

Assessing the Quality, Engagement, and Comprehension of Two Lay Summaries

Arya Patel, David Polotan-Griffiths, Isha Patel, Mikayla Rodrigues, Sahil Rizwan

Faculty of Science, McMaster University

LIFESCI/SCICOMM 3P03: Science Communication in Life Sciences

Dr. Patrycja Jazwiec

TA: Julia Tikhonov

March 9, 2025

Abstract

Increasing public awareness and participation with scientific research requires effective communication to non-expert audiences. The goal of lay summaries is to make complex research easier to understand, but their effectiveness is dependent on readability, clarity, and engagement. This study looks at how audience engagement and comprehension are affected by the caliber of lay summaries. Participants in a randomized controlled trial were given either a low-quality or an enhanced, high-quality lay summary. Independent assessors assessed the quality of the summaries, and participants had to be at least 18+ years old and not hold a post-secondary science degree. Understanding and engagement were evaluated by multiple-choice and Likert scale questions. An unpaired independent t-test was used to analyze the results. The findings show that comprehension was significantly improved by excellent lay summaries ($p < 0.001$). Participants in the intervention group were more likely to rate the summary as engaging and provide accurate answers to comprehension questions ($p = 0.0439 - 0.0002$). The participants' desire to learn more did not differ significantly ($p = 0.2220$). These results highlight the need for more precise, uniform standards for science communication in order to guarantee that research is communicated successfully and to reduce the possibility of false information.

Introduction

A lay summary is a plain-language explanation of a research article designed to enhance the accessibility to non-experts and the general public. Authors should communicate their findings in concise, clear sentences, minimize jargon, and use analogies to explain complex ideas when needed (Kirkpatrick et al., 2017). Scientists often see lay summaries as an additional burden in the peer-review process. However, adopting lay summaries can enhance the visibility, impact, and transparency of their scientific research, especially in the evolving media landscape (Kuehne & Olden, 2015). Lay summaries can create a direct pathway for dialogue between scientists and the general public, policymakers, journalists, and other experts, and can serve as a “proactive measure against the common fear that their work will be misinterpreted or misrepresented in the media” (Kuehne & Olden, 2015). Research suggests that scientists who adopt lay summaries enjoy an enhanced reputation and experience career benefits. It also promotes a positive relationship with journalists, who value scientists that can communicate clearly and accessibly (Kuehne & Olden, 2015).

These summaries are often used by various news sources, where it is the media or online journalists. to inform the public about scientific discoveries. Summaries that are poorly written can spread false information and weaken the public's trust in research (Tapia, 2020). Research results are regularly used by governments as well to inform funding and legislative choices, particularly in fields like healthcare and technology (Bero, Chiu, & Grundy, 2019). Policies that are implemented may be based on incorrect conclusions or insufficient data if these summaries are imprecise or omit important information. Similar to this, lay summaries are a crucial tool in the education system to help students understand difficult scientific ideas (Falkenberg et al., 2024). If lay summaries obtain accurate information, students may develop misconceptions and unintentionally spread misinformation. Development in technology, healthcare, and public policy

are all greatly influenced by scientific studies. However, the general public are frequently prevented from interacting with new discoveries due to the complex nature of academic language (King et al., 2017).

In general, a lay summary should explain the paper's background and significance in accessible language, answering key questions like the who, what, where, when, why, and how. One major limitation for scientists in publishing lay summaries is the lack of training in broad communication skills for public audiences outside academia. By creating training opportunities like in-person workshops, journal guidelines, and peer-review networks, it allows authors to preserve the integrity of their original work while making it accessible to a broader audience (Kuehne & Olden, 2015). Lay summaries are an essential step in successfully bridging the knowledge gap between authors and readers and increasing the public's confidence in science.

Simplifying lay summaries poses a number of difficulties for researchers. It can be challenging to summarize conclusions from scientific research without simplifying or distorting the data because of the complex procedures (Baram-Tsabari et al., 2020). Decreasing the complexity of the paper could leave out important details, which could cause miscommunications or the spread of false information (Goldstein & Krukowski, 2023). It is challenging for many researchers to put complicated concepts into understandable language while maintaining key meaning because they receive little to no formal training in good science communication (Falkenberg et al., 2024). Lay summaries need to be tailored to a wide range of readers with different degrees of background knowledge. This helps provide a balance between content richness and accessibility (FitzGibbon et al., 2020).

This study builds upon previous research that demonstrates how lay summaries enhance the accessibility of scientific findings, making them more comprehensible and engaging for broader audiences. Researchers frequently find it difficult to modify their writing for the general public, which can lead to summaries that are still too technical (FitzGibbon et al., 2020) (Freeling et al., 2021). Furthermore, there is little data on the long-term effects of various summary formats on audience engagement and understanding (Falkenberg et al., 2024). The purpose of this research study is to determine the impact of high- and low-quality lay summaries on audience comprehension and engagement. The research question for this study is: How do high- and low-quality lay summaries impact readers' comprehension and engagement in scientific research? Although readability and language challenges in scientific communication have been studied in the past, little is known about how various lay summary formats and writing styles impact audience engagement, trust, and retention in scientific research (FitzGibbon et al., 2020). Improving science communication requires an understanding of how readership and comprehension are impacted by the quality of lay summaries. The need for more precise writing instructions and improved researcher training would be highlighted if well-organized, easily readable summaries resulted in increased understanding and engagement. On the other hand, inadequate summaries draw attention to the dangers of poorly conveyed science if they cause

misunderstanding or disinterest. This study intends to shed light on these impacts in order to improve public comprehension and confidence in scientific research by refining lay summaries.

Methods

The data for the study was collected via anonymous survey participant responses. However, participants could list their names and email to draw a prize. Questions aimed at understanding lay summary quality and reading comprehension of lay summary by the general public. Two lay summaries were evaluated, one of which was assessed to be of low quality by two independent raters following a rubric versus an improved rewritten version of the same lay summary. Participants had to be at least 18 without a post-secondary degree in the sciences. Many multiple-choice questions were presented, asking participants if they believed the study was easy to understand and would benefit people with chronic hepatitis C within five years. Other questions include whether participants believe scientific information should be easily accessible if they found the lay summary interesting or if they would share it on social media. Demographic questions included level of education, birth month, age, and if English was their first language. The data was analyzed by creating a control group via birth month, where participants born from January to June represent the control group (low-rated lay summary) and participants from June to December represent the intervention group (improved lay summary). Raw data was transformed into measurable data points capable of being statistically analyzed. An unpaired independent T-test was used to analyze and investigate the significance of the two groups. A prism graph was used to calculate the results and generate graphs of the unpaired T-test. The two groups' calculated mean, distribution and standard deviation were compared, and the statistical significance was analyzed to see if there was a meaningful difference between the control and intervention groups.

Results

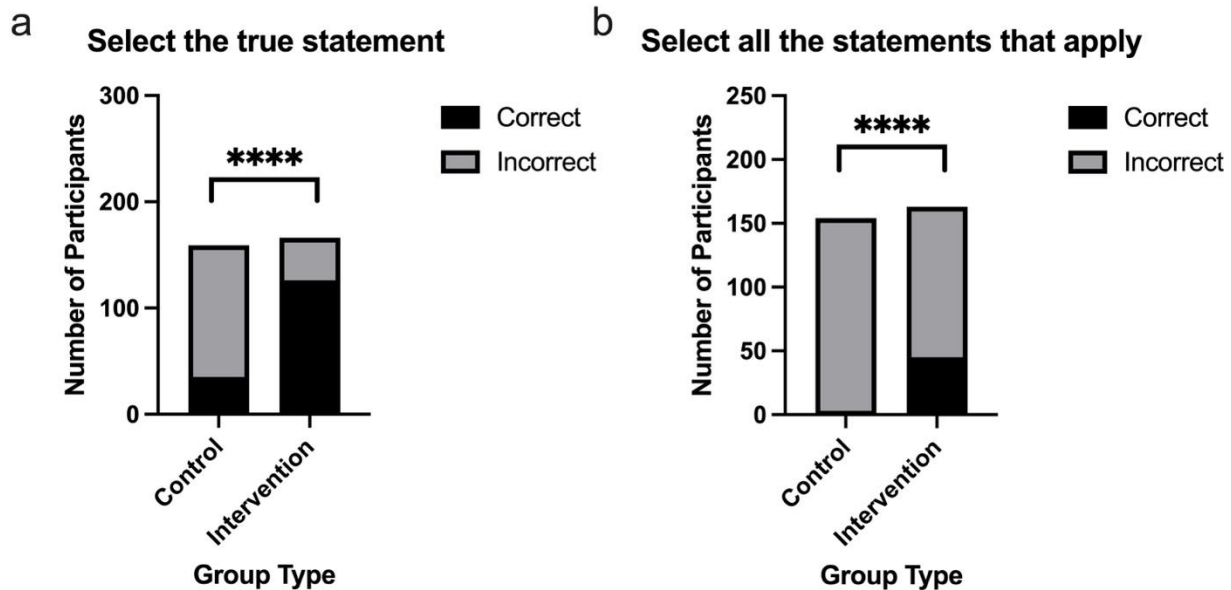


Figure 1. Audience Comprehension of Low-Quality vs High Quality Lay Summary.

Comprehension was assessed based on the number of participants in the control and intervention groups who answered the survey question correctly or incorrectly. This panel of graphs shows the number of participants who (a) were asked to select the true statement from a set of options, with only one correct answer ($n = 325$) and (b) were asked to select all the statements that apply, with only one combination of correct answers ($n = 317$). The statistical significance is represented by asterisks, ****.

To evaluate audience comprehension of low-quality versus high-quality lay summaries, participants were asked to select the correct response to two survey questions designed to assess their understanding, as shown in Figure 1. Figure 1a illustrates the responses to the first question, which had only one correct answer. In the control group, there were 163 participants, with 4 not applicable responses, 35 correct responses, and 124 incorrect responses. The intervention group consisted of 170 participants, with 4 blank responses, 126 correct responses, and 40 incorrect responses. Figure 1b shows the responses to the second question, where participants had to select all applicable answers, with only one correct combination. In the control group, there were 160 participants, with 6 no answers, 3 correct responses, and 151 incorrect responses. The intervention group had 169 participants, with 6 no answers, 45 correct responses, and 118 incorrect responses. Overall, participants in the intervention group demonstrated a significantly higher percentage of correct answers compared to the control group, with a p -value < 0.001 for both questions.

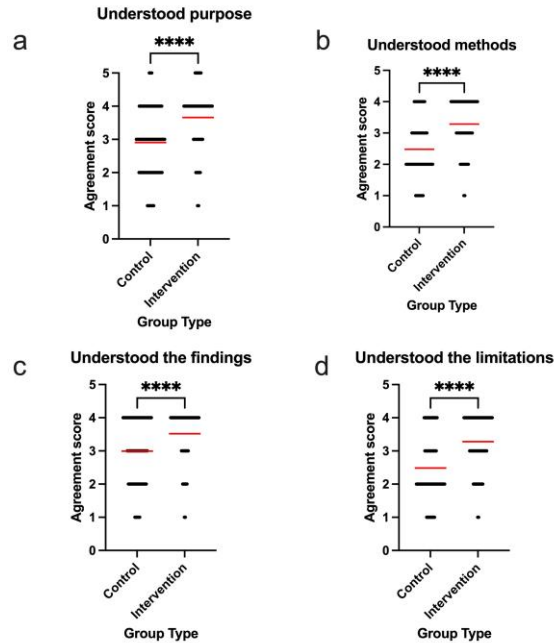


Figure 2. Assessment of Comprehension for Lay Summary Components. This panel of graphs shows the agreement scores on a scale of 0 to 5 for participants in the control and intervention groups regarding their understanding of the lay summary's purpose (a), methods (b), findings (c), and limitations (d). The red horizontal line represents the mean agreement score and statistical significance is represented by asterisks, ****.

To assess the comprehension of the lay summaries based on the components, purpose, methods, findings and limitations, participants from the control and intervention groups were asked to rank their agreement score as seen in Figure 2. On a scale of 1-5, with 0 being the least in agreement, and 5 being the most in agreement, differences in participant rankings were observed across all components. The mean values for the control group were 2.908 (a), 2.481 (b), 2.994 (c), and 2.485 (d). In contrast, the intervention group had mean values of 3.659 (a), 3.284 (b), 3.518 (c), and 3.282 (d). For the evaluation of understanding the purpose, methods, findings and limitations, participants from the intervention group demonstrated significantly higher agreement scores than the control group with a p-value < 0.001 across all four components.

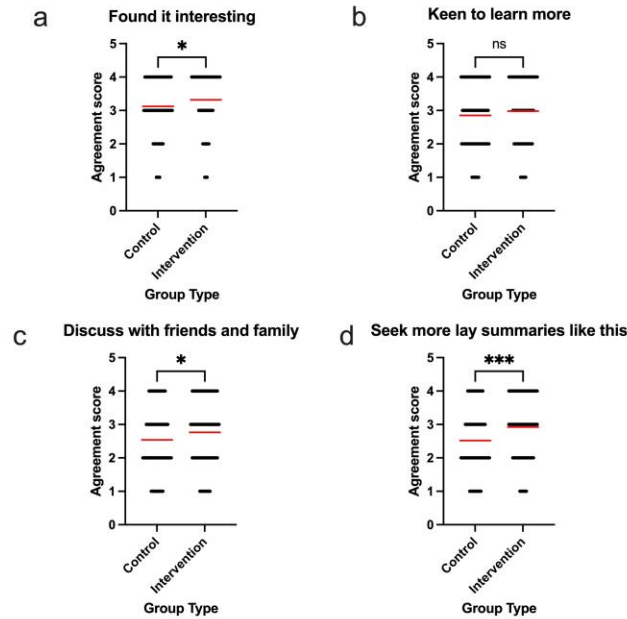


Figure 3. Assessment of Engagement with the Lay Summary. This panel of four graphs displays the agreement scores on a scale of 0 to 5 for participants in the control and intervention groups regarding their engagement with the lay summary. Each graph represents the following survey statements: (a) I found this lay summary interesting, (b) I am keen to learn more about this type of research after reading this lay summary, (c) I would discuss this study with family, friends, and/or colleagues, and (d) I would seek out more summaries of research like this one. The red horizontal line represents the mean agreement score and statistical significance is represented by asterisks and non-significance by "ns".

To assess the engagement with the lay summaries, Figure 3 highlights the differences in agreement scores between the control and intervention groups after participants were asked to rank different statements. The mean values for the control group were 3.123 (a), 2.852 (b), 2.540 (c), and 2.519 (d). In contrast, the intervention group had mean values of 3.318 (a), 2.982 (b), 2.765 (c), and 2.924 (d). Figure 3a specifically looks at whether participants found the lay summaries interesting which revealed a p-value of 0.0439. On the other hand, Figure 3b showcases the difference in control and intervention group for whether were keen to learn more, and significance tests reveal a p-value of 0.2220 indicating no significance. For Figure 3c, participants in the intervention group were more likely to demonstrate intent to discuss with friends and family compared to the control group. A p-value of 0.0393 was observed indicating significance between the groups. Finally, as seen in Figure 3d, the intervention group for the question of whether participants would seek out more lay summaries of similar research in the future revealed a significant difference as seen with a p-value of 0.0002.

Discussion

This study compared the quality of two lay summaries using survey responses to assess engagement and comprehension levels across an interventional and control group. Participants' comprehension of the lay summary was evaluated using multiple choice or “select all that apply” questions. Multiple choice evaluates if an individual can identify one correct answer among a set of options (Dell & DeVries, 2024). Contrastingly, “select all that apply” assesses if an individual can choose all correct answers while excluding incorrect ones. This method demands more cognitive load and in-depth understanding of the content compared to multiple choice (Dell & DeVries, 2024). Figure 1a concludes that only a few control participants, about 35, selected the correct multiple-choice statement. On the other hand, most of the intervention participants, about 126, correctly recognized the true statement. This variation in data stems from differences in the accessibility, clarity, and structure of both lay summaries. The low-quality summary uses a technical and passive writing style, including scientific terms such as “*extracellular vesicles*” and “*virological cure*” without definitions or context. This assumes prior knowledge of Hepatitis C infections, making it difficult for the general audience to understand. Given that the high-quality summary was designed to be more accessible, it defines jargon, connects key ideas, and provides explanations. For example, the authors defined extracellular vesicles as fluid-containing sacs that carry small molecules between cells, simplifying an unknown term. Consequently, this helped intervention participants connect new information to familiar knowledge, improving their ability to correctly answer the first question.

Self-assessment metrics evaluate abilities, processes, outcomes, and attitudes of oneself (Andrade, 2019). In this study, these metrics provided feedback on lay summary qualities, which can inform future science communication and learning improvements (Andrade, 2019). The Likert scale, a type of self-assessment, was used to evaluate comprehension of the lay summary components. The average mean of high-quality agreement scores was statistically higher in all four components compared to the control group (Figure 2). To explain these results, the intervention summary introduced Hepatitis C as an infection that leads to cancer and liver dysfunction. Here, the clear framing identifies the purpose of the study and the health consequences of the disease, which increases engagement of the non-expert audience. This summary also describes the study in a chronological step-by-step method, walking the reader through the purpose, methods, results, and limitations. This structured format improves readability by reducing information overload (Pickren et al., 2022). The readers do not need to worry about context gaps, leading to better digestion, retention and engagement of the summary (Pickren et al., 2022). Conversely, the control summary lacks details on the experimental methods, participant recruitment, and data collection. This leaves readers with a limited understanding of the research design. For instance, it fails to define scientific terms like “*fibrosis*” or “*cirrhosis*” and their implications. As a result, the reader would need to conduct a Google search to understand these key terms, which does not make the summary accessible. Past research indicates that when texts require external efforts to comprehend, readers are less likely to engage or interact with the information (Pickren et al., 2022).

This study was conducted as an RCT, where participants were randomly assigned to control or intervention groups based on their birth month. Prospective RCTs are “gold standards” of research as they reduce biases and examine cause-and-effect relationships between the outcome and intervention (Hariton & Locascio, 2018). To illustrate, the study showed strong statistical significance across all four components ($p\text{-value} < 0.001$) in Figure 2, making the data reliable. This improves internal validity of the study as the observed differences between groups arise from lay summary differences and not confounding variables. Here, participants were 18 and older, where the randomization accounts for variations in education level, background knowledge, sex, language proficiency, and other demographic factors. The purpose of this was to ensure the study sample represents a broad population from which the data was drawn, allowing for generalization of the findings (Stuart et al., 2018).

Nonetheless, there are limitations observed in this study. For one, Figure 1b illustrates that only 3 control and 45 intervention participants selected all the correct statements. “Select all that apply” questions are more challenging than multiple choice as the participant must recall more than one detail. This makes them more likely to second-guess their answers and reduces the ability to use the process of elimination. Despite the improved intervention lay summary, the low results indicate further refinements must be made. Future improvements can focus on using visual aids, such as diagrams or infographics, to make the information more accessible. An illustrative flowchart can show images with arrows connecting the different steps. More specifically, it can visualize how blood samples were collected from Hepatitis C patients and healthy donors, the extraction process of extracellular vesicles, and how RNA was analyzed. These images allow the reader to form mental representations of the processes, where they can produce a connection with the words. Another limitation is that the comprehension was self-reported. This creates the risk of response bias, where participants can overestimate or underestimate their understanding (Rosenman et al., 2011). There is also the introduction of interrater bias limitations in this study. Here, each participant will interpret and evaluate the summaries differently (O’Neill, 2017). This will affect the degree to which individuals evaluating the same lay summary agree on the Likert scale ratings (O’Neill, 2017). Since participants were not trained to evaluate lay summaries, the differences in past education, reading habits, and perceptions will influence their responses in Figures 2 and 3. To improve this, future studies can include direct assessments, such as short-answer questions or applied knowledge tests that require participants to actively recall information (Rosenman et al., 2011). Researchers can also include indirect measures that examine the reading habits and comprehension levels of participants. For perspective, they could be asked how often they require assistance from others while understanding written information in their daily activities.

Moving on, Figure 3 shows differences in engagement levels between the control and intervention groups. Intervention participants scored significantly higher in finding the summary interesting, discussing it with others, and expressing a greater chance of reading similar summaries in the future. This data suggests that the improved lay summary increases immediate interest in the topic. However, Figure 3b shows no significant difference in participants’ desire to

learn more, indicating that engagement with the content does not always foster long-term curiosity in readers. The lack of a personal connection between the summary and reader may explain this result. To account for this limitation, future research should explore interactive communication methods that incorporate individualized content and multimedia elements (Dwivedi et al., 2021). A digital tool could allow readers to insert health-related factors into a database, such as their medical history, to visualize how Hepatitis C progression varies under personal conditions (Dwivedi et al., 2021). Moreover, an animated visual showing the progression of liver scarring can make the summary easier to understand.

The following steps for this research involve replicating the study with a substantial and diverse sample size to improve statistical reliability and generalizability. A potential method for enhancing the randomization process could be stratified randomization, the highest-level gold standard for survey-based RCTs. Stratified randomization will place the cohort into strata based on a demographic trait (such as level of education) and randomly assign them after to ensure prevention of type 1 error, balancing the cohorts, and the potential for subgroup analysis. Additionally, future studies could test different lay summary formats, such as interactive summaries, visual-based explanations, or video abstracts, to assess their impact on comprehension and engagement. Researchers can collaborate with policymakers, science communicators, and subject-matter experts to take a multidisciplinary approach to refining these lay summaries. Finally, longitudinal studies may identify whether engaging with accessible summaries leads to prolonged interest and active learning over time. Therefore, it is imperative to increase research on the efficacy of lay summaries to assist with educating the public, as transparency is increasingly necessary with the rise of misinformation influencing political outcomes in democracies around the world (McLoughlin et al., 2024).

References

- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00087>
- Baram-Tsabari, A., Wolfson, O., Yosef, R., Chapnik, N., Brill, A., & Segev, E. (2020). Jargon use in Public Understanding of Science papers over three decades. *Public Understanding of Science*, 29(6), 644-654. <https://doi.org/10.1177/0963662520940501>
- Bero, L., Chiu, K., & Grundy, Q. (2019). The SSSPIN study—spin in studies of spin: meta-research analysis. *BMJ*, 367, l6202. <https://doi.org/10.1136/bmj.l6202>
- Dell, K. A., & DeVries, D. M. (2024). Effect of changing multiple choice questions from “all of the above” to “select all that apply.” *Currents in Pharmacy Teaching and Learning*, 16(3), 174–177. <https://doi.org/10.1016/j.cptl.2023.12.034>
- Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, 59, 102168. <https://doi.org/10.1016/j.ijinfomgt.2020.102168>
- Falkenberg, L. J., Patrick, & Soranno, P. A. (2024). How to write lay summaries of research articles for wider accessibility. *Limnology and Oceanography Letters*. <https://doi.org/10.1002/lol2.10373>
- FitzGibbon, H., King, K., Piano, C., Wilk, C., & Gaskarth, M. (2020). Where are biomedical research plain-language summaries? *Health Science Reports*, 3(3), e175. <https://doi.org/10.1002/hsr2.175>
- Freeling, B. S., Doubleday, Z. A., Dry, M. J., Semmler, C., & Connell, S. D. (2021). Better writing in scientific publications builds reader confidence and understanding. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.714321>
- Goldstein, C. M., & Krukowski, R. A. (2023). The importance of lay summaries for improving science communication. *Annals of Behavioral Medicine*, 57(7), 509-510. <https://doi.org/10.1093/abm/kaad027>
- Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—The gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics and Gynaecology*, 125(13), 1716. <https://doi.org/10.1111/1471-0528.15199>

- King, S. R., Pewsey, E., & Shailes, S. (2017). An inside guide to Elife digests. *Elife*, 6, e25410. <https://doi.org/10.7554/eLife.25410>
- Kirkpatrick, E., Gaisford, W., Williams, E., Brindley, E., Tembo, D., & Wright, D. (2017). Understanding Plain English summaries. A comparison of two approaches to improve the quality of Plain English summaries in research reports. *Research Involvement and Engagement*, 3(17), 1-14. <https://doi.org/10.1186/s40900-017-0064-0>
- Kuehne, L. M., & Olden, J. D. (2015). Opinion: Lay summaries needed to enhance science communication. *Proceedings of the National Academy of Sciences of the United States of America*, 112(12), 3585–3586. <https://doi.org/10.1073/pnas.1500882112>
- McLoughlin, Killian Lorcan, William J. Brady, Aden Goolsbee, Ben Kaiser, Kate Klonick, and Molly Crockett. *Misinformation exploits outrage to spread online*, November 29, 2024. <https://doi.org/10.31234/osf.io/ecxp4>.
- O'Neill, T. A. (2017). An overview of interrater agreement on likert scales for researchers and practitioners. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00777>
- Pickren, S. E., Stacy, M., Del Tufo, S. N., Spencer, M., & Cutting, L. E. (2022). The contribution of text characteristics to reading comprehension: Investigating the influence of text emotionality. *Reading Research Quarterly*, 57(2), 649–667. <https://doi.org/10.1002/rrq.431>
- Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural & Healthcare Research*, 2(4), 320–332. <https://doi.org/10.1504/IJBHR.2011.043414>
- Stuart, E. A., Ackerman, B., & Westreich, D. (2018). Generalizability of randomized trial results to target populations: Design and analysis possibilities. *Research on Social Work Practice*, 28(5), 532–537. <https://doi.org/10.1177/1049731517720730>
- Tapia, L. (2020). COVID-19 and Fake News in the Dominican Republic. *The American Journal of Tropical Medicine and Hygiene*, 102(6), 1172-1174. <https://doi.org/10.4269/ajtmh.20-0234>