

Testing Knowledge to Battle Misinformation: Quizzes as a Tool for Fact Checkers

Jessica Collier¹, Lisa Fazio², & Raunak Pillai²

Working paper. This version: August 2021

Fact-checking organizations have proliferated in the last decade as the opportunity for exposure to false or misleading information has increased. With the abundance of available information online—both true and false—fact-checkers face two difficult goals. First, they must stand out amongst this sea of online content by getting readers to see and engage with their fact checks. Second, fact-checkers aim to ensure that individuals remember and are able to recall accurate information from their fact checks. This project tests one potential method for helping fact-checking organizations achieve each of these goals. Digital journalism research has indicated that quizzes can be a useful tool to improve engagement with online news. Additionally, research on cognition finds that quizzing people can improve their memory of the information, even weeks later. To test people’s memories of fact-checked content, we conducted a series of pre-registered online experiments through Amazon’s Cloud Research platform. In Study 1, we examined whether delivering a quiz before or after reading a Snopes article improved (1) how accurate people believed the debunked claim to be, and (2) their memory for specific information within the fact check. Both articles chosen for the study were about health-related topics. We tested 450 participants immediately after reading the fact check and 450 additional participants one week after reading the fact check. Our findings indicate that even when asked specific and complex health questions such as “What is a vanished twin?” or “What is a microchimerism?,” people were significantly more likely to recall that information when they were quizzed before or after reading the fact check compared to readers who did not receive

¹ Center for Media Engagement, University of Texas at Austin

² Vanderbilt University

a quiz while reading. While individuals were able to recall this detailed information, the quizzes had no effect on readers' beliefs in the false claim that was the subject of the fact check. Overall, individuals who had seen the Snopes article tended to say the false claim was slightly inaccurate. In Study 2, we replicate this experiment and quiz participants on questions that are more directly related to the false claim itself. Using two different health-related Snopes articles, we come to similar conclusions. Participants were able to recall detailed information from the fact check even one week later, but there was no effect of quizzes on readers' beliefs in the false claim. We are currently conducting a third study to tease apart these findings further.

Method

This study explores how including a quiz before or after reading a fact check influences individuals' accuracy rating of the false claim and their memory for the information in the article. To address our hypotheses, we conducted an online survey experiment administered through Qualtrics. We used a 3 (quiz condition: quiz before, quiz after, no quiz) X 2 (article condition: DNA, marijuana) X 2 (delay: no delay, one-week delay) between-subjects design. This project received Institutional Review Board approval on December 3, 2019. Data were collected from February 4-19, 2020.

Participants

Participants were recruited using Amazon's CloudResearch (formerly TurkPrime) platform ($N = 900$). U.S. residents age 18 years or older were eligible to participate. Concerns about crowdsourcing platforms for participants have shown these samples to be more representative than convenience samples (Berinsky, Huber, & Lenz, 2012). Additionally, responses on TurkPrime are shown to be of higher quality than responses from Mechanical Turk (Litman, Robinson, & Abbercock, 2017). Participants were compensated \$1.45 for their

participation. Participants who were assigned to receive the delayed post-treatment questionnaire but did not respond were excluded from the data set. Thus, we arrived at a final sample size of $N = 837$. The demographic make-up of the sample was primarily white (73.3%) and male (52.5%). Average age was 38.33 years old ($SD = 11.42$).

Procedure

Participants were randomly assigned to one of twelve conditions in which they were shown a fact check from Snopes. In these conditions, participants received a brief, two-question quiz about the information in the fact check either before reading the article, after reading the article, or they received no quiz at all. The topic of the article was randomly assigned such that participants read one of two articles (DNA, marijuana) about a piece of health misinformation. After reading the article and answering any quiz questions, half of participants were asked to immediately rate the accuracy of the false claim from the article and answer four recall questions about the content of the fact check. The other half of participants were asked to rate the accuracy of the claim and answer the four recall questions one week later. Participants then completed a basic demographic questionnaire and were compensated for their participation.

Measures

Independent variables

Quiz condition. The primary independent variable in this study is the quiz condition, or whether participants received no quiz versus receiving a quiz prior to or after reading a fact check. The quizzes consisted of two questions about content within the fact check article (see Appendix for full wording). After answering each of the questions, participants received a notification with whether the answer option they chose was correct or incorrect.

Article condition. To avoid possible effects due to topic differences, we tested the efficacy of the quiz intervention for two different articles pertaining to a piece of health misinformation. The first was an article about DNA titled, “Do Women Retain DNA from Every Man They Have Slept With?” and the second was an article about marijuana titled, “Did a New Study Show that Marijuana Leads to a Complete Remission of Crohn’s Disease?” Both articles were fact checks from Snopes that had been rated entirely false.

Test condition. The time of testing varied to ascertain whether the effects of quizzes were immediate and/or would hold up after a one-week delay. To test this, half of participants received the post-treatment measures immediately after reading the fact check. The other half of participants received the post-treatment measures one week after reading the fact check.

Dependent variables

Accuracy. Participants were asked to rate the accuracy of the false claim that was the subject of the fact check to which they were assigned. For the DNA article, this statement read, “Women retain DNA from every man they have ever slept with,” and for the marijuana article, this statement read, “Marijuana leads to a complete remission of Crohn’s disease.” Responses were measured on an 11-point scale from *very inaccurate* (0) to *very accurate* (10).

Proportion of correct responses. Participants received a series of four open-end questions and were instructed to answer the questions in 1-2 sentences. Their responses to these questions were coded by two independent coders to identify whether the response was (1) correct, (2) partially correct, (3) incorrect, lure response from quiz, (4) other incorrect answer, or (5) don’t know. Krippendorff’s alpha for this measure was 0.89. A dummy variable of *correct* (1) and *incorrect* (0) was created for each of the questions. Correct responses were summed and divided by four to obtain a proportion of the correct responses provided by participants. This

proportion of correct responses was used as a measure of memory about the debunked information.

Results

First, we examine how quizzing affected accuracy ratings and memory for key facts on those who received the post-treatment measures with no delay. We ran a 3 (quiz: none, before, after) x 2 (article: DNA, marijuana) ANOVA on participants' accuracy rating. Results indicate a marginally significant effect of the article on accuracy such that those in the marijuana condition ($M = 3.30, SD = 3.07$) rated the false claim as more accurate than those in the DNA condition ($M = 2.82, SD = 3.23$), $F(1, 438) = 3.40, p = .07$. Results for the effects of quizzing were not significant. We ran a second 3 (quiz: none, before, after) x 2 (article: DNA, marijuana) ANOVA on participants' memory for the information in the fact check. Results here indicate a significant main effect of quizzes on the proportion of correct responses given by participants, $F(2, 438) = 12.68, p < .001$. As shown in Figure 1, those in the quiz before ($M = 0.51, SD = 0.32$) and quiz after ($M = 0.54, SD = 0.30$) conditions answered more questions correctly than those in the no quiz condition ($M = 0.39, SD = 0.29$). Results also indicate a significant main effect of article on the proportion of correct responses, $F(1, 438) = 104.70, p < .001$. Those in the marijuana article condition ($M = 0.61, SD = 0.30$) answered more questions correctly than those in the DNA article condition ($M = 0.34, SD = 0.25$). The interaction between quiz and article conditions was not significant.

[FIGURE 1 HERE]

Next, we examine whether there are similar effects on accuracy and memory for key facts for those in the delayed test condition. We again ran a 3 (quiz: none, before, after) x 2 (article: DNA, marijuana) ANOVA on participants' accuracy rating. Results indicate no significant

effects of quiz or article condition on accuracy for those in the delayed condition. We ran a second 3 (quiz: none, before, after) x 2 (article: DNA, marijuana) ANOVA on participants' memory for the information in the fact check. Results indicate a marginal effect of quizzing on proportion of correct responses, $F(2, 386) = 2.69, p = .07$. Those in the quiz before ($M = 0.27, SD = 0.24$) and quiz after ($M = 0.29, SD = 0.23$) conditions answered slightly more questions correctly than those in the no quiz condition ($M = 0.23, SD = 0.22$). There was also a main effect of article condition on the proportion of correct responses, $F(1, 386) = 35.84, p < .001$. Those in the marijuana article condition ($M = 0.33, SD = 0.25$) answered significantly more questions correctly than those in the DNA article condition ($M = 0.19, SD = 0.18$). The interaction between quiz and article conditions was not significant.

Finally, we examine differences between the no delay and delayed test conditions. We do this first with a 3 (quiz: none, before, after) x 2 (article: DNA, marijuana) x 2 (test: no delay, delayed) ANOVA on participants' accuracy rating. Results indicate a significant main effect of the article condition on accuracy, $F(1, 830) = 4.13, p < .05$. There are no other significant main effects or interactions.

Next, we ran a 3 (quiz: none, before, after) x 2 (article: DNA, marijuana) x 2 (test: no delay, delayed) ANOVA on participants' proportion of correct responses. As shown in Figure 2, there was a significant main effect of quizzing on proportion of correct responses, $F(2, 824) = 13.72, p < .001$. Again, those in the before ($M = 0.40, SD = 0.31$) and after ($M = 0.44, SD = 0.30$) quiz conditions answered a greater proportion of quiz questions correctly than those in the no quiz condition ($M = 0.31, SD = 0.27$). There was also a significant main effect of article on proportion of correct responses, $F(1, 824) = 132.80, p < .001$. Again, those in the marijuana article condition ($M = 0.49, SD = 0.32$) answered significantly more questions correctly than

those in the DNA article condition ($M = 0.27, SD = 0.23$). The main effect of the delay was also significant, $F(1, 824) = 151.49, p < .001$, where participants who answered questions immediately after reading the article ($M = 0.49, SD = 0.31$) answered more correctly than those who answered one week later ($M = 0.26, SD = 0.23$). Additionally, there was a significant interaction of article and delay, $F(1, 824) = 14.30, p < .001$ and a marginal interaction of quiz condition and delay, $F(2, 824) = 2.61, p = .08$.

Method

This study explores how including a quiz that is closely targeted to the false claims in a fact check influences individuals' accuracy rating of the false claim and their memory for the information in the article. To address our hypotheses, we conducted an online survey experiment administered through Qualtrics. We used a 2 (quiz condition: quiz, no quiz) X 2 (article condition: DNA, marijuana) X 2 (delay: no delay, one-week delay) between-subjects design. Data were collected from August 24-September 3, 2020.

Participants

Participants were recruited using Amazon's CloudResearch platform ($N = 609$). U.S. residents age 18 years or older were eligible to participate. Concerns about crowdsourcing platforms for participants have shown these samples to be more representative than convenience samples (Berinsky, Huber, & Lenz, 2012). Additionally, responses on TurkPrime are shown to be of higher quality than responses from Mechanical Turk (Litman, Robinson, & Abbercock, 2017). Participants were compensated \$1.45 for their participation. Participants who were assigned to receive the delayed post-treatment questionnaire but did not respond were excluded from the data set. Thus, we arrived at a final sample of $N = 530$. The demographic makeup of the sample was primarily white (81.1%) and male (51.1%). Average age was 41.16 years old ($SD = 12.38$).

Procedure

Participants were randomly assigned to one of eight conditions in which they were shown a fact check from Snopes. In these conditions, participants received a brief, two-question quiz about the information in the fact check after reading the article or they received no quiz at all. The topic of the article was randomly assigned such that participants read one of two articles

(DNA, marijuana) about a piece of health misinformation. After reading the article and answering any quiz questions, half of participants were asked to immediately rate the accuracy of the false claim from the article and answer two recall questions about the content of the fact check. The other half of participants were asked to rate the accuracy of the claim and answer the two recall questions one week later. Participants then completed a basic demographic questionnaire and were compensated for their participation.

Measures

Independent variables

Quiz condition. The primary independent variable in this study is the quiz condition, or whether participants received no quiz versus receiving a quiz after reading a fact check. The quizzes consisted of two questions about content directly pertaining to the claim in question (see Appendix for full wording). After answering each of the questions, participants received a notification with whether the answer option they chose was correct or incorrect.

Article condition. To avoid possible effects due to topic differences, we tested the efficacy of the quiz intervention for two different articles pertaining to a piece of health misinformation. The first was an article about DNA titled, “Do Women Retain DNA from Every Man They Have Slept With?” and the second was an article about marijuana titled, “Did a New Study Show that Marijuana Leads to a Complete Remission of Crohn’s Disease?” Both articles were fact checks from Snopes that had been rated entirely false.

Test condition. The time of testing varied to ascertain whether the effects of quizzes were immediate and/or would hold up after a one-week delay. To test this, half of participants received the post-treatment measures immediately after reading the fact check. The other half of participants received the post-treatment measures one week after reading the fact check.

Dependent variables

Accuracy. Participants were asked to rate the accuracy of the false claim that was the subject of the fact check to which they were assigned. For the DNA article, this statement read, “Women retain DNA from every man they have ever slept with,” and for the marijuana article, this statement read, “Marijuana leads to a complete remission of Crohn’s disease.” Responses were measured on an 11-point scale from *very inaccurate* (0) to *very accurate* (10).

Proportion of correct responses. Participants received two open-end questions and were instructed to answer the questions in 1-2 sentences. Their responses to these questions were coded by two independent coders to identify whether the response was (1) correct, (2) partially correct, (3) incorrect, lure response from quiz, (4) other incorrect answer, or (5) don’t know. Krippendorff’s alpha for this measure was **XX**. A dummy variable of *correct* (1) and *incorrect* (0) was created for each of the questions. Correct responses were summed and divided by two to obtain a proportion of the correct responses provided by participants. This proportion of correct responses was used as a measure of memory about the debunked information.

Results

First, we examine how quizzing affected accuracy ratings and memory for key facts on those who received the post-treatment measures with no delay. We ran a 2 (quiz, no quiz) x 2 (article: DNA, marijuana) ANOVA on participants’ accuracy rating. Results indicate no significant effect of quiz or article on accuracy. We ran a second 2 (quiz, no quiz) x 2 (article: DNA, marijuana) ANOVA on participants’ memory for the information in the fact check. Results here indicate a significant main effect of quizzes on proportion of correct responses given by participants, $F(1, 297) = 75.37, p < .001$. As shown in Figure X, those in the quiz condition ($M = 0.81, SD = 0.30$) answered more questions correctly than those in the no quiz

condition ($M = 0.48$, $SD = 0.37$). Results also indicate a significant main effect of article on the proportion of correct responses, $F(1, 297) = 20.20$, $p < .001$. Those in the marijuana article condition ($M = 0.73$, $SD = 0.33$) answered more questions correctly than those in the DNA article condition ($M = 0.55$, $SD = 0.41$). The interaction between quiz and article conditions was also significant, $F(1, 297) = 3.89$, $p = 0.05$.

Next, we examine whether there are similar effects on accuracy and memory for key facts for those in the delayed test condition. We again ran a 2 (quiz, no quiz) x 2 (article: DNA, marijuana) ANOVA on participants' accuracy rating. Results indicate no significant effects of quiz or article condition on accuracy for those in the delayed condition. We ran a second 2 (quiz, no quiz) x 2 (article: DNA, marijuana) ANOVA on participants' memory for the information in the fact check. Results indicate a significant effect of quizzing on proportion of correct responses, $F(1, 224) = 26.90$, $p < .001$. Those in the quiz ($M = 0.36$, $SD = 0.35$) condition answered more questions correctly than those in the no quiz condition ($M = 0.14$, $SD = 0.25$). There was no main effect of article condition on the proportion of correct responses. The interaction between quiz and article conditions was also not significant.

Finally, we examine differences between the no delay and delayed test conditions. We do this first with a 2 (quiz, no quiz) x 2 (article: DNA, marijuana) x 2 (test: no delay, delayed) ANOVA on participants' accuracy rating. Results indicate a significant main effect of the article condition on accuracy, $F(1, 521) = 4.86$, $p < .05$. There is also a marginally significant three-way interaction between quiz, article, and delay conditions, $F(1, 521) = 18.77$, $p = 0.10$. Next, we ran a 2 (quiz, no quiz) x 2 (article: DNA, marijuana) x 2 (test: no delay, delayed) ANOVA on participants' proportion of correct responses. As shown in Figure 2, there was a significant main effect of quizzing on proportion of correct responses, $F(1, 521) = 92.50$, $p < .001$. Again, those

in the quiz condition answered a greater proportion of quiz questions correctly than those in the no quiz condition. There was also a significant main effect of article on proportion of correct responses, $F(1, 521) = 5.92, p < .05$. Again, those in the marijuana article condition answered significantly more questions correctly than those in the DNA article condition. The main effect of the delay was also significant, $F(1, 521) = 196.59, p < .001$, where participants who answered questions immediately after reading the article answered more correctly than those who answered one week later. Additionally, there was a significant interaction of article and delay, $F(1, 521) = 12.97, p < .001$ and a significant interaction of quiz condition and delay, $F(1, 521) = 4.15, p < .05$. The three-way interaction between quiz, article, and delay was not significant.