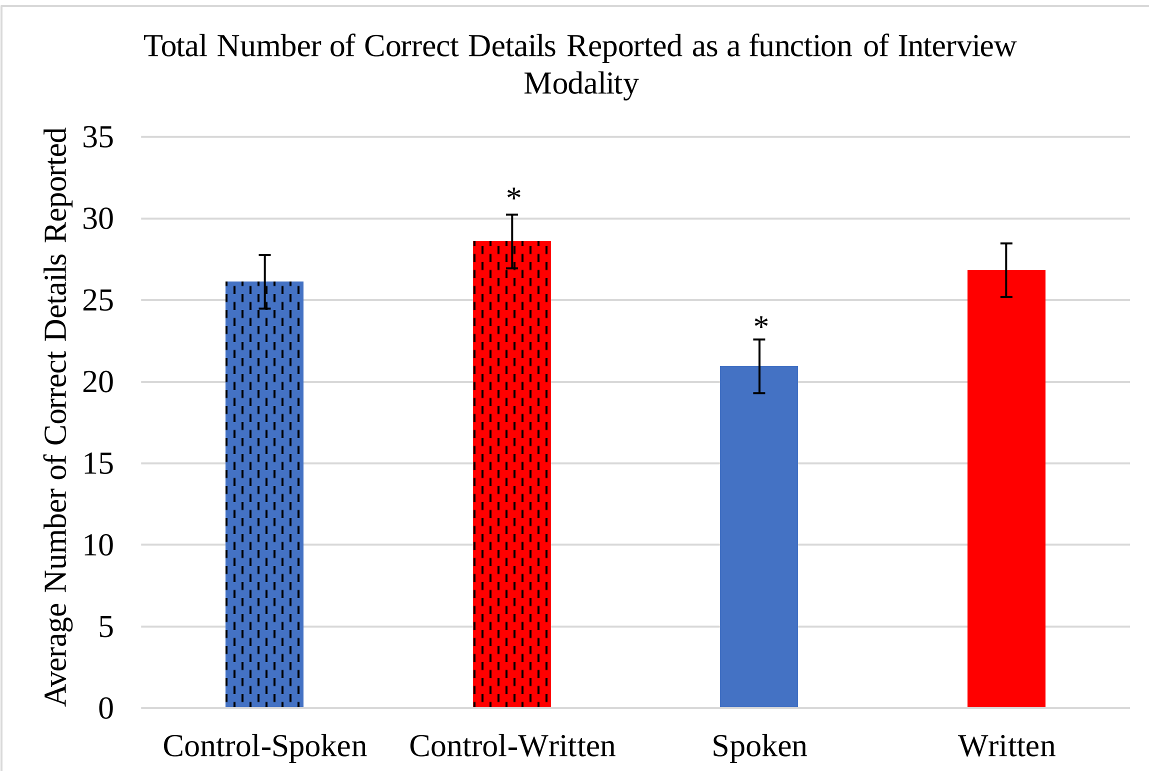
The Test 2 data were checked for outliers and transformed when necessary. The number of words spoken or written during the interviews, the time to complete the interview, along with the number of peripheral details, and intrusions, needed to be transformed. Additionally, the proportion of true interview questions answered correctly, the proportion that participants incorrectly answered a false question, and the proportion participants correctly rejected a false question, also were transformed. Following all transformations, five outlying data points were removed before conducting any analyses.

The total number of central, peripheral, and correct other details reported during Test 2 were aggregated to reflect the overall correct number of details reported for each participant. A one-way ANOVA was conducted to compare the total correct number of details reported as a function of interview modality; $F(3, 68) = 2.675$, $p = .054$, $\eta^2 = .106$. Participants in the Control-

Written condition (M = 28.59, SD = 6.87) reported (p = .057) more correct details compared to participants in the Spoken condition (M = 20.94, SD = 7.82). These findings are consistent with our hypothesis that writing improves memory reports compared to speaking (see top panel of Figure 8) even at a one-week delay. Additionally, writing, even after not being tested immediately, improves memory reports more so than when participants were initially tested by speaking.

Multiple one-way ANOVAs were conducted on the disaggregated Test 2 data to examine the effects that interview modality had on the testing effect. Multiple one-way ANOVAs were conducted instead of repeated-measure ANOVAs because the two control conditions were not tested during the Test 1 phase and the low Test 3 return rate. There is a significant difference in the number of peripheral details reported as a function of interview modality; $F(3, 69) = 4.099$, $p = .010$, $\eta^2 = .151$. Participants in the Control-Written condition (M = 5.41, SD = 2.00) reported more peripheral details compared to participants in the Spoken condition (M = 2.75, SD = 2.20). Note that in Experiment 1 it was the Written condition, not the Control-Written condition, that reported more peripheral details compared to the Spoken condition. Regardless, in this instance, the data still signal that writing overrides a possible boost from testing versus speaking because the Control-Written participants were not tested immediately but the Spoken condition participants were. This suggests that writing leads to better memory reports compared to speaking and that this enhancement remains one week following exposure to the event. There is no significant difference of interview modality on the number of central details reported ($F(3, 71) = 1.385$, $p = .254$) or the number of intrusions reported ($F(3, 68) = .370$, $p = .775$), which is consistent with the findings from Experiment 1. There is not a significant difference between interview modality on the number of correct other details reported; $F(3, 68) = 1.336$, $p = .270$, which is contrary to Experiment 1 (see bottom panel of Figure 8). Given that there is not a significant difference in the number of central details reported or the number of correct other details reported, it is possible that it is easier to report additional peripheral details as a function of increasing retention interval.

Figure 8



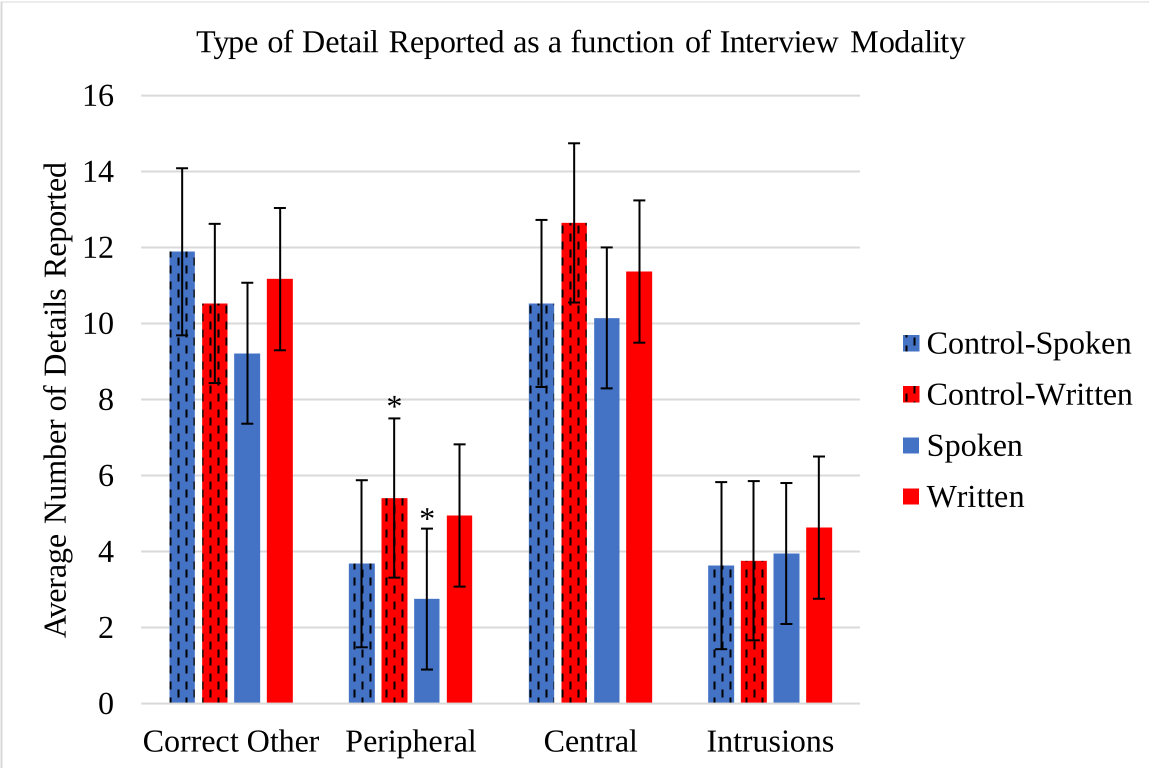


Figure 8. The max number of details that could have been reported as either central or peripheral are 21 and 18 details, respectively.

When examining the impact of interview modality on the testing effect, multiple one-way ANOVAs were conducted on the proportions recorded for the pointed questions. There is not a significant difference of interview modality for the proportion of true questions answered (F(3, 71) = 1.472, p = .230) or the proportion of false questions endorsed (F(3, 71) = 1.418, p = .245), which is consistent with the findings of Experiment 1. However, contrary to Experiment 1, there is no significant difference in the proportion of false questions correctly rejected as a function of interview modality (F(3, 71) = .884, p = .454). This suggests that participants in the Control-Written condition do not have a greater correct rejection rate of false information compared to the Spoken condition, and that interview modality does not differentially impact participants' ability to correctly identify false questions that did not occur in the video.

A series of one-way ANOVAs examined if the greater efficiency of the Written condition observed in Test 1 carried over to the Test 2 phase. The interview times were first converted to seconds before analysis. There is a significant difference in the completion times of the Test 2 interviews as a function of interview modality: $F(3, 71) = 34.802$, $p < .001$, $\eta^2 = .595$; interview completion times in the Control-Spoken condition ($M = 602.89$, $SD = 188.78$) were significantly lower compared to participants in both the Control-Written ($M = 1203.24$, $SD = 373.99$) and Written ($M = 1083.26$, $SD = 447.38$) conditions. Additionally, participants in the Control-Written condition took significantly more time to complete their interviews compared to participants in the Spoken ($M = 506.80$, $SD = 166.93$) condition. Participants in the Written condition also took significantly more time to complete the interview compared to participants in the Spoken condition. These findings are consistent with Experiment 1 and are consistent with writing being slower than speaking.

The number of words reported during the Test 2 interview were divided between the number of words reported during the open-ended portion of questioning and the pointed questioning portion of the interview. There is a significant difference in the number of words reported during the open-ended portion of questioning; $F(3, 71) = 2.780$, $p = .047$, $\eta^2 = .105$. More specifically, participants in the Control-Written condition ($M = 366.82$, $SD = 121.36$) had marginally ($p = .091$) lower word counts compared to the Control-Spoken ($M = 576.26$, $SD = 226.59$) condition. Additionally, participants in the Written condition ($M = 406.37$, $SD = 239.65$) had marginally ($p = .091$) lower word counts compared to participants in the Control-Spoken condition. There is also a significant difference in the number of words reported during the pointed portion of questioning; $F(3, 69) = 4.613$, $p = .005$, $\eta^2 = .167$; participants in the Spoken condition ($M = 114.45$, $SD = 55.09$) had significantly larger word counts compared to participants in both the Control-Written (*M* = 74.82, *SD* = 22.95) and Written (*M* = 72.32, *SD* = 25.76) conditions. These findings are consistent with Experiment 1 and are particularly interesting given the written superiority effect for peripheral details, despite the lower word counts, even one week after initial encoding. Participants who write perform more effectively than participants who speak.

Test 3

The Test 3 data were checked for outliers and transformed when necessary. The number of words reported during the interviews, along with the number of peripheral details and intrusions reported needed to be transformed. Additionally, the proportion of true interview questions answered correctly, the proportion that participants incorrectly answered a false question, and the proportion participants correctly rejected a false question were transformed. The time to complete the interview was not recorded for Test 3 because this test occurred via email. Following all transformations, one outlying data point was removed before conducting any analyses.

The total number of central, peripheral, and correct other details reported during Test 3 were aggregated to reflect the overall correct number of details reported for each participant. A one-way ANOVA was conducted to compare the total correct number of details reported as a function of interview modality; $F(3, 39) = 2.181$, $p = .106$. There is not a significant difference in the total number of correct details reported across interview modality, which is consistent with Experiment 1. These findings suggest that, in general, any written superiority effect seen immediately, or after a one-week delay, does not hold up following a one-month retention interval.

Multiple one-way ANOVAs were conducted on the Test 3 data to examine effects that interview modality had on the testing effect one month following encoding. There is not a significant difference between interview modalities on the number of central details reported ($F(3, 39) = 2.338$, $p = .089$), intrusions reported ($F(3, 37) = .410$, $p = .747$), or correct other details reported ($F(3, 39) = 1.271$, $p = .298$). These findings are consistent with what was found in Experiment 1. There is not a significant difference between interview modalities on the number of peripheral details reported; $F(3, 33) = .447$, $p = .721$. This is not consistent with Experiment 1 where participants in the Written condition reported marginally more peripheral details than in the Spoken condition. These findings suggest that interview modality does not have a differential impact on memory reports following a one-month delay.

When examining the impact of interview modality on the testing effect, multiple one-way ANOVAs were conducted on the proportions recorded for the pointed questions. There is not a significant difference between interview modalities on the proportion of true questions answered ($F(3, 39) = 1.316$, $p = .283$), the proportion of false questions endorsed ($F(3, 39) = 1.161$, $p = .337$), or the proportion of false questions correctly rejected ($F(3, 39) = .428$, $p = .734$). These findings are consistent with Experiment 1, and provide additional support that interview modality does not impact memory reports differentially over extended delay periods.

The number of words reported in each interview was also recorded. The word counts were divided between the number of words reported for the open-ended portion of questioning and the pointed portion of questioning. There is no significant difference in the number of words reported in the open-ended portion of the question ($F(3, 39) = 1.956$, $p = .137$) or the pointed portion of the interview ($F(3, 38) = .176$, $p = .912$), which is consistent with the findings of Experiment 1.

Discussion

Having individuals provide either a written or spoken memory report differentially impacts both the type and number of details reported. The present findings are consistent with research supporting a written superiority effect (Kraus et al., 2017; Sauerland et al., 2014). More specifically, participants who write usually provide more correct information compared to their counterparts. This superiority effect remains one-week following encoding and, marginally so, one-month later (Kraus et al., 2017). The present study partially supports the findings from Sauerland and Sporer (2011); participants who wrote during the Test 1 phase of both experiments reported more peripheral details compared to participants who spoke. Sauerland and Sporer (2011) posited that speaking may be more productive, but not necessarily as efficient; the present study provides support for this idea based on the longer interview times for those participants who wrote but the equivalent word counts compared to those who spoke. These findings, combined with the claim that participants who produce better memory reports, signals that writing is more effective than speaking. Additionally, the present study suggests that the written superiority effect largely diminishes by Test 3. These findings may be indicative of a shift from less verbatim recall to gist recall. This is consistent with the fuzzy-trace theory literature which suggests that, detailed (or verbatim) memories are forgotten more quickly than gist memories (Ahmad, Moscovitch, & Hockley, 2017).

Experiments 1 and 2 both suggest that writing leads to better quality memory reports compared to speaking, although some of the effects are only marginally significant in Experiment 2, likely due to the smaller sample size. Conversely, a strength of Experiment 2 is that it removed some of the social components inherent to a traditional interview. Previous research suggests that the presence of an interviewer can have both positive and negative effects on a witness (Bergman et al., 2004). Therefore, by requiring both the participant and researcher to keep their computer cameras off for the duration of the sessions, Experiment 2 may have allowed the participant to feel more comfortable reporting the details from the video. Additionally, given that there seems to be no major differences in memory reports when administering the interview in-person or via Zoom, this suggests that at a minimum, conducting remote eyewitness interviews is not detrimental to memory report quality.

For less serious crimes, interviewing a witness via Zoom from their home may reduce the resources precincts have to put forth when questioning witnesses. Additionally, allowing a witness to discuss the details of an event from their homes may improve overall memory reports. Though the present study did not show a Zoom recall advantage, other previous research does suggest that allowing individuals to be interviewed remotely leads to an increase in the accuracy of overall memory reports and a reduction in error reporting (Nash, Houston, Ryan, Woodger, & Nash, 2014; Taylor & Dando, 2018). It is important to note that the study conducted by Nash, Houston, Ryan, Woodger and Nash (2014), required participants to watch a crime film before being interviewed 1 day later (remotely or face-to-face) or 1 to 2 weeks later (face-to-face). It is possible that the present study did not show a Zoom recall advantage because the video is likely not memorable or relevant to the participants. Therefore, being interviewed remotely may not function in a similar way as to reduce the social factors thought to impact memory reports. This is plausible because the social factors here are less likely to be detrimental to a participant's memory given the neutrality of the video. Contrarily, video-mediated interviews have also been shown to be harmful to the quality of memory reports because it circumvents crucial rapport-building opportunities between the interviewer and eyewitness which may make the witness uncomfortable (Nash et al., 2014). Thus, taking into consideration which interviewing technique maintains the highest quality of memory reports is crucial. Given that conducting remote interviews seems to have differential effects on memory report qualities, research should continue to investigate the conditions in which remote interviewing is beneficial.

Writing as a means of testing one's memory seems to provide an additional benefit. This was supported by not finding a within modality (Control-Written to Written) testing effect, but rather finding a cross-modality (Control-Spoken to Written) testing effect. In other words, this suggests that writing is more beneficial than speaking because it mitigates the anticipated testing effect. For performance on Test 2, watching the video twice and then providing a written report

(Control-Written) appears to be more beneficial than watching the video twice and then providing a spoken report (Control-Spoken). At Test 2, this testing benefit is only seen between the Control-Spoken and Written conditions, which is consistent with Roediger and Karpicke (2006).

Limitations

A possible limitation of this study is that even though the research assistants were trained to be systematic in their responses to each question, it is possible that participants who spoke were more likely to look at and/or engage with the research assistant in-between questions compared to participants who wrote because participants who wrote are more likely to maintain their focus on the sheet of paper. When speaking, it is more natural to engage with the other individual, so, the speaking conditions inherently may have induced more researcher interactions than the written conditions. Thus, it is possible that the present findings could have been impacted by an increase in research assistant interactions. It is possible that, even with training, the research assistants may have inadvertently cued the participants that their responses were correct or incorrect, making it possible that accidental confirmatory feedback played a role in the present study's findings (Zaragoza et al., 2001). Research assistants that had never watched the video would have alleviated this possible concern.

Another limitation of the current study is that the video depicted events that occurred at an all-boys summer camp, and the events portrayed are largely emotionally neutral. Thus, the present findings may not necessarily generalize to more realistic scenarios. Research suggests that stress can negatively impact memory (Christianson, 1992), and given that witnessed events are likely to be stress inducing or emotionally charged, it is possible that the written superiority effect seen in Test 1 and Test 2, may not hold with more life-like stressful witnessed events. Again, it is possible that Zoom associated recall advantages may be contingent on emotional events given that the removal of social factors in this instance is not likely to provide a benefit to memory reports because when recalling a neutral or non-relevant event, individuals may not find it as important to report all critical pieces of information. Event relevance may induce greater levels of engagement from a witness because they may feel that it is important to contribute adequately to the interview.

An additional concern is the attrition rates of both experiments from Test 2 to Test 3 and whether this drop off in participation between tests is largely random or potentially due to poorer memories. It is possible that participants who believed that they had poorer memories of the video were more likely to discontinue their participation. However, the attrition rates among the four conditions do seem to occur randomly, with no one condition having a greater attrition rate compared to another.

Future Directions

Future research is needed to elucidate how the presence of the interviewer impacts eyewitness's reports, and how that may interact with interview modality. Previous research argues that interviewer presence can differentially impact memory reports. Therefore, it is important to find ways to conduct interviews, without an interviewer being present, because it may allow for law enforcement agencies to more effectively delegate their limited resources. For example, if there is a case involving multiple witnesses, a precinct must work fast and efficiently to obtain the most detailed and accurate reports. Therefore, if precincts can interview witnesses remotely, this would reduce workload by providing more efficient ways to obtain eyewitness accounts.

Additionally, it is possible that the one-month retention interval is too large for a non-memorable event. Hence, utilizing a more memorable video may be helpful to parse out if the written superiority effect extends to one month (or greater) following initial encoding. That was not the case in the present experiments, but the video used likely did not serve as a relevant or memorable event to participants. Determining if event relevance and memorability impacts memory reports as a function of interview modality is an important next step, especially given that witnesses sometimes experience long delays before being interviewed.

Though the interviews conducted in this study were semi-structured, future research should consider investigating the written superiority effect as a function of different written interview methods. Past research suggests that varied written interview structures can differentially influence the occurrence of a written superiority effect (Kraus et al., 2017). Given that the SAI is suitable for real-life scenarios, and that witnesses can experience long delays before being interviewed, parsing out the differential effects of structured written interviews for more applied scenarios is crucial. Furthermore, investigating the proficiency of implementing the SAI remotely will provide another avenue for realistically reducing police precinct resource demand. Additionally, though the SAI seemly works in applied settings, investigating how free-recall and semi-structured interviews impact the presence of a written superiority effect as well as the robustness of this effect. This is a crucial next step considering the importance of self-monitoring in inducing a written superiority effect. Writing is thought to be better than speaking because it helps individuals monitor their own reports. However, it is possible that writing may lead to better performance when a less structured interview is used because writing should allow for better self-monitoring which in turn may help to impose a structure to an already less-structured (free recall) interview.

There are a multitude of ways in which investigators can conduct interviews. Accordingly, it is of key interest to identify practices that will help investigators obtain the most accurate memory reports. In addition to the type of crime witnessed, additional factors should be taken into consideration when working to elucidate the best interview practices. For example, vantage point of the crime (Christianson, 1992), attention directed towards the crime (Hyman, Wulff, & Thomas, 2018), and the degree of structured interviews (Kraus et al., 2017), all are potential factors that should be taken into consideration when determining the most appropriate ways to interview witnesses. This study has allowed us to take an important step in identifying some factors, specifically interview modality and interview timing, which are critical to eliciting the highest quality of memory reports.

References

Ahmad, F. N., Moscovitch, M., & Hockley, W. E. (2017). Effects of varying presentation time on long-term recognition memory for scenes: Verbatim and gist representations. *Memory and Cognition*, *45*(3), 390–403. https://doi.org/10.3758/s13421-016-0672-1

Bergmann, M. M., Jacobs, E. J., Hoffmann, K., & Boeing, H. (2004). Agreement of self-reported medical history: Comparison of an in-person interview with a self-administered questionnaire. *European Journal of Epidemiology*, *19*(5), 411–416. https://doi.org/10.1023/B:EJEP.0000027350.85974.47

Bouwmeester, S., & Verkoeijen, P. P. J. L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language*, *65*(1), 32–41. https://doi.org/10.1016/j.jml.2011.02.005

Brainerd, C. J. & Reyna, V. F. (2005). *The science of false memory*. New York: Oxford University Press.

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4–5), 514–527. https://doi.org/10.1080/09541440701326097

Christianson, S. Å. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, *112*(2), 284–309. https://doi.org/10.1037/0033-2909.112.2.284

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and Testing Effects: A Deeply Critical, Lengthy, and At Times Discursive Review of the Literature. In *Psychology of Learning and Motivation - Advances in Research and Theory* (1st ed., Vol. 53). https://doi.org/10.1016/S0079-7421(10)53003-2

Geiselman, R. E. (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. *Journal of Police Science & Administration.*

Grant, T. & Sullivan, K. (1989). *Looking for Miracles*.

Hyman, I. E., Wulff, A. N., & Thomas, A. K. (2018). Crime Blindness: How Selective Attention and Inattentional Blindness Can Disrupt Eyewitness Awareness and Memory. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 202–208. https://doi.org/10.1177/2372732218786749

Kellogg, R. T. (2007). Are written and spoken recall of text equivalent? *American Journal of Psychology*, *120*(3), 415–428. https://doi.org/10.2307/20445412

Kraus, U., Zeier, F., Wagner, W., Paelecke, M., & Hewig, J. S. (2017). Comparing the quality of memory reports in different initial eyewitness questioning approaches. *Cogent Psychology*, *4*(1). https://doi.org/10.1080/23311908.2017.1403063

McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory and Cognition*, *34*(2), 261–267.

Moscovitch, M. (1994). Cognitive Resources and Dual-Task Interference Effects at Retrieval in Normal People: The Role of the Frontal Lobes and Medial Temporal Cortex. *Neuropsychology*, *8*(4), 524–534. https://doi.org/10.1037/0894-4105.8.4.524

Nash, R. A., Houston, K. A., Ryan, K., Woodger, N., & Nash, R. A. (2014). Remembering remotely: would video-mediation impair witnesses' memory reports? *Psychology, Crime & Law*, *20*(8), 756–768.

Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing. *Memory and Cognition*, *39*(1), 155–170. https://doi.org/10.3758/s13421-010-0005-8

Reyna, V. F. & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1–75. https://doi.org/10.1016/1041-6080(95)90031-4

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

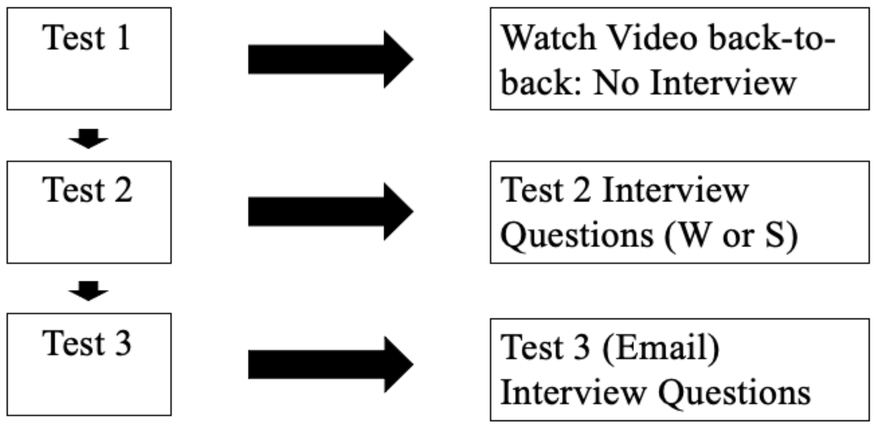
Roediger, H. L., & McDermott, K. B. (1995). Creating False Memories: Remembering Words

Not Presented in Lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803
Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist*, *57*(11), 839–849.
Sauerland, M. Sporer, S. L. (2011). Written vs. Spoken Eyewitness Accounts: Does Modality of Testing Matter? *Behavioral Sciences & the Law*, *29*, 846–857. https://doi.org/10.1002/bsl.1013
Sauerland, M., Krix, A. C., van Kan, N., Glunz, S., & Sak, A. (2014). Speaking is silver, writing is golden? The role of cognitive and social factors in written versus spoken witness accounts. *Memory and Cognition*, *42*(6), 978–992. https://doi.org/10.3758/s13421-014-0401-6
Taylor, D. A., & Dando, C. J. (2018). Eyewitness memory in face-to-face and immersive avatar-to-avatar contexts. *Frontiers in Psychology*, *9*(APR), 1–11. https://doi.org/10.3389/fpsyg.2018.00507
Warren, A. Lane, P. (1995). Effects of timing and type of questioning on eyewitness accuracy and suggestibility. *Memory and Testimony in the Child Witness*, 44–60.
Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571–580. https://doi.org/10.1080/09658210244000414
Zaragoza, M. S., Payment, K. E., Ackil, J. K., Drivdahl, S. B., & Beck, M. (2001). Interviewing Witnesses: Forced Confabulation and Confirmatory Feedback Increase False Memories. *Psychological Science*, *12*(6), 473–477. https://doi.org/10.1111/1467-9280.00388
Zoladz, P. R., Cadle, C. E., Dailey, A. M., Fiely, M. K., Peters, D. M., Nagle, H. E., … Payment, K. E. (2017). Blunted cortisol response to acute pre-learning stress prevents misinformation effect in a forced confabulation paradigm. *Hormones and Behavior*, *93*, 1–8. https://doi.org/10.1016/j.yhbeh.2017.04.003

Appendix

Figure 9

Control Conditions: Control Written (W) and Control Spoken (S)



Experimental Conditions: Written (W) and Spoken (S)