

2024

## Heed the Warning Signs: The Effectiveness of Message Popup Warnings for Deterring the Spread of Misinformation

Hollis Greenberg

Follow this and additional works at: [https://nsuworks.nova.edu/gscis\\_etd](https://nsuworks.nova.edu/gscis_etd)



Part of the [Computer Sciences Commons](#)

## Share Feedback About This Item

---

This Dissertation is brought to you by the College of Computing and Engineering at NSUWorks. It has been accepted for inclusion in CCE Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact [nsuworks@nova.edu](mailto:nsuworks@nova.edu).

Heed the Warning Signs: The Effectiveness of Message Popup Warnings for  
Deterring the Spread of Misinformation

by

Hollis J. Greenberg

A dissertation submitted in partial fulfillment of the requirements  
as part of the degree of Doctor of Philosophy  
in  
Information Systems

College of Computing and Engineering  
Nova Southeastern University

2024

Approval Page

We hereby certify that this dissertation, submitted by Hollis J. Greenberg conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Laurie P. Dringus, Ph.D.  
Chairperson of Dissertation Committee

3/5/24  
Date

  
\_\_\_\_\_  
Ling Wang, Ph.D.  
Dissertation Committee Member

3/5/24  
Date

  
\_\_\_\_\_  
Junping Sun, Ph.D.  
Dissertation Committee Member

3/5/24  
Date

Approved:

  
\_\_\_\_\_  
Meline Kevorkian, Ed.D.  
Dean, College of Computing and Engineering

3/5/24  
Date

College of Computing and Engineering  
Nova Southeastern University

An Abstract of a Dissertation Submitted to Nova Southeastern University  
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

## Heed the Warning Signs: The Effectiveness of Message Popup Warnings for Deterring the Spread of Misinformation

by  
Hollis J. Greenberg  
March 2024

As false news can propagate to others rapidly, social media platforms employ multiple methods to combat misinformation. Debunking techniques are warning features embedded into a platform's interface that alert readers of misinformation. These warnings have two goals: to "debunk" false information and to prevent the further spread of misinformation. Researchers have evaluated the effectiveness of debunking techniques to understand how users increase their awareness of misinformation, and what users do with the information given in warning messages. Message popup warnings are a newer and understudied type of debunking technique.

The overarching research question of this study was: Are message popup warnings effective for deterring the spread of misinformation? The research goal was to determine the effectiveness of message popup warnings in the context of the user's choice behaviors of liking, sharing, or commenting on misinformation. Four sub-questions explored: (RQ1) to what extent users read the full content of posting, with or without warnings (user engagement), (RQ2) what differences occur in user's choice responses to the posting, with or without warnings (effectiveness), (RQ3) how users rate the credibility of the posting, with and without warnings, and (RQ4) how users rate the usability and usefulness of message popup warnings as a debunking technique.

The study was conducted as a between-subjects experimental design using an online interactive scenario and survey (ISS). The participants ( $N = 109$ ) were recruited from the student population at Wentworth Institute of Technology (WIT) and were randomly assigned to either the control or treatment group. The ISS combined interactive Facebook-like scenarios (i.e., postings) that contained links to the full text of an article to be read and an embedded survey that prompted users to take an action or no action with the posting. By adapting questions from the Perceived Usefulness Measurement Scale (PUMS) and the System Usability Scale (SUS), the treatment group also assessed the usefulness and usability of message popup warnings.

Three hypotheses (H1-H3) were created as null hypotheses to test the first three research questions. Regarding H1, no significant differences were found for user engagement between the treatment group (presented with message popup warnings) and the control group (presented without warnings). For H2, there were no significant differences found for effectiveness between the treatment and control groups in user choice behaviors. However, when the data was compared within the treatment posting types (false postings with warnings, false postings without warnings, and true postings) and compared

responses from before and after the appearance of the warning, there were significant differences found in user choice actions. Regarding H3, there were significant differences found for credibility between treatment and control groups, suggesting that participants found postings with message popup warnings significantly less credible than postings without a warning. Additionally, data gathered from the SUS found usability unacceptable, while the PUMS indicated that the usefulness was fair with some deficiencies.

The results suggest that message popup warnings may be effective for influencing the credibility of postings and may be effective for deterring the spread of misinformation when warnings appear on some of the postings. Future studies could vary the messaging in the warning, study different participant age groups, and investigate user responses across multiple platforms. Additionally, platforms may want to add training messages to help users distinguish between true and false postings.

## Acknowledgements

Not a day goes by without news of false postings spreading on social media or other internet-based mediums. In the past decade, the prevalence of misinformation has become a topic of great interest to me, my family, and within my social circles. Furthermore, as I studied Human-Computer Interaction (HCI) and found that research on the propagation of misinformation was lacking, the direction of my dissertation became obvious.

I am deeply indebted to my family for enduring my long days on the computer. To my amazing children, Sophie and Max, who worked on their college and graduate degrees alongside me, your commitment to your own studies motivated me to persevere. I will love you forever. Many thanks to my wonderful partner, David, for listening to me and supporting me through these last three years. To my father and stepmother, Mel and Sandi, thank you for believing that I could get to the finish line. For my late mother Michele, without whom none of this would be possible, thank you for instilling the love education within me. And, in the spirit of sibling rivalry (for my brother, Myles), I would like to add that there are now two doctors in the family!

I was very lucky to have found a cheerleader in my advisor, Dr. Laurie Dringus. I would like to extend my heartfelt thanks and appreciation for your guidance, thoughtful feedback, and patience as I navigated my way through this study. In addition to my advisor, I would like to thank my committee members, Dr. Ling Wang and Dr. Junping Sun, for your expertise and guidance. Your feedback was much appreciated.

It really does take a village to embark on and complete a Ph.D. My friends and colleagues have been an incredible source of support during this journey: Rick Trilling, Michael Mozill, Santiago Umaschi, Cidhinnia Torres Campos, Betsy Kranz, Jill Kearney, Arlene Marcus, Deb Wood O'Brien, Michelle Gluck, Dana Haberman, Emily Banis Stoehrer, Willem Wallinga, Meridith Spencer, and Dean John Cribbs. Thank you for your encouragement and believing in me!

## Table of Contents

**Abstract ii**

**Acknowledgements v**

**List of Tables viii**

**List of Figures x**

### **Chapters**

#### **1. Introduction 1**

Background 1

    Effectiveness of Debunking Techniques 3

    Popups 5

Statement of the Problem, Goal, and Research Questions 6

    Problem and Goal 6

    Research Questions 9

Relevance and Significance 10

Barriers and Issues 11

Assumptions, Limitations, and Delimitations 12

    Assumptions 12

    Limitations 13

    Delimitations 13

Definition of Terms 14

List of Acronyms 15

Summary 16

#### **2. Review of Literature 18**

Introduction 18

How Misinformation Spreads 18

Effectiveness 19

User Engagement 22

Credibility 23

Usefulness 25

Usability 26

Background and Previous Studies 27

Summary of What is Known and Unknown About the Research 28

Summary 29

#### **3. Methodology 31**

Overview 31

Study Design 32

    Study Variables 34

    Treatment 35

Procedures 40

Expert Review and Pilot Test to Finalize ISS 40

Participants, Recruitment, and Selection 44

Instrumentation Setup, Development, and Quality Control 46

- Environment and Setup for the Study 46
- Instrument Development 47
- Quality Control 51
- Data Collection – Multiple Sources 52
- Data Organization and Analysis 54
- Researcher, Ethical Considerations, and IRB 59
- Resources 60
- Summary 61

#### **4. Results 63**

- Overview 63
- Data Analysis 65
  - User Engagement 66
  - Effectiveness 69
  - Credibility 74
  - Usability and Usefulness 78
- Summary of Results 81

#### **5. Conclusions 83**

- Introduction 83
- Conclusions 83
  - User Engagement 83
  - Effectiveness 85
  - Credibility 87
  - Usability and Usefulness 88
- Strengths, Weaknesses, and Limitations 89
- Implications 91
- Recommendations for Future Research 93
- Summary 95

#### **Appendices**

- A. Interactive Scenario and Survey (ISS) 102
- B. Informed Consent Form 124
- C. ISS Image Sources for Photographs and Drawings 126
- D. Institutional Review Board Documentation (NSU) 128
- E. Institutional Review Board Documentation (WIT) 130
- F. Recruitment Flyer 131
- G. Think-Aloud Checklist and Form 132

#### **References 134**



## List of Tables

### Tables

1. Constructs and Evaluation Techniques by Research Question 33
2. Interactive Scenario and Survey (ISS): Questions to Display after Each Article 38
3. Pilot Test Questions 41
4. Participant Pre-screening Questions 45
5. Usefulness Questions Based on Davis' (1989) Perceived Usefulness Measurement Scale 50
6. Usability Questions Based on Brooke's (1995) System Usability Scale (SUS) 51
7. Descriptive Statistics of Study Participants ( $N = 109$ ) 64
8. User Engagement Comparison for the Treatment and Control Groups 67
9. Comparison of Control and Treatment Groups for User Engagement by Posting Number 67
10. Means, Standard Deviations, and One-Way Analyses of Variance for User Engagement Among Posting Types 69
11. Frequencies and Chi-Square Results for Effectiveness Comparing Control and Treatment Groups 71
12. Crosstabulation and McNemar's Test of Effectiveness Comparing Pre- and Post-Test Results 72
13. Crosstabulation and Chi-Square Results for Effectiveness for All Posting Types 72
14. Frequencies for Effectiveness by Posting Type 73
15. Comparison of Control and Treatment Groups 75
16. Comparison of Control and Treatment Groups for Credibility by Posting Number 75

17. Means, Standard Deviations, and One-Way Analyses of Variance for Credibility  
Among Posting Types 77
18. Tukey's HSD Test to Compare Credibility Ratings for Postings with Popup Warnings  
with Other Posting Types 77
19. SUS Scores Using Bangor et al.'s (2008) Acceptability Ranges 78
20. SUS Scores Using Bangor et al.'s (2009) Descriptive Adjectives 79
21. Means, Standard Deviations, and Variance for PUMS Scores by Question 80
22. Results of Hypotheses 81

## List of Figures

### Figures

1. Example of Twitter's Message Popup Warning 10
2. Study Variables 35
3. Flowchart of Experiment Questions 39
4. Procedure Diagram 40
5. Revised ISS Instructions 43
6. Comparison of Original and Revised Posting Format 44
7. False News Posting in Newsfeed Prior to Message Popup Warning 48
8. False News Posting with Message Popup Warning 49
9. SUS Ratings by Adjective and Acceptability Ranges 57
10. SUS Percentile Rankings 58

## Chapter 1

### Introduction

#### Background

Although *fake news* is a contemporary term, in the past, false or incorrect news was called a hoax, conspiracy theory, or propaganda (Allcott & Gentzkow, 2017). Today with social media platforms, news outlets, and other Internet content sharing platforms, individuals are exposed to a variety of news content. Not all news sources use effective methods of fact-checking before publishing information online and, inadvertently, false information is published (Allcott & Gentzkow, 2017; Moravec et al., 2020). Some posted content contains mistakes and some posted content is purposely meant to mislead the reader (Allcott & Gentzkow, 2017; Kirchner & Reuter, 2020). False news can be categorized as *misinformation*, such as incorrect news, or *disinformation* and fake news, terminology used to describe intentional misinformation (Geeng et al., 2020; Kirchner & Reuter, 2020). Since social media platforms provide software features for users to repost or share information with others, these affordances unintentionally aid in the spread of misinformation (Allcott & Gentzkow, 2017). This study focused specifically on false information (i.e., misinformation) (Geeng et al., 2020) and warnings presented in software features to alert users that the veracity of the posting is disputed (Pennycook et al.).

Social media platforms employ multiple methods to combat misinformation. Debunking techniques are message features embedded into a platform's interface that

alert readers of misinformation with the goal to ‘debunk’ false information and to prevent the further spread of misinformation. Facebook uses several techniques to alert users of false content such as a warning screen covering the content of posting, providing disputed fact-checking information, and placing the false content lower in users’ newsfeeds (Meta, n.d.). Comparatively, Twitter alerts users of false content in the following ways: a label and/or warning on the tweet, a message popup warning when attempting to sharing or liking, turning off the user’s ability to reply, like, or retweet the message, providing links for additional explanation, and reducing visibility and/or not recommending the posting (Twitter, 2021). In certain cases, both Meta and Twitter may remove questionable postings from the platform (Meta, n.d.; Twitter, 2021).

While debunking techniques alert the user about the credibility or accuracy of posted information, social media platforms also utilize algorithms for both curating newsfeed content and alerting users of misinformation (Geeng et al., 2020). On Facebook, algorithms determine where false postings appear in a user’s newsfeed, as well as detect misinformation (Kirchner & Reuter, 2020). Unlike human misinformation detectors, algorithms typically do not explain to users why the posting is false (Kirchner & Reuter, 2020). Furthermore, Facebook algorithms learn users’ interests; if users share or comment on undetected false news, other postings with the same false message will appear higher in the user’s newsfeed (Ardèvol-Abreu et al., 2020). By prioritizing false postings, these algorithms, created to minimize user engagement with misinformation, can have the opposite intended effect (Ardèvol-Abreu et al., 2020).

Additionally, users control their newsfeeds and content management. On social media, news is filtered through a *filter-bubble* by the *friends* of an individual (Bentley et

al., 2019). Individuals see only the posts of the others they follow, acting as a filter (Allcott & Gentzkow, 2017). With filters, individuals may be following other like-minded people, leading to newsfeed postings that often share the same opinions and political-leanings (Bentley et al., 2019). Users tend to not fact-check what is read (Jun et al., 2017), but as members of social media groups, the amount of time spent in an online group and the number of likes or comments on a posting predicts an individual's trust in what is read (Ma et al., 2019). Additionally, individuals tend to trust articles viewed multiple times or that validate pre-existing beliefs (Spezzano et al., 2021). Alas, users do not always detect misinformation since false news can corroborate their previously held beliefs (Moravec et al., 2020). This circumstance, called *confirmation bias*, erroneously confirms the 'truthfulness' of the false news to the reader. With confirmation bias, the user is more likely to share and spread a false news story that aligns with prior beliefs (Geeng et al., 2020; Jun et al., 2017).

#### *Effectiveness of Debunking Techniques*

Researchers have evaluated the effectiveness and various affordances of debunking techniques to understand how users increase their awareness of misinformation, and what users do with the information given in the warning's message. Kirchner and Reuter (2020) compared multiple debunking techniques, such as reducing posting size, adding warnings underneath postings, confirmation popup alerts, and displaying related contradictory headlines, by evaluating user preference and the effectiveness of changing users' beliefs. Other studies have assessed effects of user comments on attitudes and actions (Colliander, 2019), content and political stance labels on article selection and perception (Gao et al., 2018), and debunking techniques and

political ideologies on user perceptions (Nyhan & Reifler, 2010). Newer social media debunking techniques, such as message popups that warn users about liking or sharing content marked as misinformation (Alonzo, 2020), have yet to be evaluated on the constructs of user engagement, usability, usefulness, and effectiveness for preventing misinformation propagation.

Gao et al. (2018), Kirchner and Reuter (2020), and Colliander (2019) investigated the effects of various debunking techniques on misinformation by focusing on user-centric attributes such as: likelihood to share, change in attitude, acceptance, influence, article selection, and article perception. In Gao et al.'s (2018) study, labels for credibility and political classification were observed to be ineffective for correcting misinformation and led to an opposite influence for both perception and reading selection. Additionally, Kirchner and Reuter's (2020) findings suggested that warning-type techniques, peer corrections, and fact-checking articles positively influenced individual's disbelief in misinformation. Likewise, Colliander (2019) found individuals were less likely to share misinformation when faced with peer corrections (from comments) than when faced with disclaimers.

For a debunking technique to be effective, it must sow doubt in the reliability of the posting (Colliander, 2019). Furthermore, the misinformation must not be accepted as the social norm so that users are not inclined to further spread the message (Kirchner & Reuter, 2020). Additionally, in Kirchner and Reuter's (2020) study, users expressed preference for the false news warning-type techniques over other more subtle debunking techniques.

### *Popups*

Popups are a traditional interface tool that Twitter has repurposed for combatting misinformation. Popups on social media may yield different results than on other websites; social media platforms are unique as users interact with content, rather than passively watch or read content (Wash & Lampe, 2012). The unique feature of popups is that the message interrupts a user's task and necessitates an immediate user action (Wash & Lampe, 2012). Responding to a popup inevitably slows down users and distracts from the intended task (Abdulin & Billman, 2016). Research on popups reveals that it can be a useful debunking tool for some users; however, for users to doubt the validity of the posting and to avoid sharing with others, the wording of the message is critical (Ardèvol-Abreu et al., 2020). Popups can be effective; research has found that popups detecting offensive and threatening language influenced 93% of users to modify message content before sending (Anitha et al., 2019). However, there are drawbacks from using popups; experienced users may not read the popup's message or may respond to the popup automatically without processing the message (Abdulin & Billman, 2016). For some users, closing popups may be a repeated behavior and the user's response to close without reading may be done impulsively, with little cognitive processing (Gu & Hong, 2019). Users also find popups to be annoying and disruptive (Abdulin & Billman, 2016). Consequently, frequent users are encouraged to avoid reading or responding to popups (Abdulin & Billman, 2016) which may affect the effectiveness of popups long term (Gu & Hong, 2019).

However, popups are used for positive messaging and remain an understudied debunking technique. Research is needed to determine if popups, as one type of



debunking design, are effective in increasing users' awareness of misinformation and in deterring spread. Effectiveness is critical to understand the usefulness of the various debunking techniques within social media platforms. Most debunking research has not focused on deterring or preventing spread of misinformation, but rather on evaluating users' affect towards misinformation and debunking techniques such as belief change, attitude, and user preferences (Colliander, 2019; Kirchner & Reuter, 2020). Assessing the effectiveness of message popups will lead to recommending ways that social media platforms can design better debunking techniques to help users recognize misinformation and deter the spread of false news (Moravec et al., 2020).

### **Statement of the Problem, Goal, and Research Questions**

#### *Problem and Goal*

To date, research has understudied the effectiveness of debunking techniques; specifically, experiments evaluating user responses and actions when debunking techniques are presented to users in real-time. Gao et al. (2018) studied participants' selection and perception of news articles with both political stance and credibility labels. Colliander's (2019) study focused on users' attitudes towards misinformation when user comments were displayed that corrected, criticized, or supported the article in question. Comparatively, Kirchner and Reuter (2020) investigated various debunking techniques for their effectiveness of changing user's belief in misinformation and user acceptance.

The goal of this study was to measure the effectiveness of message popup warnings in the context of the user's choice behaviors of liking, sharing, or commenting on misinformation, not on changing the degree of belief in the misinformation as in Kirchner and Reuter's (2020) study. Although it has been found that warning-type

techniques lessen a user's belief in false news, it is still unknown whether a user will still share and propagate the misinformation (Ecker et al., 2010; Kirchner & Reuter, 2020).

This study is unique in that message popup warnings were studied for their effectiveness in deterring the spread of misinformation. The debunking technique, message popup warnings, was shown to participants (treatment group) via postings that contain misinformation. The effectiveness of integrating message popup warnings (i.e., comparing treatment and control groups, with or without message popup warnings, respectively) was evaluated by the user's choice response (actions or inactions), of liking, sharing, and commenting (actions) or not liking, sharing, and commenting (inactions) (Oh et al., 2018) on the false news story in a Facebook-like interface. Patterns of these actions or inactions were analyzed to determine if message popup warnings lead users to make informed choices about liking, sharing, or commenting or not liking, sharing, or commenting on misinformation.

Effectiveness is a multi-faceted construct that is often evaluated in concert with other usability attributes, such as user engagement. In literature the definition of user engagement varies, depending on the technology in question, and can refer to interaction, emotional connection, involvement or immersion, style of interface, and motivation of user (Doherty & Doherty, 2018). In this study, user engagement refers to the user's involvement with debunking techniques and newsfeed postings in terms of the extent users read the full content contained in the posting. User engagement was measured by a 4-point Likert-scale decision on a survey question. If the user fully read the content of the posting, the variable was equal to 4. "Reading somewhat" was equal to 3 and "read little of article" was equal to 2. Ignoring the article, or responding "no, not at all", the variable

was equal to 1. Effectiveness refers to the user choice response (liking, sharing, or commenting) to false news. The user choice response was a binary decision; action (liking, sharing, or commenting) was equal to 1 or inaction (not liking, sharing, or commenting) was equal to 0.

Ultimately, an effective debunking technique should be useful and usable by the intended users and be appropriate for the intended purpose (Brooke, 1995). A useful tool is one in which aids in the task being performed (Davis, 1989). To be useful, the debunking technique should help users detect false postings. Relatedly, usability is defined as a system that does what is intended (effective), performs the task expeditiously (efficient), engages the user (engaging), handles errors well (error tolerant) and is easy to learn (Quesenbery, 2003). For a message warning to be deemed usable as a debunking technique, users should be able to easily use and recognize the warning as intended.

The aim of this study was to evaluate a more recent debunking technique, message popup warnings, to determine if this interface is effective to help deter the spread of misinformation. The design of the warning's message itself is an important consideration; the design impacts both the visibility of the warning and user behavior (Fan et al., 2018; Wogalter & Mayhorn, 2005). Warning-type messages can result in the intended effect of changing user's behavior (Moravec et al., 2020; Wogalter & Mayhorn, 2005). To be effective, the warning and its message must be noticed by the user and presented at the appropriate time (Wogalter & Mayhorn, 2005).

However, not all debunking techniques are created equal. Facebook designed a technique in 2017 that was ineffective and pulled off the platform later that year (Moravec et al., 2020). Debunking techniques are ever-changing; since Kirchner and

Reuter's (2020) study was published, Facebook and Twitter have designed new debunking techniques and discontinued the use of others (Meta, n.d.; Twitter, 2021).

Research on message popup warnings is necessary because debunking techniques and the ways users share information are changing within the platforms.

### *Research Questions*

The overarching research question of this study was: Are message popup warnings effective for deterring the spread of misinformation?

To address this question, the study was guided by the following sub-questions:

RQ1: Regarding user engagement, to what extent do users read the full content contained in postings when comparing postings that are presented with or without message popup warnings?

RQ2: Regarding effectiveness, what is the user-choice response (actions or inactions) to postings when comparing postings that are presented with or without message popup warnings? Note: actions (1=liking, sharing, or commenting on a false posting) or inactions (0=not liking, sharing, or commenting on a false posting).

RQ3: Regarding credibility, how does the debunking technique of message popup warnings influence the users' credibility rating of the posting, compared to ratings given for postings without message popup warnings?

RQ4: How do users rate the usability and usefulness of message popup warnings as a debunking technique?

## Relevance and Significance

Thus far, message popup warnings used as a debunking technique have not been evaluated for the effectiveness of deterring the propagation of misinformation and for evaluating user engagement. Twitter has begun using message popup warnings when users attempt to share or like a posting that is either unread or contains disputed content (see Figure 1) (Ahmed, 2020; Twitter, 2021). Since popup warnings distract and slow down users (Abdulin & Billman, 2016), the technique holds promise for being an effective debunking technique. However, for any debunking technique to be effective, it must limit the spread of false news; message popup warnings need further examination of their effectiveness (Kirchner & Reuter, 2020). User engagement, usability (O'Brien & Toms, 2008), and usefulness may also determine each technique's level of effectiveness (Quesenbery, 2003).

**Figure 1**

*Example of Twitter's Message Popup Warning*



*Note.* From “Twitter is Expanding Its Warning Pop-ups for the Liking of Disputed Tweets Now,” by A. Ahmed, 2020, *Digital Information World*. Retrieved Oct 20, 2022, from <https://www.digitalinformationworld.com/2020/11/twitter-is-expanding-its-warning-pop.html#>. Copyright 2020 by digitalinformationworld.com. Reprinted with permission.

Researchers have studied the effectiveness of various strategies for users to devalue misinformation (Kirchner & Reuter, 2020; Tahir et al., 2021). The most effective techniques were also ones preferred by participants; individuals preferred transparent methods that warned of false news and included an explanation of the inaccuracies of the article (Kirchner & Reuter, 2020). Techniques, such as simply making the false news posting size smaller or hidden further in the newsfeed, was found less effective and less preferred, even though this was Facebook’s false news counter strategy in 2020 (Kirchner & Reuter, 2020). Regarding the technique of user comments, Colliander’s (2019) study suggested that other’s comments led to users siding with the majority opinion when adding comments to false news. Additionally, studies evaluating debunking techniques without detailed explanations disputing a posting’s truthfulness demonstrated limited effectiveness (Colliander, 2019; Fan et al., 2018).

### **Barriers and Issues**

Design challenges exist for researchers who want to evaluate interactive functionality such as having workable clickable interfaces and message popup functionality. This study presented similar design challenges in that a simulated environment was developed to foster an experiment that evaluates user responses and actions when debunking techniques (i.e., message popup warnings) were presented to

users in real-time. Developing mock-ups or functional prototypes for research purposes can be costly and time consuming. With limited interactive functionality of specific debunking methods such as message popup warnings and the use of full article content, research has been limited on evaluating effectiveness, usefulness, usability, and user engagement (Colliander, 2019; Kirchner & Reuter, 2020; Moravec et al., 2020; Pennycook et al., 2020).

Likewise, researchers face design challenges to present a realistic scenario of a typical posting. For example, some posting identifiers found in Facebook have been excluded to isolate the effects of the debunking techniques; poster name and source have been modified or masked, as to not influence the participant (Kirchner & Reuter, 2020; Moravec et al., 2020). Like these studies, this study modified poster name and the source to present a neutral or de-identified poster.

Although every effort was made to have a representative sample for the study, individuals who are prone to believe false news were not isolated and targeted for this study (Kirchner & Reuter, 2020). Furthermore, false news stories chosen were apolitical to avoid the divisive U.S. political landscape and to avoid the influence of the participants' previously held beliefs. Consequently, the participants may not have had equal interest in all the stories displayed in the questionnaire.

### **Assumptions, Limitations, and Delimitations**

#### *Assumptions*

This study was conducted as an experiment of an interactive scenario and online survey; therefore, participants needed access to a laptop, tablet, or smartphone to complete the tasks involved. Additionally, it was assumed that a significant portion of the

Wentworth Institute of Technology's population had both an active social media account and engaged in interactivity (liking, sharing, and commenting) on the social media platform.

### *Limitations*

This study faced two potential threats to validity: testing effects and selection bias. Within the ISS, participants were questioned about effectiveness before and after each scenario encountered that contained a message popup warning. However, posing the question prior to the treatment (message popup warning), participants may have been unduly influenced to pay more attention to the treatment itself (interactive testing effects) or may have chosen to keep their answers consistent between the pre- and post-tests (main testing effect) (Sekaran & Bougie, 2016). Furthermore, participants were restricted to the available pool of students at Wentworth Institute of Technology's Boston campus. As such, the results are subject to selection bias based on the use of a convenience sample and the results cannot be generalized (Sekaran & Bougie, 2016).

### *Delimitations*

This study did not include political news stories, which may have been less interesting to the participants, but avoided the need to evaluate political bias that would have been expected to ensue from using accurate and false political news stories in the study. Furthermore, the researcher did not distinguish confirmation bias or decision quality between the different user choice responses of liking, sharing, and commenting. Consequently, effectiveness was limited to binary interactions.



## Definition of Terms

The following terms are used within this study:

- Choice response - When confronted with a debunking technique, the user either took an action or took no action. An action of choice response was represented by the user's choice to like, share, or comment. An inaction of choice response was represented by the user's choice to not click on like, share, or comment.
- Credibility – Users evaluated the truthfulness or believability of social media postings; also known as the accuracy of the posting (Geeng et al., 2020).
- Debunking technique – Instead of displaying a false news story without any formatting, social media platforms display a false news story with message warning users of false content (Chan et al., 2017).
- Disinformation – Purposeful false information, also defined as fake news (Allcott & Gentzkow, 2017; Kirchner & Reuter, 2020).
- Effectiveness – For a debunking technique to be effective, the technique must help the user to understand that the message is false and stop the propagation of misinformation (Chan et al., 2017). This study evaluated user choice response (clicking or not clicking on like, share, or comment) as effectiveness.
- Engagement – On social media platforms, user engagement refers to click on like and share (Geeng et al., 2020) or reading and analyzing credibility before taking action (Kirchner & Reuter, 2020). This study refers to user engagement as a single type of user action, reading, and was evaluated on the extent users read the full content contained in the posting.

- Interaction – Otherwise known as direct manipulation, interaction is the user's direct communication with the system interface (O'Brien & Toms, 2008; Shneiderman, 1997).
- Fake news – Purposeful false information, also defined as disinformation (Allcott & Gentzkow, 2017; Kirchner & Reuter, 2020).
- Misinformation- False information without evaluating the poster's intent (Geeng et al., 2020).
- Usability – A usable system (or debunking technique) must be effective, efficient, engage the user, handle errors well, and is easy to learn (Quesenbery, 2003).
- Usefulness – A debunking method is useful if the users believe that the tool helps them make better decisions (Fan et al., 2018).
- Warning – A warning is used to alert users that the veracity of the posting is disputed (Pennycook et al., 2020).

### **List of Acronyms**

- HCI – Human Computer Interaction
- IRB – Institute Review Board
- ISS – Interactive Scenario and Survey
- NSU – Nova Southeastern University
- PUMS – Perceived Usefulness Measurement Scale
- SUS – System Usability Scale
- WIT – Wentworth Institute of Technology

## Summary

Controlling the spread of misinformation on social media platforms is challenging as it can propagate to others rapidly. Postings spread by user choice behaviors and algorithms, each affecting which postings will populate users' newsfeeds. Newsfeeds contain postings by like-minded people, filtered through filter-bubbles, where postings are created by, or are of interest to, the individual's contacts.

To combat the spread of misinformation, researchers focus on debunking, or decreasing, the credibility of false claims. Users tend to trust postings that are posted by their contacts or viewed multiple times. When false postings validate pre-existing beliefs, the false posting confirms the 'truthfulness' to the user through an effect called confirmation bias.

Various debunking techniques have been devised and applied to false news postings on social media with little impact on effectiveness. However, message popup warnings as a debunking technique are understudied to date. This study expands upon both Pennycook et al.'s (2020) and Kirchner and Reuter's (2020) research; Pennycook et al. (2020) focused on debunking techniques' effects on credibility ratings and Kirchner and Reuter (2020) suggested that users prefer warning-type debunking techniques rather than more subtle techniques. By assessing the effectiveness of message popup warnings, platforms can devise better debunking techniques to deter the spread of misinformation.

The overarching research question of this study was: Are message popup warnings effective for deterring the spread of misinformation? The research sub-questions focus on (RQ1) how message popup warnings affect user engagement (extent users read the full content contained in posting), (RQ2) how effective message popup

warnings are in deterring the spread of misinformation (user choice response of like, share, or comment), (RQ3) how users rate posting credibility, with and without message popup warnings, and (RQ4) how users rate the usability and usefulness of message popup warnings as a debunking technique.

## Chapter 2

### Review of Literature

#### Introduction

This chapter reviews literature on users' involvement with misinformation on social media platforms. Studies have delved into the constructs of credibility, user engagement, and debunking techniques' effectiveness to help to deter the spread of misinformation (Colliander, 2019; Kirchner & Reuter, 2020; Moravec et al., 2020; Pennycook et al., 2020). This chapter concludes with what is known and what is unknown from the research on message popup warnings.

#### How Misinformation Spreads

The postings that appear in an individual's social media newsfeed are not curated by individuals but by algorithms (DeVito, 2017). According to a patent filed by Facebook/Meta, all news articles and other content are placed in users' newsfeeds by algorithms (Sittig & Zuckerberg, 2013). Actions taken by social media users, such as liking, sharing, and commenting, positively influence those postings by appearing higher in other's newsfeeds (Ardèvol-Abreu et al., 2020). Although users may comment on a false posting to alert others of misinformation, this action may have the unintended effect of propagating the misinformation to other's newsfeeds (Ardèvol-Abreu et al., 2020; Jahanbakhsh et al., 2022). As an example, during the Zika crisis, false postings about the virus appeared more often than truthful postings on Facebook (Bode & Vraga, 2017).

On social media, news is filtered through a *filter-bubble* by the online contacts, or *friends*, of an individual (Bentley et al., 2019). As such, individuals see posts of the

others they follow, acting as a filter. Often, individuals follow other like-minded people, and information often is biased with the same opinions and political-leanings (Bentley et al., 2019). As such, social media news postings may contain bias; Twitter is known to display articles with bias and misinformation more than fact-checking content (Geeng et al., 2020).

### **Effectiveness**

Social media platforms must devise debunking techniques that limit the spread of misinformation (Moravec et al., 2020). Prior studies have evaluated techniques to minimize the influence of misinformation on social media platforms, once news is deemed to be false (Colliander, 2019; Kirchner & Reuter, 2020). However, to be effective, the debunking technique must first convince the user that a posting's content is untrue (Kirchner & Reuter, 2020; Moravec et al., 2020). Kirchner and Reuter's (2020) study suggested that users want to be warned of false postings and be provided explanations of why the content is disputed so that users can self-evaluate the questionable content.

Debunking effectiveness has multiple meanings. At a high level, a tool is effective when it does what it was meant to do (Quesenbery, 2003). In debunking literature, some researchers use the term *effectiveness* to evaluate the difference between the users' pre-existing beliefs and their beliefs after viewing a false article with its attached debunking technique (Fan et al., 2018; Kirchner & Reuter, 2020; Pennycook et al., 2020). Other studies define effectiveness as the user's choice response to the debunking technique (Moravec et al., 2020). Various studies have looked at metrics, such as reposting counts or average reposting time, to measure the effectiveness of posting propagation (Son,

2023; Spiro et al., 2012; Zeng et al., 2016). To evaluate the effectiveness of a debunking technique, a user must engage by reading the false posting and viewing the debunking technique, evaluate credibility, and then choose whether to take action, either share or not share the false posting (Moravec et al., 2020). Simply put, the debunking technique is evaluated on whether it affected the further spread of the false posting. Many social media users are passive users (approximately 90% of users); these individuals tend to not interact or create postings (Papakyriakopoulos & Goodman, 2022). Therefore, debunking techniques need to be effective with the active users, who create the majority of the platform's content (Papakyriakopoulos & Goodman, 2022).

Pre-existing beliefs and user's choice response can influence each other; Pennycook et al.'s (2020) study suggested that a user's perceived accuracy of an article does play a role in the user's choice response. Misinformation that confirms a user's pre-existing beliefs tends to be believed, even if tagged with a warning message (Papakyriakopoulos & Goodman, 2022). This phenomenon, known as *confirmation bias*, can influence a user even when the social media platform is attempting to debunk or discredit false information (Jun et al., 2017). When pre-existing beliefs are *confirmed* by others, individuals tend to believe and not question the credibility of the posting (Jun et al., 2017; Spezzano et al., 2021).

Some users also consider the belief system of their online contacts (Pennycook et al., 2020) or how much they trust the members in their online groups (Ma et al., 2019) before choosing a response, such as liking, sharing, or commenting. Some social media users will take action (like, share, or comment) on a posting to show alliance with a group or to obtain followers (Boehm, 1994). As an example, Ma et al. (2019) found that

individuals were more likely to take action (like or comment) in groups they trusted. Some users mistrust centralized fact-checking or warnings and prefer a peer-to-peer system to evaluate a posting's accuracy (Jahanbakhsh et al., 2022). Individuals are prone to trust family, friends, identity-based, education, and work groups more than interest and location-based groups (Ma et al., 2019).

Effectiveness of warning messages may wane over time; as individuals are exposed to the same warning message repeatedly, less attention is given and response to the message becomes habitual (Wash & Lampe, 2012; Wogalter & Mayhorn, 2005). When a user interacts with a warning message, the following steps are necessary for the message to be effective: 1. the user must notice the warning, 2. the user must mentally encode, or process, the message, 3. the user must comprehend the intended meaning of the warning, and 4. the user must comply with the warning (Rogers et al., 2000). To avoid habitual effects and ensure an individual's attention to a warning message, the presentation timing and warning's appearance should vary (Wogalter & Mayhorn, 2005).

Misinformation, which has been later recanted, can continue to live on in individual's memories and affect behavior (Ecker et al., 2010; Johnson & Seifert, 1994). Research suggests that repeated exposure to misinformation strengthens an individual's recall and the post's credibility (Eakin et al., 2003). Some studies have observed a *backfire effect*; individuals' beliefs in the misinformation increased with the appearance of a warning message (Ardèvol-Abreu et al., 2020; Nyhan & Reifler, 2010). Other studies have disputed the presence of a such a backfire effect (Papakyriakopoulos & Goodman, 2022; Wood & Porter, 2019). Nevertheless, the timing of the warning message and the message itself is critical; specific warning messages and warnings delivered at the same



time as the misinformation are more effective in reducing the influence of misinformation (Ecker et al., 2010).

### **User Engagement**

Moreover, the level of user engagement may impact the effectiveness of the debunking technique (Kirchner & Reuter, 2020). As defined by Quesenbery (2003), engagement occurs when the user is pulled into or engrossed in their task. Engagement in this context refers to involvement. On social media, a user can demonstrate engagement or interest in a news story by reading or other choice responses with the posting (Moravec et al., 2020).

In Argo and Main's (2004) study, five engagement elements were assessed to determine effectiveness of warnings: attention, reading and comprehension, recall, judgments, and behavioral compliance. Argo and Main (2004) posited that individuals must cognitively process information (thus be involved) to fully engage in decision making. Furthermore, individuals who are confident in their pre-existing knowledge are less likely to engage in actions to obtain further information (Beatty & Smith, 1987).

Involvement is intertwined with credibility assessment; for individuals to embark on credibility decision making, the posting must first get their attention (Geeng et al., 2020). In Geeng et al.'s (2020) study, short postings (i.e., short text or memes) and postings with many likes or comments tended to gain users' attention. The ability to identify false postings increases when users read the full text of the posting (Spezzano et al., 2021). If the user ignores the debunking technique, by scrolling past or clicking without reading, this action could directly affect other aspects of the user experience, such as effectiveness (Barnum, 2021).

## Credibility

How do users know to question the accuracy and credibility of a news posting? Pennycook et al. (2020) posit that postings without a warning tag may be perceived to be truthful, even though a posting without a warning may fall into one of two categories: 1. posting has been checked and verified as accurate or 2. the posting may have not yet been assessed for accuracy. As such, this *implied truth effect* may make a posting appear to be more credible since it has no warning (Pennycook et al., 2020). However, there is some dispute whether it exists; in subsequent studies, others have been unable to replicate the appearance of an implied truth effect (Clayton et al., 2020; Kirchner & Reuter, 2020).

When a user finds a social media posting credible or disbelieves a disputed content warning, specific user behavior can be observed, such as reposting or liking the posting (Ginsca et al., 2015). Credibility is often judged by the posting source's expertise and trustworthiness (Pasquetto et al., 2022; Vraga & Bode, 2018). Gao et al.'s (2018) study evaluated user perception of *The Wall Street Journal's* labels of credibility (disputed vs. trustworthy) and political stance (conservative vs. liberal). They found that credibility labeling by itself was not an effective technique to debunk misinformation (Gao et al., 2018).

Assisting users to recognize misinformation may help to limit the spread of false news. Research has led to questions about the users' attention to postings and warnings prior to taking an action (liking, sharing, or commenting) (Pennycook et al., 2020). Some users simply forget to evaluate the new story for accuracy, but when reminded to do so, the resulting action shows a decrease in false news propagation (Pennycook et al., 2021). Additionally, individuals consider other factors such as existing user comments

(Colliander, 2019), the poster's name, the news source, and fact-checking websites when evaluating the credibility of a posting (Geeng et al., 2020). Although, users may not take the time to assess credibility, often making quick decisions to share content based on the posting's headline, image, news source, or poster's name (Spezzano et al., 2021).

Individuals often depend on the poster to vet the news posting for credibility. Credibility is derived from an individual's belief in the expertise and trustworthiness of the poster (Pasquetto et al., 2022). Additionally, credibility may include a third element, such as goodwill or looking out for the recipient's best interests (McCroskey & Teven, 1999). Close relationships, or *strong ties*, to a poster positively influences the likelihood of an individual to trust that the news posting is accurate and credible (Pasquetto et al., 2022). Even though the poster may not be a subject-matter expert, individuals often believe that the poster is looking out for their contacts' best interests and would not post false information (McCroskey & Teven, 1999; Pasquetto et al., 2022).

Perceived consensus makes misinformation more believable (Papakyriakopoulos & Goodman, 2022). Social media displays metrics of actions (likes, shares, and comments) and these metrics give the perception to users that others hold the same belief as the posting (Avram et al., 2020). Individuals will hold onto unpopular beliefs, or misinformed beliefs, when they have 'proof' that others believe the same (Hills, 2019).

As false news is repeated, individuals may experience a *validity effect*. The term, validity effect, describes the phenomenon of individuals believing misinformation simply because it is read numerous times (Spezzano et al., 2021). Boehm (1994) found the validity effect present for participants who found statements true, false, or neutral prior to repeated exposure.

## Usefulness

In order for an individual to use a tool, the tool must be considered useful or helpful in completing the task at hand (Davis, 1989). An agreed-upon definition of usefulness has been evasive to HCI researchers since the term must be applied in-context to the technology in question (MacDonald & Atwood, 2014). However, when applied to debunking techniques, usefulness implies that the debunking technique must be helpful and valuable for decision making (Fan et al., 2018). MacDonald and Atwood (2014) suggested that usefulness encompasses both practical and hedonic elements; the tool must do what is expected and be visually attractive to use. Although usability and usefulness are closely related, a tool cannot be fully evaluated without considering its usefulness (MacDonald & Atwood, 2014).

For a debunking technique to be useful, the message should influence the user's decision for a choice response of action or inaction (Fan et al., 2018). As such, the specific text message in the warning should be clear and easily understood (Fan et al., 2018; Wogalter & Mayhorn, 2005). There are two categories of misinformation warning messages: warnings that contain a corrective statement (state that the posting's content is disputed) and warnings that provide additional information about the post's subject matter (Papakyriakopoulos & Goodman, 2022). Fan et al. (2018) found that detailed warning messages had a positive influence on perceived usefulness and individuals believed they were more empowered to make better decisions. Papakyriakopoulos and Goodman (2022) found that individuals were more likely to make an actionable user choice response (like, share, or comment) when encountering a generalized warning

message, rather than when encountering a warning message containing specific details about the misinformation.

## **Usability**

There are five dimensions to fully describe a tool's usability. Usability was first defined to include the qualities of learnability, efficiency, memorability, a low rate and good handling of errors, and satisfaction (Nielsen, 1994). Quesenbery (2003) refined the usability dimensions to describe systems that are effective, efficient, engaging, error tolerant, and easy to learn. A user's perceived usability of a system measures both a user's affective and cognitive responses (O'Brien et al., 2018).

Usability problems with the system and external distractions have resulted in user disengagement (O'Brien & Toms, 2008). If a task cannot be completed (effective), cannot be done with reasonable effort and time (efficient), does not have a satisfying experience (engaging), is not free of errors or handle error well (error tolerant), and is complex to use (easy to learn), the tool is not usable (Quesenbery, 2003). O'Brien and Tom (2008) posited that usability and poor tool design may act as an impediment to user engagement. Engaging tools have a common characteristic of usability, however, not all tools that are usable are engaging (O'Brien & Toms, 2008).

On social media platforms, user interface issues persist, as individuals grapple with distinguishing truthful postings from false postings. In Geeng et al.'s (2020) study, participants had difficulty separating the user from the news source they shared or liked, as the platform made it appear that the two entities (poster and news source) were associated with each other. Additionally, comments are another source of user confusion,

as some users rely on comments as a debunking or fact-checking method (Colliander, 2019; Geeng et al., 2020).

Users are accustomed to how popup warnings function, thus all the instruction necessary to use should be in the warning's embedded message (Wash & Lampe, 2012). However, as in the aforementioned topic of habitual behaviors, users do grow accustomed to message popup warnings and may click on the popup without cognitive awareness (Wash & Lampe, 2012; Wogalter & Mayhorn, 2005).

### **Background and Previous Studies**

Previous studies have limited the specific debunking techniques studied or the actions possible by participants. Additionally, Pennycook et al. (2020) suggested that future debunking studies should not be limited to story headlines, rather user engagement and debunking effectiveness should be evaluated after a user reads the false news story. Kirchner and Reuter (2020) eliminated the debunking technique of message popup warnings in their study, due to difficulties in simulating its interactive nature. Researchers have avoided assessing user clicks and user choice responses, opting for studying static debunking techniques and evaluating various constructs through flat questionnaires or interviews (Colliander, 2019; Fan et al., 2018; Kirchner & Reuter, 2020; Moravec et al., 2020; Pennycook et al., 2020). However, when mouse clicks are used as data to capture user interactions in interactive scenarios, they can lead to meaningful insights when comparing actual user behavior (actions) to user perceptions (Geeng et al., 2020).

Existing research has focused on participants evaluating postings on headlines, rather than enabling participants to evaluate an article on its full content (Colliander, 2019; Fan et al., 2018; Kirchner & Reuter, 2020; Moravec et al., 2020; Pennycook et al.,

2020). On the actual Facebook platform, users can evaluate articles on a plethora of posting identifiers, such as headline, article content, poster, and various debunking techniques (Kirchner & Reuter, 2020; Moravec et al., 2020). Previous studies have modified or stripped these posting identifiers as to not muddy the influence of other variables (Kirchner & Reuter, 2020; Moravec et al., 2020).

### **Summary of What is Known and Unknown About the Research**

The interactive nature of message popup warnings has resulted in this debunking technique to be understudied (Kirchner & Reuter, 2020). In Twitter, users only encounter a popup warning after the user's choice response of attempting to take action (like, retweet, quote tweet, or comment) and will not encounter the popup warning if they choose to not take action (Papakyriakopoulos & Goodman, 2022). While Twitter uses both static text warnings and message popup warnings to warn users of spreading false content, Facebook does not use static text warnings in concert with message popup warnings (Kirchner & Reuter, 2020; Twitter, 2021). Each social network platform uses different messaging in the misinformation warnings. The variation of text in these messages can influence the user differently (Fan et al., 2018).

To date, few studies have included an interactive scenario and survey. As message popup warnings can only be tested in an interactive environment, it is unknown how effective popup warnings are in deterring the spread of misinformation. Furthermore, with the use of an interactive scenario, credibility, usefulness, and usability of message popup warnings can be assessed. As the influence of the full article's text in the posting has been understudied, it is also unknown whether the availability of the full article may promote further user engagement or assist users in their choice response.

## Summary

Misinformation spreads quickly on social media platforms. Platforms utilize algorithms to populate newsfeeds (Sittig & Zuckerberg, 2013). Algorithms are influenced by user actions, such as liking, sharing, or commenting, and each of these actions cause the posting to appear higher in other's newsfeeds (Ardèvol-Abreu et al., 2020; Jahanbakhsh et al., 2022).

The spread of misinformation is affected by user engagement and the perceived credibility of the posting (Kirchner & Reuter, 2020; Pennycook et al., 2020). An individual judges posting credibility with their pre-existing beliefs, group alliances, perceived group consensus, and trust in the poster (Ma et al., 2019; Papakyriakopoulos & Goodman, 2022; Pennycook et al., 2020). Often, users may not take the time to assess credibility, making quick decisions to share content based on the posting's headline, image, news source, or poster's name (Geeng et al., 2020). Additionally, the level of user engagement may impact whether misinformation spreads further (Kirchner & Reuter, 2020). To fully be engaged in decision making for choice responses, users must cognitively process information by reading and comprehending the posting (Argo & Main, 2004).

Debunking techniques should be perceived by users as both useful and usable to be effective. To be considered useful, the debunking technique must be helpful and valuable for decision making, and should influence the user's decision for a choice response of action or inaction (Fan et al., 2018). To be considered usable, the users must find the debunking technique effective, efficient, engaging, error tolerant, and be easy to learn (Quesenbery, 2003).



To combat the spread of misinformation, social media platforms have developed various debunking techniques to mark postings as false. An effective debunking technique is one that affects the user's choice response, i.e., whether to take action or no action (click or not click on like, share, or comment) (Moravec et al., 2020). Newer debunking techniques, such as message popup warnings, are understudied. It is unknown how effective message popup warnings are for deterring the spread of misinformation, or how it will affect user engagement or posting credibility. Additionally, research is needed to assess the influence of providing the user with the full article contents.

## Chapter 3

### Methodology

#### Overview

This chapter outlines the research methodology, design, and procedures for addressing the following research question: Are message popup warnings effective for deterring the spread of misinformation? Specifically, this study sought to evaluate the following research questions:

- RQ1: Regarding user engagement, to what extent do users read the full content contained in postings when comparing postings that are presented with or without message popup warnings?
- RQ2: Regarding effectiveness, what is the user-choice response (actions or inactions) to postings when comparing postings that are presented with or without message popup warnings? Note: actions (1=liking, sharing, or commenting on a false posting) or inactions (0=not liking, sharing, or commenting on a false posting).
- RQ3: Regarding credibility, how does the debunking technique of message popup warnings influence the users' credibility rating of the posting, compared to ratings given for postings without message popup warnings?
- RQ4: How do users rate the usability and usefulness of message popup warnings as a debunking technique?

To help answer research questions 1-3, the following hypotheses were developed:

H1: Regarding user engagement, there will be no significant differences when comparing the extent users read the full content contained in postings that are presented with or without message popup warnings.

H2: Regarding effectiveness, there will be no significant differences in user choice response (actions or inactions) to postings, when comparing postings that are presented with or without message popup warnings.

H3: Regarding credibility, there will be no significant differences in how users rate the credibility of the posting when receiving the message popup warning, compared to ratings given for postings without message popup warnings.

### **Study Design**

To address the above research questions, the researcher employed a between-subjects experimental study design. The study had two randomly assigned groups of participants, one as a treatment group and the other as a control group. Specifically, the study consisted of individuals viewing postings that simulate Facebook's platform. The experiment was scenario-based, showing interactive news postings and posing questions after every news article. In the treatment group, the Interactive Scenario and Survey (ISS) contained a newsfeed with both true and false postings; the misinformation posts first appeared without any warnings and after deciding to take action (click on like, share, or comment) or take no action (not click on like, share, or comment), these posts were tagged with a message popup warning. Conversely, the control group received the same newsfeed postings but did not receive the message popup warnings with false postings.

Table 1 lists the constructs and evaluation techniques applied in the study by research question.

**Table 1**

*Constructs and Evaluation Technique by Research Question*

<b>RQ</b>	<b>Construct</b>	<b>Evaluation Technique</b>
1	Engagement (extent users read full content of posting)	Interactive Scenario and Survey (ISS)
2	Effectiveness (user's choice response – actions or inactions of liking, sharing, or commenting)	ISS
3	Credibility	ISS
4	Usefulness	Perceived Usefulness Measurement Scale (PUMS)
4	Usability	System Usability Scale (SUS)

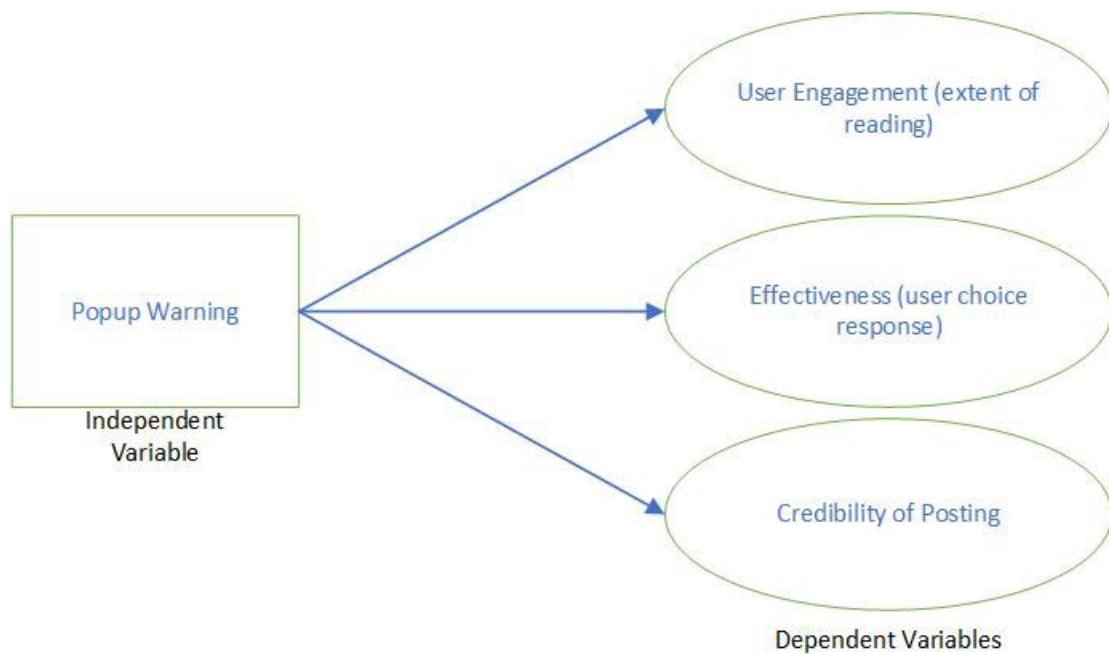
A pilot test was performed prior to the ISS rollout to test the validity of the instrument (Lazar et al., 2017). After validation, participants were recruited from the student population at Wentworth Institute of Technology (WIT) in Boston, MA. Quantitative data was captured from the Interactive Scenario and Survey as the user navigated through the Facebook-like platform. Prior to the treatment of message popup warnings, users were asked if they want to take an action (click like, share, or comment) or choose not to take an action (not click like, share, or comment). After the appearance of message popup warnings, embedded questions related to the user's engagement (extent users read the full content contained in posting) with the article, an assessment of the posting's credibility, and provided an opportunity to evaluate the message popup warning's effectiveness (the user's choice response of action or inaction.) The questions

related to both credibility and effectiveness of the message popup warnings were adapted from Pennycook et al.'s (2020) study.

To evaluate the usefulness and usability of the message popup warnings, this study used two validated and widely used instruments, the Perceived Usefulness Measurement Scale (PUMS) and the System Usability Scale (SUS). Users in the treatment group were prompted to complete both instruments after receiving the message popup warning. Users in the control group were not administered the PUMS and SUS as they did not experience the message popup warnings.

### *Study Variables*

This study focused on the effects of the independent variable, message popup warnings, and on the three dependent variables: user engagement (extent users read the full content contained in posting), effectiveness (user choice response), and credibility of the posting. User engagement was measured by asking the user if they read the message content. The effectiveness of the message popup warning was measured by the user choice response of action (liking, sharing, or commenting) or inaction (not liking, sharing, or commenting). Additionally, credibility was measured by asking users to rate the credibility of the posting. Figure 2 illustrates the relationship of variables.

**Figure 2***Study Variables**Treatment*

A between-subject experimental design was used to evaluate aspects of user engagement (extent users read the full content contained in the posting), effectiveness (user choice response), credibility, usability, and usefulness as users are presented with true and false news postings. Like Pennycook et al.'s (2020) study, this study compared a control group to a single treatment group. As in Pennycook et al. (2020)'s study, the participants were given 24 postings, consisting of an equal mix of true and false randomly ordered postings. The control group saw no difference in the appearance of any of the true or false news articles; no warnings appeared to indicate disputed content. For the treatment group, the group evaluated true and false news postings; half of the false postings (six postings) had message popup warnings appear after the user took an action (like, share, or comment) or no action (not like, share, or comment) with the posting.

Unlike Pennycook et al. (2020) and Kirchner and Reuter (2020), the treatment group received effectiveness questions twice, before and after receiving a message popup warning. For the postings without a message popup warning, the effectiveness question was only presented once.

Comparatively, in the studies by Pennycook et al. (2020) and Kirchner and Reuter (2020), participants were asked to assess accuracy of article headlines, not the full article. To expand upon Pennycook et al.'s (2020) research, the articles' full text, which included as much as a page or more of content, was available to the participants. The full article may have affected participants' credibility judgment, by providing additional information beyond headline, picture, and poster's name.

As on any social media platform, not all false news can be detected (Pennycook et al., 2020). Therefore, false news was not initially tagged as misinformation in this study. Due to the nature of popups, the message popup warning only displayed after the user answered the effectiveness (pre-treatment) question, which recorded the user's likelihood to take action (like, share, or comment) or take no action (not like, share or comment) without a message popup warning present. To ensure that the false news stories used were untrue, Snopes.com was consulted. Snopes.com is a reputable U.S. based, third-party fact-checking website, commonly used by concerned citizens and by researchers (Geeng et al., 2020; Kirchner & Reuter, 2020; Pennycook et al., 2020; *Snopes is the internet's definitive fact-checking resources*, n.d.). As in Pennycook et al. (2020), participants were instructed that only some of the articles have been fact-checked, to add doubt to the credibility of the unmarked postings.

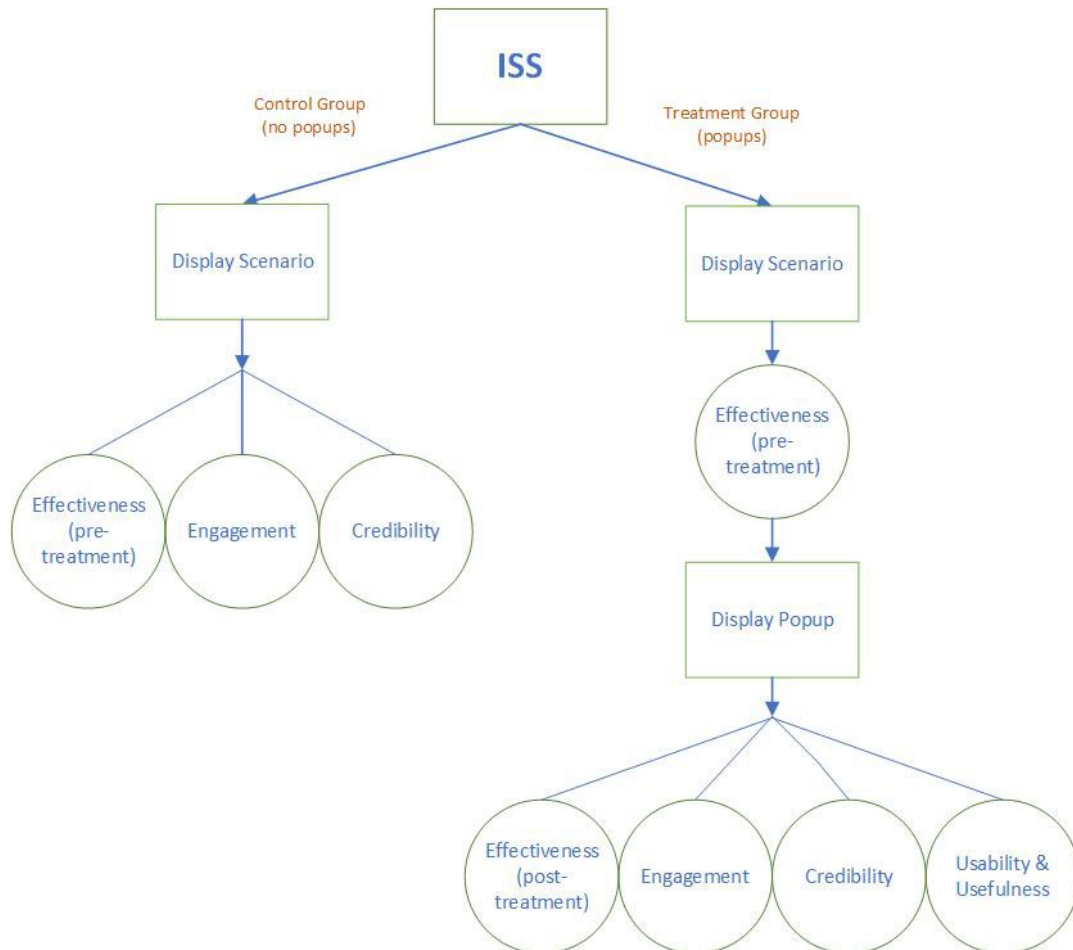
To address the overarching research question of “Are message popup warnings effective for deterring the spread of misinformation?”, this study evaluated the constructs of user engagement, effectiveness, and credibility between the treatment group (presented with message popup warnings) and the control group (presented with no warnings). On the real Facebook platform, the actions of liking, sharing, or commenting trigger the posting to spread to their contacts’ newsfeeds (Kirchner & Reuter, 2020). To measure the effectiveness of the popup warnings, the treatment group had to first decide on their user choice response (like, share, comment, or no action) prior to viewing the popup. The treatment group participants received the effectiveness (pre-treatment) question prior to the popup warning message, then received the effectiveness (post-treatment) question after viewing the popup. The questions required users to consider liking, sharing, or commenting both before and after the popup. Although the effectiveness questions (pre-treatment and post-treatment) were posed twice and were almost identical contextually, the responses were keyed as separate data points to assess the change in user choice response. Additionally, all scenarios queried the level of user engagement (extent users read the full content contained in posting) and credibility. See Table 2 for questions in the ISS.



**Table 2***Interactive Scenario and Survey (ISS): Questions to Display after Each Article*

<b>RQ</b>	<b>Construct</b>	<b>Question</b>	<b>Source</b>
1	Engagement (extent users read full content of posting)	Did you read the previous article? (4-pt Likert scale: No, did not read; Read little of article; Read somewhat; Yes, read fully)	
2	Effectiveness (pre-treatment) (user choice response)	If you were to see this article on Facebook or another social media platform, would you consider liking, sharing, or commenting on it? (Multiple choice: Like, Share, Comment, Would not click on above choices)	Adapted from Pennycook et al. (2020)
2	Effectiveness (post-treatment) (user choice response)	Regarding the previous posting, would you reconsider liking, sharing, or commenting on it? (Multiple choice: Like, Share, Comment, Would not click on above choices)	Adapted from Pennycook et al. (2020)
3	Credibility	To the best of your knowledge, how accurate is this article ? (4-pt Likert scale: Not at all accurate, Not very accurate, Somewhat accurate, Very accurate)	Adapted from Pennycook et al. (2020)

Since message popup warnings are interactive in nature, users did not view these warnings until an action (user choice response) was taken (see Figure 3). Prior to the choice response, users saw an interactive Facebook-like scenario with a link to access the full text of the article. Once the user answered the effectiveness (pre-treatment) question, the message popup warning appeared. Figure 3 displays the logic of the appearance of the warnings and question appearance.

**Figure 3***Flowchart of Experiment Questions*

*Note.* Only the treatment group was presented with a message popup warning (for half of the false scenarios) after answering the first effectiveness (pre-treatment) question.

In addition, participants in the treatment group completed an evaluation of message popup warnings for usefulness and usability, using the PUMS and SUS instruments. The control group did not view message popup warnings on the false postings, thus was unable to evaluate this debunking technique for usefulness and usability. Mouse clicks from both groups were captured through the interactive scenarios to compare user choice responses (actions).

## Procedures

To prepare and run the experiment was a four-step process, as depicted in Figure 4. To begin, an expert review and pilot test of the ISS was first conducted to ensure that the test questions and instructions were clear, the layout was appropriate, and the interface was functional (Lazar et al., 2017). Once the ISS was in its final form, participants were recruited from WIT's campus (see Appendix F). Participants were pre-screened for eligibility to participate in this study. Qualified participants were randomly assigned to either the control or treatment group. Within each group, the 24 true and false postings were presented in a random order. After the participants completed the ISS, the treatment group was asked to take the PUMS and SUS instruments, to evaluate the usability and usefulness of the message popup warnings.

**Figure 4**

*Procedure Diagram*



### Expert Review and Pilot Test to Finalize ISS

The study began with an expert review and pilot test of the ISS (see Appendix A). First, the ISS was reviewed by two research experts to ensure that the questions and instructions were clear. To qualify as experts, these individuals had technical expertise in instrument design and test measurement. For convenience purposes, colleagues of the researcher, who are skilled in creating instruments to evaluate data, were recruited. These

experts determined the validity of the survey, confirming that the questions and possible answers were measuring the constructs appropriately (Lazar et al., 2017). The expert reviewers were given a copy of the research questions to evaluate the validity of the constructs embedded in the survey. At the end of the ISS, the experts were asked about validity concerns, interface usability, and system issues. Different screen sizes (laptop, tablet, and phone) were tested for compatibility issues. The experts were asked to provide written feedback about the ISS to highlight areas of concern. Additionally, the experts helped to determine the approximate time frame for participant survey completion. (See Table 3 for Pilot Test questions.)

**Table 3**

*Pilot Test Questions*

<b>Audience</b>	<b>Construct to Test</b>	<b>Pilot Test Questions</b>
Expert	Validity	Will the survey questions generate data to answer the research questions?
Experts, Pilot Participants	Interface Usability	Were the instructions clear? Are the scenarios readable? Was the survey straightforward and intuitive? Does the survey conform to basic accessibility standards and is technically and functionally usable for people with disabilities?
Experts, Pilot Participants	System Issues	Did hyperlinks open quickly? Did the survey move efficiently from screen to screen? Any other issues not specifically mentioned?
Experts, Pilot Participants	Time on Task	Did any of the questions take a particularly long time to answer? How long did it take you to complete the survey?

Two experts reviewed the ISS between September 14 and September 26, 2023 and found the ISS to be both reliable and valid for measuring the outlined constructs. The

experts made suggestions for user clarification which were implemented, such as adding the text “I consent to participate in this study” to the button on the Informed Consent form, listing multiple types of social media in the directions (Facebook, Instagram, TikTok, LinkedIn), removal of arrow on page when presented with the control or treatment survey link, and replacing the word “tool” with “popup” on the SUS instrument. One expert also found a bug where one of the false news articles did not open in a separate window; this bug was rectified. The experts took 20 to 30+ minutes to complete the survey. Both experts found the instructions to be clear, hyperlinks opened quickly, and that the instrument was valid and reliable for answering the research questions. The “thank you” page was well-received by one expert, commenting that she appreciated the list of false news articles at the end of the survey.

Once the ISS was modified as recommended by the experts’ feedback, 11 pilot test participants were recruited (see Participants, Recruitment, and Selection) to complete the survey and provide verbal feedback through a thinking aloud process. These pilot participants were recruited from the representative population, however, they were not included in the formal study (Lazar et al., 2017). From September 29 to October 3, 2023, pilot participants tested both the treatment and control versions of the ISS. The pilot participants verified that the survey functionality worked as intended (Barnum, 2021; Lazar et al., 2017). However, the pilot participants, through the thinking aloud process, made note of some additional bugs, concerns regarding the choice of a Facebook interface, and confusing instructions. Additionally, the researcher discovered numeric coding issues after the pilot surveys were completed. Consequently, the survey instructions were modified to include a picture of a sample news posting, highlighting

where the user should click for their choice response (see Figure 5). Additionally, the news posting images were cropped to make the postings look more generic, as some pilot participants commented about preferences for other non-Facebook social media platforms (see Figure 6). Furthermore, the pilot participants helped determine that the length of the ISS was acceptable (24 scenarios). (See Table 3 for Pilot Test questions.) After a pilot test of 10 participants, the survey tool was again modified. On October 9, 2023, an 11<sup>th</sup> pilot tester was recruited to ensure that the modified survey worked as expected. Each pilot test participant was compensated with a \$20 Amazon gift card.


## Figure 5

### *Revised ISS Instructions*

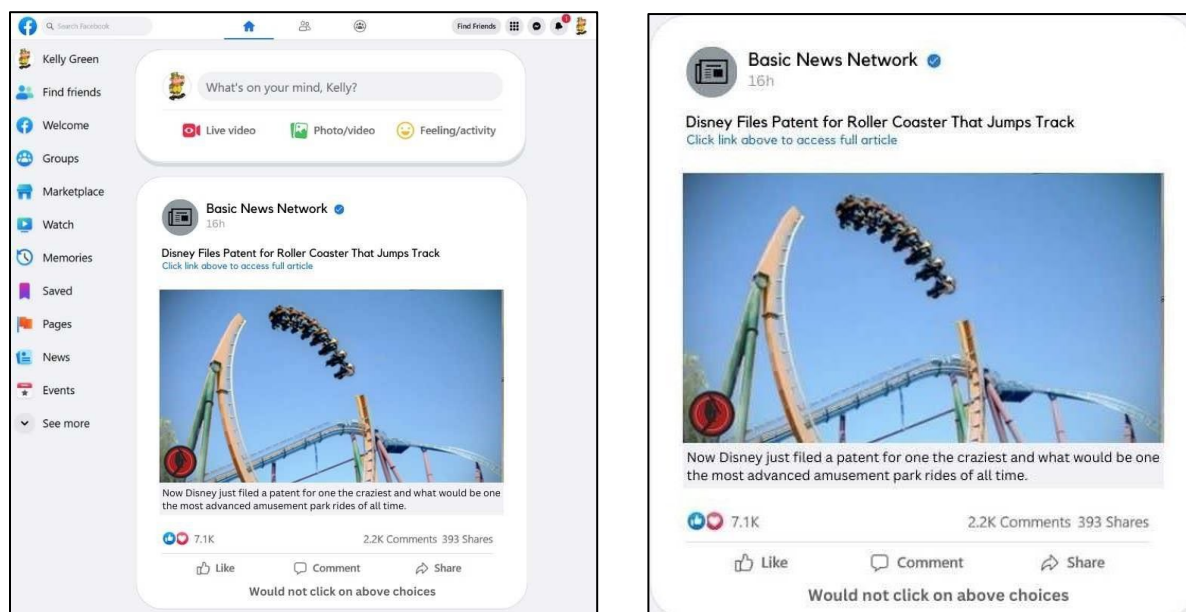
Instructions: You will be presented with a series of news stories from 2022 and 2023 (24 in total.) We are interested in two things: (1) whether you think the articles are accurate or not, and (2) whether you would be willing to like, share, or write comments on the article on **any** social media platform, such as Facebook, Instagram, TikTok, LinkedIn, etc. Fact-checking has not been performed on all of the articles in this survey.

If you would like to read the full article, click on the [link](#) above the posting.

Below is an **example** of a posting. At the bottom of each social media posting, click on ONE of the buttons ("Like", "Comment", "Share", or "Would Not Click on Above Choices"):



The screenshot shows a social media post from 'Basic News Network' with 1.6k likes. The headline is 'Unbelievable - A missing plane has returned after 37 years.' with a link to the full article. The image shows a Pan Am flight 914 aircraft. Below the image, the text reads: 'Unbelievable - A Missing Plane Has Landed After 37 Years Pan Am Flight 914 took off from an airport in New York City bound for Miami, Florida. Howeve...'. The post has 7.1K likes, 2.2K comments, and 393 shares. At the bottom, there are buttons for 'Like', 'Comment', 'Share', and 'Would not click on above choices'. A large black arrow points from the text above to the 'Would not click on above choices' button.

**Figure 6***Comparison of Original and Revised Posting Format*

*Note.* Original posting is on left and revised (generic) posting is on right.

### **Participants, Recruitment, and Selection**

Although the population of Facebook users is in the billions (Dixon, 2023a), not all individuals have an account or participate as interactive users. As in Pennycook et al.'s (2021) study, potential participants were pre-screened to ensure that they were active on a social media platform. Since this study was evaluating the effectiveness of a debunking technique to deter the spread of misinformation, it was important to ensure that participants were not mere observers on the platform but interactive users. Interactive users are defined as individuals who like, share, or comment on postings that appear in their newsfeed. Table 4 contains the participant pre-screening questions. WIT's students were recruited to participate in the experiment; however, participants were required answer in the affirmative to both screening questions to be included in the study's sample. If eligible, the participants received the informed consent form and then were

randomly directed to either the control or treatment ISS, beginning with survey instructions. If ineligible, the participants were thanked for participating and the survey ended.

**Table 4**

*Participant Pre-screening Questions*

Question	Source
Do you have a social media account? (Binary answer: Yes or No)	Pennycook et al., 2021
Have you liked, shared, or commented on a news article on social media? (Binary answer: Yes or No)	

Sample size for studies have varied in literature; studies have used from 44 to over 5,000 participants for their debunking surveys (Colliander, 2019; Fan et al., 2018; Kirchner & Reuter, 2020; Moravec et al., 2020; Pennycook et al., 2020). According to Salkind (2017), studies with multiple groups should have at least 30 participants in each subgroup. Since the population that uses Facebook is estimated to be almost three billion users (*Number of monthly active Facebook users worldwide as of 3rd quarter 2022*, 2022), a smaller population must be studied. This study recruited participants from the Wentworth Institute of Technology's population of approximately 4,000 students. Only WIT students were asked to participate and self-selected, thus, qualifying as a convenience sample (Sekaran & Bougie, 2016). Human-Computer Interaction (HCI) research commonly uses convenience samples, as often researchers are unable to claim a representative sample (Lazar et al., 2017). For HCI survey research, it has been suggested that for a population of 4,000, a sample of approximately 90 participants will achieve a confidence level of 95% with a 10% margin of error (Müller et al., 2014). From October



10 to October 13, 2023, 206 participants started the survey, resulting in 109 participants ( $N = 109$ ) who were eligible and completed the survey. The participants in this sample were randomly assigned to the control or treatment group, using Qualtrics' randomization tool, with a total of 54 participants in the control group and 55 participants in the treatment group (Müller et al., 2014). To aid in the recruitment of participants, individuals were compensated with a \$10 Amazon gift card.

### **Instrumentation Setup, Development, and Quality Control**

#### *Environment and Setup for the Study*

This experiment functioned as an interactive scenario and survey (ISS) of a simulated Facebook newsfeed. (See Appendix A for screenshots of the interactive survey. Appendix C lists the sources for all the photographs and drawings used in the ISS.) Like Colliander (2019), Pennycook et al. (2021), and Moravec et al. (2020), Qualtrics was used as the survey platform, and Canva images were added. Qualtrics software provided the interactive functionality needed for hotspot and popup functionality and for mouse click captures of user actions. Additionally, Qualtrics provided functionality to present postings in both a random order and randomly assign participants to control and treatment groups. Wix was used to host the full content contained in the posting, appearing in the ISS as a popup window (Wix, n.d.). Canva was used to develop screen mockups containing true and false news on a simulated Facebook platform.

The ISS was administered in-person for the pilot test and online for survey participants. As the ISS was an online survey, the participants determined the place where the survey was taken, and the device used (laptop, tablet, or smartphone). Conversely, in the pilot phase, the participants completed the survey in a conference

room at Wentworth Institute of Technology so that the researcher could observe issues and receive timely feedback about the survey (Lazar et al., 2017). To obtain this feedback, pilot testers participated in a thinking aloud process; these participants were asked verbalize thoughts about the survey as they moved from screen-to-screen (Nielsen, 2012). The users' thoughts were recorded manually by the researcher (See Appendix G for the Think-Aloud Checklist and Form used).

### *Instrument Development*

Prior to the start of the ISS, participants received instructions. In line with Pennycook et al. (2020), instructions informed participants that only some of the articles have been fact-checked for accuracy, although all were fact-checked using Snopes.com (see Appendix A). As in Pennycook et al.'s (2020) and Kirchner and Reuter's (2020) studies, participants evaluated the credibility of the posts encountered during the survey.

Questions for engagement, effectiveness, and credibility (pre-treatment and post-treatment, variation based on control or treatment group) were posed to the participants after every news article (see Table 2) and were based on questions used in both Pennycook et al.'s (2020) and Kirchner and Reuter's (2020) studies. As in previous studies, true and false postings were presented to users in a random order. These questions were used to evaluate the article's credibility and the effectiveness of the debunking technique. Like Pennycook et al. (2020), the questions regarding credibility and effectiveness were posed after the disputed warning message. Message popup warnings may have affected the user's credibility rating, but have previously failed to influence user choice response (Pennycook et al., 2020).

To understand the level of user engagement, one question was added to ascertain how much of the article was read. This study expanded on Pennycook et al.'s (2020) and Kirchner and Reuter's (2020) studies by providing the full text of the article (when requested by user). The full article gave the users more information to judge credibility. Furthermore, the effectiveness of message popup warnings may have been influenced by whether the warning message was read or ignored.

Sample screenshots from the ISS are shown in Figures 7 and 8. For both the control and treatment groups, Figure 7 depicts a sample false news posting in a participant's newsfeed, prior to the user taking any action. After the user answered the effectiveness (pre-treatment) question, Figure 8 shows the message popup warning that users encountered in the treatment group only.

### Figure 7

#### *False News Posting in Newsfeed Prior to Message Popup Warning*



*Note.* Photograph and drawing from “Elon Musk Shocks the World With Plans for New Twitter HQ on Alcatraz Island” by Architizer Editors, 2023, Architizer.com,

(<https://architizer.com/blog/inspiration/stories/elon-musk-twitter-hq-alcatraz-island.>)

Copyright 2023 by Architizer, Inc.

## Figure 8

*False News Posting with Message Popup Warning*



*Note.* Photograph and drawing from “Elon Musk Shocks the World With Plans for New Twitter HQ on Alcatraz Island” by Architizer Editors, 2023, Architizer.com,

(<https://architizer.com/blog/inspiration/stories/elon-musk-twitter-hq-alcatraz-island.>)

Copyright 2023 by Architizer, Inc.

False postings without an embedded message popup warning did not contain any markings, flags, or warnings. In the interactive Facebook scenarios, as in reality, false news is not detected immediately (Pennycook et al., 2020). In 2017, Facebook took three

days to detect false news (Pennycook et al., 2020). Thus, users must rely on their own credibility judgment.

The usefulness of message popup warnings was assessed using Davis' (1989) Perceived Usefulness Measurement Scale (PUMS) (see Table 5). The participants used a 7-point Likert scale to assess the questions within the usefulness scale (Davis, 1989). The score was used to evaluate the tool's usefulness and user acceptance (Doll et al., 1998) of message popup warnings as a debunking technique.

**Table 5**

*Usefulness Questions Based on Davis' (1989) Perceived Usefulness Measurement Scale*

<b>RQ</b>	<b>Construct</b>	<b>Question</b>
4	Usefulness	Using popup warning on social media would enable me to detect false news more quickly. Using popup warnings would improve my performance in detecting false news. Using popup warnings would increase my productivity on social media. Using popup warnings would enhance my effectiveness on social media. Using popup warnings would make it easier to use social media.  I would find popup warnings useful on social media.

Next, the System Usability Scale was employed to evaluate the message popup warning's usability (see Table 6). This scale has been widely used in HCI research, as it has been found to be a valid and reliable tool for measuring usability (Brooke, 1995).

**Table 6**

*Usability Questions Based on Brooke's (1995) System Usability Scale (SUS)*

<b>RQ</b>	<b>Construct</b>	<b>Question (5-pt Likert Scale)</b>
4	Usability	<p>I think that I would like to use popup warnings frequently.</p> <p>I found popup warnings unnecessarily complex.</p> <p>I thought popup warnings were easy to use.</p> <p>I think that I would need the support of a technical person to be able to use popup warnings.</p> <p>I found the various functions in popup warnings were well integrated.</p> <p>I thought there was too much inconsistency in popup warnings.</p> <p>I would imagine that most people would learn to use popup warnings very quickly.</p> <p>I found popup warnings very cumbersome/difficult to use.</p> <p>I felt very confident using popup warnings.</p> <p>I needed to learn a lot of things before I could get going with popup warnings.</p>

### *Quality Control*

Instruments are often judged on their reliability and validity. If reliability or validity is lacking, the data obtained is not reliable (Salkind, 2017). Reliability refers to the instrument giving consistent results as the experiment is repeated (Lazar et al., 2017; Salkind, 2017). Validity refers to whether an instrument measures the construct that it is attempting to measure (Salkind, 2017).

In the ISS, questions were adapted from Pennycook et al. (2020). In Pennycook et al.'s study (2020), the researchers measured perceived accuracy and then asked participants if they would share the news story. Kirchner and Reuter (2020) reused Pennycook et al.'s (2020) question, regarding accuracy, to evaluate the effect of various debunking techniques on the users' perception of accuracy. This study expanded the work of these researchers by reusing the survey questions and adapting as necessary.

Additionally, two experts independently evaluated the ISS and deemed it to be reliable and valid.

The SUS scale was created in 1995 to evaluate the usability of systems at Digital Equipment Corporation. Researchers have demonstrated it to be valid and reliable, even though this scale has fewer questions than other usability scales (Brooke, 2013). The scale has even indicated reliability with as few as eight to 12 users (Stetson & Tullis, 2004). Within the tool, every other question is written in a positive tense, and the others are written negatively (Brooke, 1995). The alternating positive and negative statements were created to prevent a user from using a single rating for all questions (Brooke, 2013). Since the question-types alternate, a scoring tool must be used to calculate the evaluation, which can be found in the Data Organization and Analysis section.

Davis' (1989) Perceived Usefulness scale has been tested by many researchers to confirm its validity and reliability (Doll et al., 1998). Some researchers believe that some of the six questions should be removed from the scale, yet others confirm the original scale's reliability (Hendrickson et al., 1993), good fit, and construct validity (Doll et al., 1998). Furthermore, the scale has been shown to be effective in evaluating multiple types of productivity software, across genders, and for both novice and expert users (Doll et al., 1998).

### **Data Collection – Multiple Sources**

To capture the data, the online interactive scenarios and survey (ISS) utilized Qualtrics to record the survey answers. User engagement (extent users read the full content contained in posting) and credibility were evaluated with the participants' responses to the first and fourth questions in Table 2. The engagement variable was

evaluated with a 4-point Likert scale; if the user fully read the content contained in posting, the variable was equal to 4. “Reading somewhat” was equal to 3 and “read little of article” was equal to 2. Ignoring the article, or responding “no, not at all”, the variable was equal to 1. Additionally, credibility was assessed by asking participants to rate the posting’s credibility after each news posting. Credibility was rated on a 4-point Likert scale, in line with Pennycook et al.’s (2020) and Kirchner and Reuter’s (2020) studies. Like Pennycook et al. (2020), the credibility ratings were the following: “very accurate” was equal to 4, “somewhat accurate” was equal to 3, “not very accurate” was equal to 2, and “not at all accurate” was equal to 1.

To assess the message popup warning’s effectiveness, participants were questioned about the user’s choice response, i.e., the user actions or inaction, to the false news articles (Kirchner & Reuter, 2020; Pennycook et al., 2020). The effectiveness (pre-treatment) and effectiveness (post-treatment) questions inquired about the user’s intent to like, share, or comment on the posting (the user’s choice response). The effectiveness pre-treatment and post-treatment questions were similar; however, pre-treatment was presented prior to the message popup warning and post-treatment was posed after the user views the message popup warning. The action variable was binary; if the user took any action (like, share, or comment), action equaled 1. If the user did not choose like, share, or comment (clicked on “Would not click on above choices”), action equaled 0.

For the treatment group participants, two additional tests were administered with questions regarding the usability and the usefulness of the message popup warnings. The usability of the message popup warnings was assessed by the Systems Usability Scale (Brooke, 1995). This ten-item test uses a scoring rubric to compute the usability score;



the scores can range from 0 to 100 (see Table 5) (Brooke, 1995). The SUS was administered to participants in the treatment group only. To evaluate the usefulness of message popup warnings, Davis' Measurement Scale for Perceived Usefulness (1989), consisting of six items, was used (see Table 6). These 7-point Likert questions were also administered to the treatment group. This measurement determined if the participants found the message popup warnings useful for detecting false news articles.

### **Data Organization and Analysis**

Beginning with the pilot test, data was captured within and about the ISS. Two expert reviewers created written feedback about test validity, interface usability, system issues and length of instrument. Modifications to the survey were made, according to the experts' findings. Next, 11 pilot test participants were recruited to test interface usability, system issues, and length of instrument through a thinking aloud process, while the researcher manually recorded the users' observations.

Qualtrics captured the data for the experiment. For the Interactive Scenario and Survey (ISS), data was captured through a series of questions (see in Table 2) which were coded as either binary choices (effectiveness) or Likert scales (engagement and credibility). Additionally, the treatment group answered questions from PUMS (see Table 5) and SUS (see Table 6), and the unique scoring guidelines were used for coding. Once data collection was complete, data was exported from Qualtrics to Microsoft Excel. Excel was used to prepare the data for analysis and for general data cleanup. Only surveys with user completed sections for ISS, PUMS, and SUS were analyzed; incomplete user sections were removed from the data set. Once ready, the data was either analyzed in Excel or imported to SPSS for more complex data analyses.

Once all the pilot tests were completed, the users' comments were analyzed for common themes or patterns. These themes were coded and assigned to categories (interface usability, system issues, and time on task). As applicable, quantitative or qualitative analysis was used to study the observations. The feedback was compared to expert findings. Modifications were made to the ISS address the found issues for interface usability and system issues.

Statistical tests were run on the data collected by the ISS. To answer RQ1 and H1 regarding user engagement (extent users read the full content contained in posting), a *t*-test was used to compare the differences in means between the control and treatment groups, specifically comparing the postings in the treatment group with message popup warning with the same postings without a message popup warning in the control group. Furthermore, posting types were compared. Posting types were broken down to three treatment types (false postings with message popup warnings = "Treatment False-Warning", false postings without warnings = "Treatment False-No Warning", and true postings = "Treatment True") and two control posting types (false postings without warnings = "Control False" and true postings = "Control True"). An Analysis of Variance (ANOVA) was used to compare the differences in means between the posting types in the control (Control True and Control False posts) and treatment (Treatment False-Warning, Treatment False-No Warning, Treatment True posts) groups (Sekaran & Bougie, 2016). Since six to 12 postings were in each category (Treatment False-Warning, Treatment False-No Warning, Treatment True, Control False, Control True), mean values were computed within each category for each user. Additionally, *p*-values were calculated to ensure the results are statistically significant at 0.05 or less, in order to reject

or fail to reject the hypothesis (Salkind, 2017). In a similar fashion, RQ3 and H3 which evaluated credibility, *t*-tests, ANOVA, *F* test statistics, degrees of freedom, *p*-values, and effect sizes were calculated to determine the statistical significance of differences in means between the control and treatment groups and between the five posting types (Control False and Control True posts) and treatment (Treatment False-Warning, Treatment False-No Warning, and Treatment-True posts) groups.

To answer R2 and H2, Chi-square tests were used to confirm whether the frequency distribution of effectiveness (action or inaction) is significant for treatment group postings with message popup warnings and the corresponding postings in the control group without warnings. To further analyze the data, posting types (Treatment False-Warning, Treatment False-No Warning, Treatment True, Control False, and Control True) were tested for significance in frequency distribution. Additionally, McNemar's test was used to calculate distributions for the effectiveness (pre-treatment) question administered prior to the treatment of popup warnings and the effectiveness (post-treatment) question administered after the warning post treatment (Sekaran & Bougie, 2016). As the Chi-square test cannot compare pre- and post-test results, McNemar's test must be used for the effectiveness binary variable ("SPSS Tutorials: Chi-Square Test of Independence," 2023). The exact significance was computed to determine significance for the McNemar's test; whereas, asymptotic and exact significance were computed to determine statistical significance for the Chi-square test (Sekaran & Bougie, 2016).

To avoid type I and type II errors by incorrectly rejecting or failing to reject the null hypotheses, effect size was computed for all three hypotheses (H1-H3) (Sekaran &

Bougie, 2016). As  $p$ -values and degrees of freedom are affected by the sample size, conversely, effect size focuses on the strength of the relationship or how large the difference is between groups (Sekaran & Bougie, 2016).

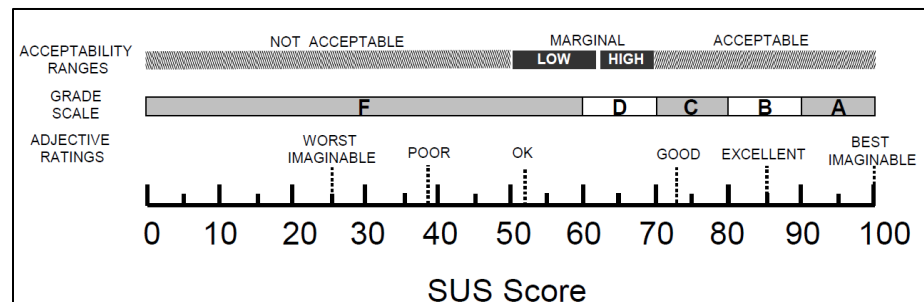
Regarding the SUS evaluation for the treatment group, the 10 questions can be found in Table 6. These questions were scored in the following manner:

- Questions 1, 3, 5, 7, and 9 were scored by noting the score position (on the 4-point Likert scale, ranging from 0 to 4) and subtracting 1.
- Questions 2, 4, 6, 8, and 10 were scored starting at 5 and subtracting the score position.
- The scores for the ten questions were added and then multiplied by 2.5 to calculate the total usability value (Brooke, 1995).

To understand what the total SUS score implies, researchers have devised a chart to give meaning to the users' perceptions of a system's usability (Bangor et al., 2009). Figure 9 breaks down the overall score by a descriptive adjective. According to Bangor et al. (2009), scores from 52 and above are positive usability experiences.

**Figure 9**

*SUS Ratings by Adjectives and Acceptability Ranges*



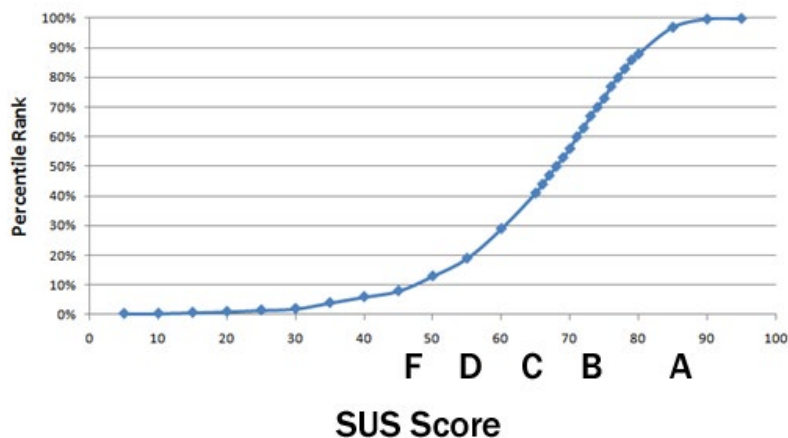
*Note.* From “Determining what individual SUS scores mean: Adding an adjective rating scale,” by A. Bangor, P. Kortum, and J. Miller, 2009, *Journal of Usability*

*Studies*, 4(3), p. 121. Copyright 2008-2009 by the Usability Professionals' Association and the authors.

Researchers have attempted to further define the SUS scores, by creating Percentile Rankings, as seen in Figure 10 (Sauro, 2011). For example, the passing score of 52 would score in the “D” range using the SUS Percentile Rankings, depicting that the score is more positive than negative but is not highly usable.

**Figure 10**

*SUS Percentile Rankings*



*Note.* From “Measuring Usability with the System Usability Scale,” by Jeff Sauro, 2011, *MeasuringU.com*. Copyright 2023 by MeasuringU.

The survey for the treatment group also includes Davis’ (1989) Perceived Usefulness Measurement Scale. This scale contains six questions to assess the perceived usefulness of a tool, using a 7-point Likert scale for responses (Davis, 1989). The answers ranging from *Extremely Likely* (1) to *Extremely Unlikely* (7) are scored, added together, and then averaged (Venkatesh & Davis, 1996).

## **Researcher, Ethical Considerations, and IRB**

Prior to the pilot test, Institutional Review Board (IRB) approval was granted by Nova Southeastern University's (NSU) IRB and Wentworth Institute of Technology's IRB. The researcher completed the pre-requisite CITI Human Subjects Protection training course prior to submitting the application for NSU's IRB approval. IRB approval process is necessary to ensure that no harm will come to the participants of this study (Lazar et al., 2017). NSU's IRB approved this study on July 21, 2023, and WIT's IRB approved this study on August 30, 2023.

The experiment was conducted as an online survey and, thus, did not require the physical presence of the participant; it was expected that its contents would cause little to no harm to the participants. Before beginning the survey, participants viewed the informed consent form, which listed the risks and contained information regarding participation (Salkind, 2017). Participants were informed that participation was voluntary and could be terminated at any time. As participants were asked to read headlines, full articles, and answer questions, it was possible that survey fatigue could have been experienced, as the survey was approximately 25 minutes in length (see Appendix B). As shown in Appendix B, the consent form also contained information, such as time commitment, purpose of study, researcher's name and contact information, how data would be used, benefits of the research, costs to participants, and compensation for participation (Lazar et al., 2017; Salkind, 2017).

As privacy was of utmost importance, participants' identifying information was only recorded on a single master file (Salkind, 2017). This master file contained participant information that was not accessible to anyone but the researcher (Salkind,

2017). Identifying information collected was comprised of demographics (gender, age, and race) and email addresses. Participants were informed how their personal information will be stored and used (Lazar et al., 2017). In the data cleanup phase, all identifying information was stripped from the data to ensure privacy of all participating individuals.

At the completion of the ISS, participants were debriefed. Like Geeng et al.'s (2020) study, participants were unaware at the survey's outset which of the news postings are false. The debriefing immediately followed the survey, explaining which of the scenarios contain false information, so that participants would not share misinformation unknowingly (Geeng et al., 2020; Salkind, 2017; Sekaran & Bougie, 2016). To reduce the risk of participant distress, the debriefing also stated that identifying misinformation is difficult and many are tricked by false news (Geeng et al., 2020). Additionally, a list of all false postings was shared with the participants (See Appendix A).

## **Resources**

For the experiment, the researcher acquired the following resources:

- A survey platform to host the surveys and provide interactive functionality (Qualtrics)
- Software to design the screen mockups (Canva)
- Software to build and host website (Wix) - \$215
- License for images (cost for images and Dreamstime license) - \$79
- Participants
- Participants must have access to a digital device to take the survey (laptop, tablet, or smartphone)
- Gift card compensation for participants (\$10 Amazon gift cards) - \$1,100

- Gift card compensation for pilot participants (\$20 Amazon gift cards) - \$220
- SPSS for data analysis
- Total cost was approximately \$1,600.

## **Summary**

The overarching research question of this study was: Are message popup warnings effective for deterring the spread of misinformation? Message popup warnings were evaluated on effectiveness, credibility, and user engagement using the ISS. The interactive interface of the survey captured user engagement (extent users read the full content contained in posting) with the article. The users' choice response (clicking or not clicking on like, share, or comment) determined the effectiveness of popup warnings. Message popup warnings were assessed on the user's perception of credibility, usefulness, and usability for detecting misinformation.

Employing a between-subjects experimental study design, this study had two randomly assigned groups, a treatment and a control group. The Interactive Scenario and Survey (ISS) simulated a Facebook platform, containing a newsfeed with true and false postings. For the treatment group, half of the false postings initially appeared without warnings and after answering one question, the user received a message popup warning. Conversely, the control group did not receive message popup warnings during the ISS.

The ISS was tested for reliability and validity. The constructs of usefulness and usability were evaluated using the Perceived Usefulness Measurement Scale (PUMS) and System Usability Scale (SUS) instruments, respectively. By adapting questions from Kirchner and Reuter's (2020) and Pennycook et al.'s (2020) studies and using the PUMS



and SUS tools, these well-established instruments satisfied some initial reliability and validity concerns. By utilizing both evaluation experts and eligible participants, the pilot test ensured instrument validity. Additionally, to ensure participants were not harmed in this study, participants were briefed about the study and the potential risks regarding participation in the study, then completed informed consent forms. Approval was received from the IRBs of both NSU and WIT prior to the experiment. Participants were debriefed at the end of the survey to bring successful closure and to reduce any potential distress they may have experienced regarding their responses to the survey.

Data analysis was quantitative in nature, using the Chi-square test, McNemar's test, *t*-tests, ANOVA, and the SUS and PUMS scoring guidelines. The test results were analyzed, taking into consideration the sample size when rejecting or accepting the null hypotheses. Additionally, the SUS score was interpreted by using Bangor et al.'s (2009) SUS Descriptive Adjectives, Bangor et al.'s (2008) SUS Acceptability Ranges, and Sauro's (2011) Percentile Ranking.

## Chapter 4

### Results

#### Overview

Chapter 4 presents the results and analysis for the overarching research question of: Are message popup warnings effective for deterring the spread of misinformation? This study employed a between-subject experimental design, with users randomly assigned to either treatment or control groups. The ISS, SUS, and PUMS were used to collect quantitative data to evaluate the constructs of user engagement, effectiveness, credibility, usefulness, and usability. Qualtrics hosted the survey and was the platform used for data collection. This chapter outlines the results of the survey questions and corresponding research questions.

#### Descriptive Statistics

The survey was made available to the students at Wentworth Institute of Technology over four days from October 10 through October 13, 2023. During that period, 206 individuals volunteered to participate. There were 109 participants who were qualified through pre-screening questions and age restrictions and successfully completed the survey, resulting in a completion rate of nearly 53%. Qualtrics captured demographic data for the eligible participants ( $N = 109$ ). Participants were randomly assigned to either the control or treatment group, with 54 participants in the control and 55 participants in the treatment group.

Wentworth Institute of Technology students were the target audience for this research study. Table 7 contains several demographic attributes. As the campus is predominantly male, the breakdown of gender was representative of the student body: 81 male, 21 female, 1 transgender female, 3 non-conforming genders, 1 gender not listed, and 2 preferred to not answer. Regarding age, the participants were primarily of traditional college-age: 106 participants were 18-24 years old, 2 participants were 25-34 years old, and 1 participant was 35-44 years old. Additionally, these participants were heavy social media users with 88% of participants use social media daily and 77% reported that they click on like, share, or comment on social media news postings at least once a week.

**Table 7**

*Descriptive Statistics of Study Participants (N = 109)*

Descriptive Statistic	N	%
<b>Age</b>		
18-24	106	97.25%
25-34	2	1.83%
35-44	1	0.92%
45-54	0	0.00%
55-64	0	0.00%
65+	0	0.00%
<b>Gender</b>		
Female	21	19.27%
Male	81	74.31%
Transgender female	1	0.92%
Transgender male	0	0.00%
Non-conforming	3	2.75%
Not listed	1	0.92%
Prefer to not answer	2	1.83%
<b>How often do you use social media?</b>		
Less than monthly	0	0.00%
1-2 times a month	3	2.75%
Once a week	3	2.75%
2-3 times a week	7	6.42%

Descriptive Statistic	N	%
Daily	96	88.07%
How often do you click on like, share, comment on social media?		
Less than monthly	15	13.76%
1-2 times a month	10	9.17%
Once a week	19	17.43%
2-3 times a week	15	13.76%
Daily	50	45.87%

### Data Analysis

The ISS contained 24 postings, consisting of 12 false and 12 true postings, each included hyperlinks to the full content of the articles. Participants were asked to answer questions on effectiveness (click on like, share, comment, or chose to take no action), user engagement (extent users read the full content contained in the posting), and rate the credibility of the posting. In the treatment group, the 12 false postings were split in half; six postings presented with message popup warnings and six postings presented with no warning. For the six postings with a message popup warning, participants received an effectiveness question twice, once prior to the message popup warning and once after viewing the popup warning. The remaining six false postings with no message popup warning only presented the effectiveness question a single time. Additionally, the treatment group completed usability (SUS) and usefulness (PUMS) questions about message popup warnings at the end of the survey.

The results from the ISS were analyzed to evaluate effectiveness, user engagement, credibility, usability, and effectiveness. User engagement and credibility were analyzed using independent samples *t*-tests and Analysis of Variance (ANOVA) tests. Both McNemar's and Chi-square tests were utilized for effectiveness. The SUS and PUMS scoring guidelines were used to evaluate usability and usefulness, respectively.

### *User Engagement*

The ISS was employed to gather data for user engagement. RQ1 posed the following question: Regarding user engagement, to what extent do users read the full content contained in postings when comparing postings that are presented with or without message popup warnings? To evaluate this research question, H1, as a null hypothesis, stated the following: Regarding user engagement, there will be no significant differences when comparing the extent users read the full content contained in postings that are presented with or without message popup warnings. For each of the 24 scenarios in the ISS, participants were given the opportunity to open and read the full content contained in the posting. After each posting, participants were asked “Did you read the previous article?” and answered using a 4-pt Likert scale (1=“no, did not read”, 2=“read little of article”, 3=“read somewhat”, 4=“yes, read fully”).

Data was analyzed by comparing user engagement with the postings of the treatment group with the control group. Similar to Kirchner and Reuter’s (2020) data analysis, this study compared the same postings in the treatment group presented with a message popup warning (postings #1-6) with those in the control group presented without a message popup warning (postings #1-6). For each participant, the answers to the six postings were averaged. An independent samples *t*-test was performed in IBM SPSS to evaluate whether there was a difference between the user engagement of the treatment and control groups (see Table 8). The results indicated that there was no significant difference,  $t(107) = 0.429$ ,  $p = .669$ , in user engagement between the control group ( $M = 2.18$ ,  $SD = .84$ ) and treatment group ( $M = 2.12$ ,  $SD = .79$ ). The effect size, as measured by Cohen’s *d*, was  $d = 0.08$ , indicating a very small effect (Cohen, 2013). As such, H1 was

failed to reject; there was no significant difference in user engagement (extent users read the full content) when a message popup warning was received compared with postings with no popup warning.

**Table 8**

*User Engagement Comparison for the Control and Treatment Groups*

	Control		Treatment		<i>t</i> (107)	<i>p</i>	Cohen's <i>d</i>
	M	SD	M	SD			
User Engagement	2.182	0.843	2.115	0.786	0.429	0.669	0.082

Furthermore, findings are consistent between the control and treatment groups, when comparing user engagement for each of the 24 postings. Table 9 shows that none of the postings had a significant difference in user engagement (no *p*-values fall at or below .05) when *t*-tests were executed for each posting. Only two postings came close to a significant difference yet neither presented with a warning; posting #12 was a false posting without a warning (*p* = .086) and posting #18 was a true posting (*p* = .101). For the 24 postings, effect size ranged from very small to small (*d* ranges from .03 to .33), as measured by Cohen's *d*.

**Table 9**

*Comparison of Control and Treatment Groups for User Engagement by Posting Number*

Posting Number	Control ( <i>n</i> =54)		Treatment ( <i>n</i> =55)		<i>t</i> (107)	<i>p</i>	Cohen's <i>d</i>
	Mean	SD	Mean	SD			
1	2.26	1.200	2.05	1.026	0.958	0.340	0.183
2	2.11	1.160	1.87	0.924	1.188	0.238	0.228
3	2.46	1.193	2.20	1.112	1.191	0.236	0.228
4	2.04	1.228	2.07	1.034	-0.164	0.870	-0.031
5	2.28	1.235	2.24	1.053	0.188	0.851	0.036

Posting Number	Control (n=54)	Treatment (n=55)	<i>t</i> (107)	<i>p</i>	Cohen's <i>d</i>	Posting Number	Control (n=54)
6	1.94	1.140	1.95	1.044	-0.005	0.996	-0.001
7	1.93	1.096	1.95	0.931	-0.100	0.920	-0.019
8	2.35	1.200	2.18	0.964	0.816	0.416	0.156
9	2.35	1.305	2.22	1.166	0.564	0.574	0.108
10	2.07	1.163	2.11	1.012	-0.168	0.867	-0.032
11	2.22	1.254	2.11	1.012	0.519	0.605	0.099
12	1.78	0.984	2.11	1.012	-1.732	0.086	-0.332
13	2.00	1.099	1.93	1.016	0.359	0.720	0.069
14	2.33	1.149	2.13	1.139	0.940	0.349	0.180
15	2.00	1.064	2.31	1.034	-1.538	0.127	-0.295
16	2.19	1.100	2.20	1.129	-0.069	0.945	-0.013
17	1.98	1.073	1.95	0.931	0.187	0.852	0.036
18	2.48	1.145	2.13	1.090	1.655	0.101	0.317
19	2.24	1.288	2.07	0.979	0.768	0.444	0.147
20	2.20	1.105	2.09	0.986	0.562	0.575	0.108
21	2.06	1.140	2.13	0.862	-0.371	0.711	-0.071
22	2.02	1.124	2.27	0.990	-1.254	0.213	-0.240
23	2.00	1.116	1.95	0.970	0.272	0.786	0.052
24	2.35	1.119	2.22	1.049	0.644	0.521	0.123

To further analyze the data, the postings were sorted into five posting types. As in Pennycook et al. (2020) and Kirchner and Reuter (2020), other posting types were compared after analyzing the false postings with warnings (treatment) and without warnings (control). Comparing posting types enables researchers to look for residual effects of message popup warnings on postings without warnings (both true and false postings). Pennycook et al. (2020) observed significant differences for postings without warnings in the treatment group compared with the same postings in the control group. For this study, the treatment group split into three posting types: (1) treatment group had six false postings that presented with a message popup warning (Treatment False-Warning), (2) treatment group had six false postings with no warnings (Treatment False-No Warning), and treatment group had 12 true postings (Treatment True). The control

group split into two posting types: (1) control group had 12 false postings with no warnings (Control False) and (2) control group had 12 true postings (Control True). All treatment posting types had 55 participants ( $n = 55$ ) and all control posting types had 54 participants ( $n = 54$ ). For each participant, the answers to the credibility question for each posting type were averaged. These posting type averages were then compared in IBM SPSS using a one-way ANOVA test to evaluate the effect of message popup warnings on user engagement. As seen in Table 10, the one-way ANOVA shows that there is no statistically significant difference in user engagement (the extent the participants read the full content message) between the five posting types (Treatment False-Warning, Treatment False-No Warning, Treatment True, Control False, Control True) ( $F(4, 268) = .124, p = .974$ ). Additionally, the effect size was found to be small ( $\eta^2 < .01$ ).

**Table 10**

*Means, Standard Deviations, and One-Way Analyses of Variance for User Engagement Among Posting Types*

Posting Type	<i>N</i>	Mean	SD	Mean Square	<i>F</i> (4, 268)	$\eta^2$
Between Groups				0.073	0.124*	.00
Treatment False-Warning	55	2.064	0.732			
Treatment False-No Warning	55	2.112	0.744			
Treatment True	55	2.114	0.687			
Control False	54	2.150	0.825			
Control True	54	2.154	0.840			
Within Groups				0.589		

\*  $p > .05$

### *Effectiveness*

Effectiveness of the message popup warnings were evaluated using data from the ISS. The second research question (RQ2) asked the following question: Regarding effectiveness, what is the user choice response (actions or inactions) to postings when



comparing postings that are presented with or without message popup warnings? The corresponding null hypothesis for H2 stated the following: Regarding effectiveness, there will be no significant differences in user choice response (actions or inactions) to postings, when comparing postings that are presented with or without message popup warnings.

Effectiveness was assessed by tracking the participants' user choice response (clicking on "like", "share", "comment", or "would not click on above choices") to each posting. Data was captured as a binary number; if the user clicked on either like, share, or comment, the value was equal to 1, otherwise choosing "would not click on the above choices" was assigned the value of 0. In the treatment group only, participants were asked twice if they would consider taking an action, before and after the message popup warnings. For all non-popup postings (postings #7-24 in the Treatment group and all postings in the Control group), this question was only posed a single time.

To evaluate H2, the Chi-Square test was employed. Responses to the treatment postings that contained a message popup warning were compared to the same postings in the control group. A Chi Square test of independence was performed to evaluate the relationship between message popup warnings and effectiveness. The relationship between these variables was not significant and had a negligible association,  $\chi^2(1, 654) = .03, p = .863, \phi = -.007$  (see Table 11). The control group was not more likely to click on like, share, or comment than the treatment group. Consequently, H2 was failed to reject as user choice response (actions or inactions) to postings was not significantly different when comparing postings that are presented with or without message popup warnings.

**Table 11**

*Frequencies and Chi-Square Results for Effectiveness Comparing Control and Treatment Groups*

	Control		Treatment		$\chi^2$	$\phi$
	N	%	N	%		
Effectiveness					0.030*	-.007
Would not take action	209	64.5%	215	65.2%		
Clicked on like, share, or comment	115	35.5%	115	34.8%		

\*  $p > .05$

To further evaluate effectiveness, McNemar's test was used. When comparing pre- and post-test results, the Chi-square test was not used as the observations are not independent of each other ("SPSS Tutorials: Chi-Square Test of Independence," 2023). Therefore, McNemar's test is the appropriate test to use to compare the pre- and post-test results from the treatment group ("SPSS Tutorials: Chi-Square Test of Independence," 2023). Treatment group postings #1-6 were used for the pre- and post-test comparison. A McNemar's test was performed to evaluate the difference in effectiveness as affected by message popup warnings. The McNemar's test determined that there was a statistically significant difference in the proportion of participants who took action (clicked on like, share, or comment) before receiving a message popup warning than those who took action after receiving a message popup warning,  $p < .001$ , (see Table 12). The Phi value ( $\phi = .653$ ) indicates a strong association between message popup warnings and effectiveness (Rea & Parker, 2014).

**Table 12**

*Crosstabulation and McNemar's Test of Effectiveness Comparing Pre- and Post-Test Results*

		Post-Treatment				$\phi$
		No Action		Action		
		N	%	N	%	
Pre-Treatment	No Action	169	78.6%	12	10.4%	0.653*
	Action	46	21.4%	103	89.6%	

\*  $p < .001$

Chi-square tests were performed to evaluate the difference between all five posting types: Treatment False-Warning, Treatment False-No Warning, Treatment True, Control False, and Control True. Table 13 demonstrates that there was a significant difference ( $p < 0.01$ ) in user actions between all the treatment posting types: false postings with a message popup warning (Treatment False-Warning), false postings without a message popup warning (Treatment False-No Warning), and true postings (Treatment True).

**Table 13**

*Crosstabulation and Chi-Square Results for Effectiveness for All Posting Types*

		Treatment False-Warning				$\chi^2$	$\phi$
		No Action		Action			
		N	%	N	%		
Treatment False-No Warning	No Action	145	67.4%	43	37.4%	27.600*	0.289
	Action	70	32.6%	72	62.6%		
Treatment True	No Action	126	58.6%	37	32.2%	20.939*	0.252
	Action	89	41.4%	78	67.8%		
Control False	No Action	143	66.5%	70	60.9%	1.042	0.056
	Action	72	33.5%	45	39.1%		
Control True	No Action	115	53.5%	59	51.3%	.143	0.021
	Action	100	46.5%	56	48.7%		

\*  $p < .01$

The differences are demonstrated in the action and no action frequencies in Table 14. There were significant differences in user choice responses between all the treatment posting types (Treatment False-Warning, Treatment False-No Warning, and Treatment True) ( $p < .01$ ). The participants took action on 34.8% of the treatment group postings with message popup warnings (Treatment False-Warning), compared with the participants who took action on 43% of the treatment group false postings without message popup warnings (Treatment False-No Warning). Likewise, there was a significant difference,  $p < .01$ , between the user choice responses for treatment group false postings with a message popup warning (Treatment False-Warning) with 34.8% choosing to take action and treatment group true postings (Treatment True) with 48.8% choosing to take action. Comparatively, there was no significant difference ( $p = 0.307$  and  $p = 0.705$ , respectively) between the treatment group false postings with a message popup warning (Treatment False-Warning) and either the true or false postings in the control group (Control False and Control True).

**Table 14**

*Frequencies for Effectiveness by Posting Type*

Posting Type	No Action		Action	
	N	%	N	%
Treatment False-Warning	215	65.2%	115	34.8%*
Treatment False-No Warning	188	57.0%	142	43.0%*
Treatment True	338	51.2%	322	48.8%*
Control False	414	63.9%	234	36.1%
Control True	344	53.1%	304	46.9%

\*  $p < .01$

### *Credibility*

Research question (RQ3) evaluated the credibility of the posting, with the following question: Regarding credibility, how does the debunking technique of message popup warnings influence the users' credibility rating of the posting, compared to ratings given for postings without message popup warnings? The null hypothesis H3 stated the following: Regarding credibility, there will be no significant differences in how users rate the credibility of the posting when receiving the message popup warning, compared to ratings given for postings without message popup warnings. To gather data for this construct, the participants were asked after each posting, "To the best of your knowledge, how accurate is this article?" and responded using a 4-point Likert scale (1="not at all accurate", 2="not very accurate", 3="somewhat accurate", 4="very accurate").

To analyze credibility, the data from the control and treatment groups were compared. Consistent with the analysis of the other constructs, this study compared the postings in the treatment group which presented with a message popup warning with the same postings in the control group which presented with no message popup warning. The responses to these six postings were averaged, then an independent samples *t*-test was performed to assess whether there was a difference of credibility between the control and treatment groups. (see Table 15). The results indicated that there was a significant difference,  $t(107) = 2.46, p = .015$ , for credibility between the control group ( $M = 2.33, SD = .84$ ) and treatment group ( $M = 1.99, SD = .6$ ). Cohen's *d* was  $d = 0.47$ , indicating that the effect size was medium. Consequently, H3 was rejected; there was a significant difference in credibility ratings when a message popup warning was received compared with postings without a popup warning.

**Table 15***Comparison of Credibility for Control and Treatment Groups*

	Control		Treatment		<i>t</i> (107)	<i>p</i>	Cohen's <i>d</i>
	M	SD	M	SD			
Credibility	2.333	0.835	1.991	0.601	2.460	0.015	0.471

Additionally, when comparing by individual posting number, only four postings had a significantly different credibility rating among control and treatment groups (posting #3, #6, #13, and #22,  $p < .05$ ) (See Table 16). Two of the four, #3 and #6, were postings that received a message popup warning in the treatment group; this finding suggests that message popup warnings may have affected credibility ratings in two out of six instances (six postings had message popup warnings out of 24 total postings) or 33.3% of the time. Additionally, there was a third posting with a message popup warning that came close to a significant difference (posting #1,  $p = .108$ ). For the postings without message popup warnings, credibility was significantly different between the control and treatment group for two out of 18 instances (18 postings did not present with message popup warnings) or 11.1% of the time. As measured by Cohen's *d*, effect size varies from very small to large (*d* ranges from .003 to .861).

**Table 16***Comparison of Control and Treatment Groups for Credibility by Posting Number*

Posting Number	Control ( <i>n</i> =54)		Treatment ( <i>n</i> =55)		<i>t</i> (107)	<i>p</i>	Cohen's <i>d</i>
	Mean	SD	Mean	SD			
1	1.94	1.071	1.63	0.886	1.619	0.108	0.315
2	2.17	0.927	2.21	0.871	-0.257	0.798	-0.050
3	2.89	0.984	2.31	0.981	3.044	0.003	0.591
4	1.87	1.010	1.92	1.026	-0.266	0.79	-0.052

Posting Number	Control (n=54)	Treatment (n=55)	<i>t</i> (107)	<i>p</i>	Cohen's <i>d</i>	Posting Number	Control (n=54)
5	2.22	1.058	2.06	0.938	0.846	0.399	0.164
6	2.91	2.672	1.90	0.869	2.579	0.011	0.501
7	2.06	0.960	2.38	0.844	-1.872	0.064	-0.364
8	1.83	0.906	2.27	1.848	-1.551	0.124	-0.301
9	1.81	1.117	1.88	1.003	-0.338	0.736	-0.066
10	2.33	0.971	2.38	0.911	-0.28	0.78	-0.054
11	2.35	1.152	2.31	0.961	0.214	0.831	0.042
12	2.11	0.904	2.37	0.817	-1.517	0.132	-0.295
13	2.44	1.058	2.02	0.939	2.185	0.031	0.425
14	2.15	1.089	1.98	1.000	0.824	0.412	0.160
15	2.69	0.948	2.65	0.883	0.176	0.861	0.034
16	2.37	1.033	2.15	0.998	1.097	0.275	0.213
17	2.44	1.022	2.56	0.938	-0.594	0.554	-0.115
18	2.67	1.046	2.40	1.015	1.312	0.192	0.255
19	2.89	1.040	3.27	2.206	-1.142	0.256	-0.222
20	2.56	0.945	2.63	0.841	-0.454	0.65	-0.088
21	2.52	1.005	2.48	0.874	0.206	0.837	0.040
22	1.98	1.000	2.77	2.193	-2.394	0.018	-0.465
23	1.96	1.045	1.69	0.853	1.458	0.148	0.283
24	2.54	0.985	2.44	0.938	0.507	0.613	0.098

To further analyze credibility, postings were categorized into five posting types (Treatment False-Warning, Treatment False-No Warnings, Treatment True, Control False, Control True). Each posting type consisted of either six or 12 postings, and these answers were then averaged. An ANOVA was used to assess the differences between these posting types. The one-way ANOVA shows there was a statistically significant difference in credibility ratings between the five posting types. Table 17 demonstrates a significant difference between the posting types (Treatment False-Warning, Treatment False-No Warning, Treatment True, Control False, Control True) ( $F(4, 268) = 4.516$ ,  $p = .002$ ). The effect size was found to be medium ( $\eta^2 < .06$ ).

**Table 17**

*Means, Standard Deviations, and One-Way Analyses of Variance for Credibility Among Posting Types*

Posting Type	<i>N</i>	Mean	SD	Mean Square	<i>F</i> (4, 268)	$\eta^2$
Between Groups				1.796	4.516*	0.063
Treatment False-Warning	55	1.991	0.601			
Treatment False-No Warning	55	2.264	0.607			
Treatment True	55	2.421	0.651			
Control False	54	2.208	0.697			
Control True	54	2.434	0.590			
Within Groups				0.398		

\*  $p < .01$

Tukey's HSD Test for multiple comparisons found that the mean value of credibility was significantly different between Treatment False-Warning and Treatment True ( $p = .004$ , 95% C.I. = -.761, -1.000) and Treatment False-Warning and Control True ( $p = .003$ , 95% C.I. = -.774, -0.111) (See Table 18).

**Table 18**

*Tukey's HSD Test to Compare Credibility Ratings for Postings with Popup Warnings with Other Posting Types*

Posting Types		Mean Difference	Std. Error	<i>p</i>	95% Confidence Interval	
					Lower Bound	Upper Bound
Treatment False-Warning	Treatment False-No Warning	-0.273	0.120	0.159	-0.603	0.058
	Treatment True	-0.430	0.120	0.004	-0.761	-0.100
	Control False	-0.217	0.121	0.376	-0.549	0.114
	Control True	-0.443	0.121	0.003	-0.774	-0.111



### *Usability and Usefulness*

RQ4 focused on the constructs of usability and usefulness, posing the following question: How do users rate the usability and usefulness of message popup warnings as a debunking technique? Usability and usefulness of message popup warnings were evaluated using the instruments, SUS and PUMS, respectively. As only the treatment group received message popup warnings, the SUS and PUMS were administered to only these participants. Within the treatment group, all participants ( $n = 55$ ) completed the SUS and PUMS.

With regard to usability, the SUS contained 10 statements which participants evaluated on a 5-point Likert scale (SUS questions can be found in Table 6). Within the SUS, every other question is written positively and the opposite questions are written negatively. For assessment, scores were imported to Microsoft Excel to calculate individual and aggregate scores. Participants' scores ranged from 20 to 75 for the usability of the popup warning; the mean score of usability was 46. By using Bangor et al.'s (2008) acceptability ranges, message popup warnings score in the unacceptable range (see Table 19). Looking at the individual results, 56.36% of participants scored in the unacceptable range, 36.36% scored low marginal, and 3.64% scored high marginal.

**Table 19**

*SUS Scores Using Bangor et al.'s (2008) Acceptability Ranges*

Acceptability Ranges	Acceptability Scores	Acceptability Counts	Acceptability Percentage
Unacceptable	0 to less than 50	31	56.36%
Low Marginal	50 to less than 63	20	36.36%
High Marginal	63 to less than 70	2	3.64%
Acceptable	70 to 100	2	3.64%

Conversely, if the scores are interpreted using Bangor et al.'s (2009) descriptive adjectives, a composite score of 46 on the SUS scores message popups in the "OK" category (see Table 20). In their study, Bangor et al. (2009) discussed that the "OK" terminology may be better labeled as "fair" or "so-so", to imply that the usability was somehow deficient or not fully acceptable. By breaking down the numbers, nearly 31% of participants found message popup warnings to not be usable, scoring either worst imaginable or poor on Bangor et al.'s (2009) descriptive adjective chart. However, over 69% of participants found message popup warnings' usability either OK (27.27%), good (40%), or excellent (1.82%). No participants found the popup's usability in the range of best imaginable.

**Table 20**

*SUS Scores Using Bangor et al.'s (2009) Descriptive Adjectives*

Descriptive Adjective	Descriptive Scores	Descriptive Counts	Descriptive Percentage
Worst Imaginable	Less than 25	1	1.82%
Poor	25 to less than 39	16	29.09%
OK	39 to less than 52	15	27.27%
Good	52 to less than 73	22	40.00%
Excellent	73 to less than 85	1	1.82%
Best Imaginable	85 to 100	0	0.00%

To understand the SUS score, researchers can use the SUS percentile rankings, acceptability ranges, and descriptive adjectives. Using the SUS Percentile Rankings (see Figure 10), the composite score of 46 for message popup warnings scores as an "F". In summary, the composite score of 46 for message popup warnings scored "unacceptable" on the acceptability range, "OK" on the descriptive adjectives range, and an "F" on the

SUS percentile rankings, clearly demonstrating that message popup warnings are not usable.

Comparatively, participants completed PUMS to evaluate usefulness. Usefulness was assessed using a 7-point Likert scale (1=extremely likely, 2=moderately likely, 3=slightly likely, 4=neither likely or unlikely, 5=slightly unlikely, 6=moderately unlikely, and 7=extremely unlikely). Perceived usefulness of message popup warnings was slightly positive ( $M = 3.294$ ,  $SD = 1.894$ ) (see Table 21). Some studies have calculated perceived usefulness into scores with a 100-point denominator (Lewis, 2019) in order to use Bangor et al.'s (2009) descriptive adjective ratings created for interpreting usability (van der Nat et al., 2022). Following van der Nat et al's (2022) example, scores were added together for each participant, subtracting 1 from the score's position. These scores were then multiplied by 2.7778 (ensuring a maximum score of 100) to calculate the overall perceived usefulness score (van der Nat et al., 2022). The overall usefulness score falls in the "OK" rating in the descriptive adjective ranking, with an score of 43.539. Similar to the usability rating, the usefulness of the message popup warnings is found to be fair but not fully acceptable (Bangor et al., 2009).

**Table 21**

*Means, Standard Deviations, and Variance for PUMS Scores by Question*

Question	Mean	SD	Variance
Using popup warnings on social media would enable me to detect false news more quickly.	3.091	1.937	3.751
Using popup warnings would improve my performance in detecting false news.	2.964	1.885	3.554
Using popup warnings would increase my productivity on social media.	3.818	1.806	3.263
Using popup warnings would enhance my effectiveness on social media.	3.200	1.690	2.855

Question	Mean	SD	Variance
Using popup warnings would make it easier to use social media.	3.491	2.081	4.329
I would find popup warnings useful on social media.	3.200	1.966	3.867
Overall Score	3.294	1.894	

## Summary of Results

This chapter reviewed the results of the four research questions and three hypotheses in this study. Table 22 outlines the results of the three hypotheses. For the construct of user engagement, no significant differences were found between postings with message popup warnings (treatment group) and postings without message popup warnings (control group); users were not prone to read more or less of the postings with the presentation of popup warnings. Thus, H1 was failed to reject. Likewise, H2 was failed to reject as no significant differences were found for user choice behaviors between postings with message popup warnings (treatment) group and postings without warnings (control group). Conversely, H3 was rejected as credibility ratings were found to be significantly different between postings with message popup warnings (treatment group) and postings without warnings (control group).

## Table 22

### *Results of Hypotheses*

Hypothesis	Results
H1: Regarding user engagement, there will be no significant differences when comparing the extent users read the full content contained in postings that are presented with or without message popup warnings.	Failed to reject
H2: Regarding effectiveness, there will be no significant differences in user choice response (actions or inactions) to postings, when comparing postings that are presented with or without message popup warnings.	Failed to reject

Hypothesis	Results
H3: Regarding credibility, there will be no significant differences in how users rate the credibility of the posting when receiving the message popup warning, compared to ratings given for postings without message popup warnings.	Rejected

Usability and usefulness ratings were similar, with interpretation varying based on scale used. Usability was found to be lacking; using the Bangor et al.'s (2008) acceptability ranges and Sauro's (2011) percentile rankings, the usability of message popup warnings was found to be unacceptable and scored an "F", respectively. Using Bangor et al.'s (2009) descriptive adjectives, message popup warnings were found to be OK but not qualifying as fully acceptable. Likewise, usefulness scored in the OK category in Bangor et al.'s (2009) descriptive adjective ranking, which can also describe the warning's usefulness fair yet somehow deficient.

## Chapter 5

### Conclusions

#### Introduction

In Chapter 4, the sample and data gathering techniques were described, and data was analyzed to address the research questions and hypotheses. Chapter 5 discusses the constructs of user engagement, effectiveness, credibility, usefulness, and usability and how these constructs have been assessed to aid in answering the overarching research question of: Are message popup warnings effective for deterring the spread of misinformation?

#### Conclusions

In this section, each construct is discussed in regard to message popup warnings in social media. The results of the data are synthesized to better understand what was learned from this study.

#### *User Engagement*

The first research question, RQ1, focused on user engagement. Within this study, user engagement was defined as the extent users read the full content contained in the posting. Extant research has not offered access to full posting content, but rather has asked users to make decisions based on a posting's headline and picture (Kirchner & Reuter, 2020; Pennycook et al., 2020). This study provided a hyperlink to the full content (similar functionality to current-day Facebook) to determine whether users would seek out additional information after a message popup warning was presented. Data collected

indicated that message popup warnings had no effect on the extent the users read the full content contained in the posting. However, during the pilot test's think-aloud process, participants commented that they would perform Internet searches on their own to check the posting's veracity, not trusting the full content provided. The researcher did not ask participants to list which sources that they would trust. Studies have shown that trust can be based on trust in the poster or group, number of likes or comments, number of times the user views the posting, or even if the posting corroborates a previously held belief (Ma et al., 2019; Moravec et al., 2020; Spezzano et al., 2021).

Data was further analyzed to look for differences in user engagement between the five posting types (Treatment-Warning, Treatment-No Warning, Treatment True, Control False, and Control True). The ANOVA results were consistent with the *t*-tests run between the treatment and control groups; the results showed no significant differences in user engagement among the groups. As user engagement is understudied, it is unknown why there was little variation. It is possible that the subject matter of postings were not of interest to the participants, or that the participants' age or motivation to participate in the survey was a factor. In summary, this data suggests that presentation of message popup warnings did not change the users' behavior regarding user engagement, as the extent the participants read the full content contained in the posting was not significantly different from postings without warnings. Future studies could evaluate user engagement with greater participant age variation and the use of political postings or other big news stories for comparison.

### *Effectiveness*

The second research question, RQ2, compared effectiveness between postings with and without message popup warnings. To assess effectiveness, participants chose to take action (click on “like”, “share”, or “comment”) or not take action (click on “would not click on above choices”) for each posting. When comparing effectiveness of message popup warnings for the treatment and control groups, there were no significant differences ( $p > .05$ ). These results may be due to participants deducing that the false postings were indeed false. If the posting was perceived to be false (regardless of whether or not a message popup warning was received), the participants may have decided to not take action. Also, participants may have been disinterested in the subject matter of the first six postings. If the posting held little interest, participants may have decided to not share the posting with others.

Conversely, there was a significant difference ( $p < .01$ ) found for user choice response between the three treatment posting types (Treatment False-Warning, Treatment False-No Warning, and Treatment True) for postings with message popup warnings and postings without warnings. This is an important finding, as it corroborates results from Pennycook et al. (2020). In that study, participants were less likely to share the posting with others when some postings were labeled as false (Pennycook et al., 2020). Consistent with Pennycook et al. (2020), in this study, participants took action less often when message popups were presented on some postings.

Additional analysis found a significant difference ( $p < .001$ ) in user choice responses between the users’ actions prior to receiving a message popup warning and after receiving the warning (treatment group). This data suggests that the appearance of



the message popup warning influenced some users to decide to not take action on a posting. Although this study used a convenience sample ( $n = 55$ ), these findings are promising as they suggest that, when present, message popup warnings may help in deterring the spread of misinformation.

This interactive survey evaluated the action or inaction of user choice response, and it is this physical action which is responsible for the spread of misinformation on social media. Some studies have shown that accuracy and credibility are affected by warnings, but did not evaluate user choice responses (Geeng et al., 2020; Kirchner & Reuter, 2020). Other studies limited participants to a single user choice response, sharing, when faced with true or false postings (Colliander, 2019; Pennycook et al., 2020).

Users may still choose to like, share, or comment even when they know the posting is not accurate. Studies have suggested that users may choose to disseminate known false news for ideological and entertainment purposes (Allcott & Gentzkow, 2017; Ardèvol-Abreu et al., 2020). Correspondingly, during the pilot test's think-aloud process, participants commented that even though the posting was not true, they found the posting to be humorous and wanted to share it with a friend or family member. Based on this observation, future research could explore if there is a correlation between credibility and effectiveness. Also, future research may choose to differentiate among the different user choice responses of like, share, or comment.

In this study, the language in the message popup warnings was kept static. Fan et al. (2018) noted that user behaviors may vary, based on the wording on the warning. Future studies could investigate effectiveness (user choice response) for message popup warnings with text variation.

### *Credibility*

The third research question, RQ3, focused on the credibility of the posting. Participants were asked to evaluate the credibility of each posting. The aim was to determine whether the appearance of message popup warnings affected credibility ratings. The data indicates that there was a significant difference in credibility ratings when comparing postings with a message popup warning to postings without the warning ( $p = .015$ ). This finding suggests that users tend to doubt the credibility of a posting when a popup warning is presented, as was suggested in Ardèvol-Abreu et al.'s (2020) qualitative study. Other studies have found similar decreases in credibility when users were presented with a warning label (Kirchner & Reuter, 2020; Koch et al., 2021; Pennycook et al., 2020). From the think-aloud observations during pilot testing, the researcher noted that some participants, after seeing a message popup warning, commented that they thought the posting was false and that the warning confirmed their suspicions.

The comparison of the five posting types (Treatment False-Warning, Treatment False-No Warning, Treatment True, Control False, and Control True) found significant differences in credibility ( $p < .01$ ). With further analysis, significant differences were found between Treatment False-Warning and Treatment True ( $p < .01$ ) and also Treatment False-Warning and Control True ( $p < .01$ ). This finding is important as it suggests that message popup warnings are more believable than a true posting with no tag or warning. The participants were prone to trust the popup warning more than their own judgment.

These findings suggest that message popup warnings are effective in debunking false news; participants tended to trust that the warning was accurate and they found the messaging credible. Users do not always trust fact-checkers; some believe that these independent fact-checking organizations are biased or intentionally participating in fraud (Ardèvol-Abreu et al., 2020; Brandtzaeg et al., 2018). Unlike studies that included political postings, this study used non-political news stories. This is an important differentiator; individuals tend to read political news stories with established pre-existing beliefs. In this study, participants may not have had pre-existing beliefs about the non-political news stories contained in the interactive survey. Without pre-existing beliefs, the message popup warnings were found to be credible and believable. It is unknown whether message popup warnings would be found credible when pre-existing beliefs influence credibility, as users tend to trust postings that confirm pre-existing beliefs (Spezzano et al., 2021). Future studies may choose to evaluate credibility when pre-existing beliefs are present.

#### *Usability and Usefulness*

The final research question, RQ4, reflected on the usability and usefulness of message popup warnings as a debunking technique. Users in the treatment group completed the SUS to measure usability and the PUMS to evaluate usefulness. The results suggest that most participants found message popup warnings were somewhat useful (fair) for identifying misinformation. As the overarching research question focused on deterring the spread of misinformation, this usefulness rating supports the goal of this study since most participants found the message popup warnings somewhat effective for the detection of misinformation.

However, the usability results were less favorable, with usability scales indicating popup warnings were not highly usable, as they were rated unacceptable, OK, and an “F” by various SUS interpretation scales. This usability finding is in line with Abdulin and Billman’s (2016) study which stated that users found popups to be annoying and disruptive. As these participants are frequently online and have previously encountered popup warnings, usability may have been evaluated on annoyance, not on the knowledge needed to operate popups. Future studies which include multiple debunking techniques or other participant age groups may provide further insight into usability of message popup warnings.

### **Strengths, Weaknesses, and Limitations**

Developing and rolling out an interactive survey (ISS) was a strength of this study. The Qualtrics platform allowed for interaction with individual postings. Thus far, studies regarding debunking techniques have avoided interactive surveys, as these surveys are more difficult to produce. The hotspot functionality in Qualtrics provided users with a clickable interface, similar to affordances found in Facebook’s platform. Additionally, this study provided the full content for each posting, using hyperlinks within the interactive Qualtrics platform. To date, offering additional information to users to assess credibility is understudied. The clickable hyperlinks connected to a separate website, which kept the participants within the study’s environment, and allowed easy return to the survey questions.

The social media platform choice was an unexpected weakness to this study. In the planning stages, Facebook was chosen as the platform for the ISS. As Facebook is the most popular social media platform with over three billion users (Dixon, 2023a),

Facebook was expected to be the best platform choice for the survey. During pilot testing, the researcher used the Think-Aloud Protocol (TAP) for obtaining feedback about the survey. Several participants made comments during the pilot test that they would take certain actions (like, share, or comment) if the platform was Instagram, but would not take those actions on Facebook. Dixon (2023a) explains that Instagram is the preferred choice for younger users, as 57.6% of Instagram in the U.S. users are 34 years of age or younger. Comparatively, only 45.1% of Facebook's users are in the same age group (Dixon, 2023b).

This study was limited by the sample population and the small sample size. The sample was restricted to students at Wentworth Institute of Technology and only 109 participants completed the ISS. As such, the results from this study cannot be generalized. Furthermore, this sample was primarily of traditional college-age students, and results may have been influenced by the age of the participants.

The ISS asked participants to self-report the degree to which they read the articles. While observing participants during the pilot test, the researcher noted that these self-reported rankings were not always accurate. Without collecting data on user clicks or time on task, the self-reported answers cannot be verified; however, data triangulation could confirm the veracity of the user engagement data.

Additionally, this study may have been influenced by a main testing effect. Participants in the treatment group received the effectiveness question twice, both before and after the message popup warning. Although data analysis found a significant difference between user choice responses for the effectiveness questions, it is likely that

some of the participants chose to keep their answers consistent affecting the results of the ISS (Sekaran & Bougie, 2016).

Finally, participant compensation was a weakness to this study. Each participant who completed the ISS was compensated with a \$10 Amazon gift card. While the gift card was important for participant recruitment, some participants took the survey very quickly or attempted to take the survey more than once. Removing anonymity or lessening the compensation amount may alleviate the weakness caused by the compensation.

### **Implications**

This study has implications for current and future social media platforms. In recent times, it has become more popular for individuals to obtain news from social media platforms, rather than traditional news sources. As such, social media platforms have the responsibility to provide content, while helping users to differentiate true and false postings. If believed, false postings can range from benign to dangerous for individuals and groups. In Pennycook et al.'s (2020) study, the researchers tested tagging all fact-checked postings with labels, marking them either "true" or "false". Social media platforms may consider adding warnings or labels on all fact-checked postings, regardless of the posting's veracity. Providing additional information to users, whether it is a warning, link to a fact-checking site, or link to the full content of the posting, may be necessary for users to determine the credibility of the posting.

Social media platforms may want to experiment with message popup warnings to alert users of fact-checked false postings. Testing may include variations in the warning's

wording and operation. Additionally, platforms should explore including fact-check links within the warning and experiment with the timing of the popup's appearance.

Social media platforms should attempt to educate users about unchecked postings. Postings without a warning or label may appear to be true, if some of the more obviously false content is labelled as false (Pennycook et al., 2020). Moravec et al.'s (2020) study experimented with brief training messages within social media and found that this education had significant effects for increased effort (cognition) in credibility decisions. Social media platforms and researchers often assume that users understand how to use social media's affordances, however, this assumption may not be correct (Moravec et al., 2020).

Additionally, algorithms populate newsfeeds based on other's actions, the user's own actions, and truthfulness of the posting (Geeng et al., 2020). User engagement or action with a false posting will cause the posting and similar postings to appear higher in the user's newsfeeds (Ardèvol-Abreu et al., 2020). False postings spread quickly as users are less proficient at detecting false postings as they believe (Moravec et al., 2020). For those reasons, social media platforms may consider adjusting algorithms to devalue known false postings causing these postings to appear lower on users' newsfeeds.

This study has implications for researchers. First, no known studies exist with interactive interfaces and allow for user choice response and engagement (reading the full content) with the posting. As technology has rapidly advanced, interactivity is now possible without programming skills. Second, the relationship between the constructs of credibility and effectiveness needs further exploration. Extant studies have co-mingled the use of "credibility" and "effectiveness" when the terms are quite different from one

another (Kirchner & Reuter, 2020). Researchers would benefit from standard terminology in this field. Third, research should be extended to include both politically charged postings and non-political postings. Pre-existing beliefs and other motivations may influence user engagement levels, user choice responses, and credibility ratings. Fourth, slowing user responses to message popup warnings and other labels is worthy of continued research. Responses to message popup warnings can become routine over time, losing their effectiveness (Wash & Lampe, 2012; Wogalter & Mayhorn, 2005). Slowing down users to take the time to fully read and comprehend postings allows users to be more engaged in decision making (Argo & Main, 2004). Finally, researchers should experiment with varying appearances of postings, warning designs, and warning types to slow down users for better quality decisions.

### **Recommendations for Future Research**

This study represents opportunities for future research. As this study faced several limitations and weaknesses, additional research can extend the body of knowledge on misinformation detection and deterrence. Suggestions for extending this study are listed below:

First, as the sample was primarily college-age students, this research could be expanded to include other age groups. Other age groups may respond differently (user choice behaviors) to message popup warnings, have increased or decreased levels of engagement with the full content contained in the posting, or possess wide-ranging levels of prior knowledge for assessing the credibility of a posting. Diverse age groups may respond differently to the postings simply due to the Facebook-like platform used in the survey. Additionally, usability and usefulness ratings may vary with other age groups.



Second, future studies should consider adding postings that look like Instagram or another social media platform preferred by younger users. Research might compare a control group with a treatment group with Facebook-like scenarios and a second treatment group with Instagram-like scenarios. It would be interesting to note how users respond to postings based on platform differences.

Third, reliability of the self-reported engagement ratings could be verified by capturing another data point, such as time-on-task. If the ISS captured time-on-task, measuring from the time when the user opens the full content of the posting until advancing to the next question, the time stamp could be analyzed alongside the self-reported engagement question. Data triangulation would strengthen the veracity of the findings (Sekaran & Bougie, 2016).

Fourth, future studies may want to differentiate between the type of user choice response. User responses of “like”, “comment”, and “share” trigger different responses from algorithms, and do not appear in users’ newsfeeds equally (Kim & Yang, 2017). On Facebook, “share” offers multiple choices to disseminate a posting to all contacts, select groups, a single contact, or place on their own or a contact’s profile. Although choosing a user response may seem like a subtle difference to the user, each user choice response triggers unique actions within the platform (Kim & Yang, 2017). Future studies may analyze the user choice responses with predetermined weightings, for the degree in which each user choice will cause the false news to propagate.

Fifth, future studies could expand to include different messaging on the popup warning. By experimenting with different messaging, studies could compare the constructs of user engagement, credibility, and effectiveness. Fan et al.’s (2018) study

found that users found detailed warning messaging more useful. While studies have focused on the credibility or accuracy of postings, user choice responses (effectiveness) are understudied to date.

Sixth, during the pilot test phase, users anecdotally indicated that there is no correlation between credibility and effectiveness. Participants commented that they wanted to share postings, even though they believed them to be false. However, results from this study suggest that message popup warnings affected the credibility of the postings, as well as affected the user choice response. Future studies should explore whether there is a correlation between credibility and effectiveness regarding message popup warnings.

Finally, future research should consider extending this study to a survey that contains political news postings. As pre-existing beliefs may affect the effectiveness and credibility of a posting (Fan et al., 2018; Kirchner & Reuter, 2020; Pennycook et al., 2020), extending this study to politically charged postings would test message popup warnings with pre-existing viewpoints. This study focused on misinformation (unintentional false postings), whereas a future study may find different results with disinformation (intentional false postings).

## **Summary**

Misinformation in social media is a vexing problem. Social media platforms and researchers alike have attempted to debunk false postings and prevent its spread. Platforms have experimented with multiple debunking techniques, such as fact-checking links, placing false postings lower in users' newsfeeds, labels and warning text, and confirmation popup warnings (Meta, n.d.; Twitter, 2021). Newsfeeds are curated by

algorithms (Kirchner & Reuter, 2020); algorithms use the users' interests, online contacts, and detection of false news to uniquely populate each newsfeed (Allcott & Gentzkow, 2017; Ardèvol-Abreu et al., 2020).

To assess message popup warnings as a debunking technique, several constructs were evaluated: user engagement, effectiveness, credibility, usefulness, and usability. Regarding the first construct of user engagement, it has not been included in studies to date, as it requires an interactive interface for users to read the full content contained in the posting. Additionally, studies suggest that user engagement may influence user choice behaviors (Kirchner & Reuter, 2020). The second construct, effectiveness, has various definitions in literature; this study adopted Moravec et al.'s (2020) version, defining effectiveness as the user choice response (clicking or not clicking on like, share, or comment) for each posting. Evaluating effectiveness was the focus of this study, as misinformation is spread when a user takes action on a posting (Allcott & Gentzkow, 2017; Kim & Yang, 2017). The third construct credibility, research has focused on trust and debunking false news. Users tend to trust postings if the poster is an online contact, when posting displays in a trusted group, when posting confirms a pre-existing belief, when posting has a high number of likes or comments, or the posting is viewed multiple times (Ma et al., 2019; Moravec et al., 2020; Spezzano et al., 2021). Next, the fourth construct of usefulness was assessed to determine whether the message popup warnings were helpful and valuable for decision making (Fan et al., 2018). Finally, the fifth construct, usability, was assessed to ensure that users know how the tool (message popup warnings) functions (Wash & Lampe, 2012) and that users respond with cognitive awareness (Wash & Lampe, 2012; Wogalter & Mayhorn, 2005).

Message popup warnings are a traditional interface tool, yet social media platforms have been slow to embrace this technology. Popups can influence user behavior by requiring immediate user action (Wash & Lampe, 2012). Studies have shown that message popup warnings can be useful as a debunking tool (Ardèvol-Abreu et al., 2020) but user responses to these warnings may become automatic (Abdulin & Billman, 2016). Nevertheless, the use of message popup warnings as a debunking technique remains understudied.

The goal of this study was to measure the effectiveness of message popup warnings in the context of the user's choice behaviors of liking, sharing, or commenting on false postings. Consequently, the overarching research question of this study was: Are message popup warnings effective for deterring the spread of misinformation?

To address this question, the study was guided by the following sub-questions:

RQ1: Regarding user engagement, to what extent do users read the full content contained in postings when comparing postings that are presented with or without message popup warnings?

RQ2: Regarding effectiveness, what is the user-choice response (actions or inactions) to postings when comparing postings that are presented with or without message popup warnings? Note: actions (1=liking, sharing, or commenting on a false posting) or inactions (0=not liking, sharing, or commenting on a false posting).

RQ3: Regarding credibility, how does the debunking technique of message popup warnings influence the users' credibility rating of the posting, compared to ratings given for postings without message popup warnings?

RQ4: How do users rate the usability and usefulness of message popup warnings as a debunking technique?

To help answer research questions 1-3, the following hypotheses were developed:

H1: Regarding user engagement, there will be no significant differences when comparing the extent users read the full content contained in postings that are presented with or without message popup warnings.

H2: Regarding effectiveness, there will be no significant differences in user choice response (actions or inactions) to postings, when comparing postings that are presented with or without message popup warnings.

H3: Regarding credibility, there will be no significant differences in how users rate the credibility of the posting when receiving the message popup warning, compared to ratings given for postings without message popup warnings.

This study employed a between-subjects experimental design to address these research questions. Prior to the interactive scenario and survey (ISS) rollout, two experts and 11 pilot test participants tested the survey for validity, reliability, and functionality. The ISS was created to allow participants to view and interact with true and false news postings. Participants were randomly assigned to either the control or treatment group. The treatment group received message popup warnings on half (six) of the false postings and made a user choice response (click or not click on like, share, or comment) both before and after the appearance of the warning. Additionally, the treatment group was administered the PUMS to measure usefulness, and the SUS to measure usability of the message popup warnings. All the instruments (ISS, PUMS, and SUS) used in this study collected quantitative data.

Participants were recruited from the WIT campus; 109 qualified participants completed the survey. The sample ( $N = 109$ ) was distributed nearly equally between the treatment (55 participants) and the control (54 participants) groups. In addition to the listed instruments, demographics, social media usage, and frequency of post interaction were collected. User engagement and credibility were analyzed using  $t$ -tests and ANOVA tests and effectiveness was analyzed using McNemar and Chi-square tests. The SUS and PUMS were analyzed using their respective scoring guidelines.

The results from the ISS were able to answer research questions #1-4 and hypotheses #1-3. RQ1 and H1 focused on user engagement and the data indicated that message popup warnings had no effect on the users' likelihood to read the full content contained in the posting. For RQ2 and H2, the message popup warning's effectiveness was analyzed; this finding suggests that there was no difference in user choice response between postings with message popup warnings in the treatment group and postings without message popup warnings in the control group. However, within the treatment group, the data analysis found significant differences in user choice response when message popup warnings were presented for only some postings and when the effectiveness question was posed both before and after the warning. Regarding R3 and H3, the data analysis found significant differences in credibility when comparing postings with message popup warnings and those without warnings, suggesting that credibility of the posting is lessened when the message popup warnings appear. For RQ4, the results of the PUMS suggest that most participants found the message popup warnings somewhat useful in the detection of false postings. However, data from the SUS suggests that the usability of the warnings is lacking. The overall findings indicate that message popup

warnings positively affect credibility and effectiveness (user choice response) but have no effect on user engagement.

This study has implications for academics and social media platforms. The findings suggest that message popup warnings may be effective for credibility of the message and deterring the spread of misinformation when warnings appear on some of the postings. For academics and researchers, this study represents opportunities to embrace interactive technologies in future studies. Social media platforms may want to experiment with labelling all fact-checked postings as true or false. Also, message popup warnings could be tested with a variation of wording, operation, and timing. Additionally, platforms may want to post brief training messages to ensure users know how to detect false postings. Furthermore, algorithms could be adjusted to give lower priority to known false postings, despite user engagement or interactivity.

Recommendations for future research include involving other age groups, adding postings from other social media platforms, differentiating between the type of user choice responses, and exploring different messaging on the popup warnings. As the sample was mostly comprised of college-age students, comparing those results with participants in other age groups may influence the results. Younger age groups prefer social media platforms other than Facebook, and by varying the age group and the social media platform used in the survey, future studies may yield important discoveries. Moreover, user choice responses do not spread misinformation equally (Kim & Yang, 2017). If focused on specific user choice responses, future studies may evaluate effectiveness with a rating system, giving the choices of like, share, or comment different weights. Additionally, future studies may want to explore a variety of messages within

the popup; research has suggested that the wording within the popup may also influence user behavior (Fan et al., 2018; Papakyriakopoulos & Goodman, 2022).



## Appendix A –Interactive Scenario and Survey (ISS)

Image credit can be found in Appendix C

### Welcome Survey

#PS1 - Pre-Screening questions:

Thank you for agreeing to participate in this research study. Your responses will help in the fight against false news.

Do you have a social media account?

Yes  No

Have you liked, shared, or commented on a news article on social media?

Yes  No

[→](#)

#End1 - If both answers were not “yes”, then the next screen appeared:

We thank you for your time spent taking this survey.

#IC1 - If both answers were “yes”, the survey continued with the Informed Consent form. Entire form is shown in Appendix B, this is the top half of the form in the window below:

**Participant Letter for Anonymous Surveys**  
**NSU Consent to be in a Research Study Entitled**  
*Detering the Spread of Misinformation*

**Who is doing this research study?**

The person doing this study is Hollis Greenberg with the College of Computing and Engineering at NSU. She will be helped by Dr. Laurie Dringus.

**Why are you asking me to be in this research study?**

You are being asked to take part in this research study because your opinions will help shape how social media users are informed of false postings.

**Why is this research being done?**

The purpose of this study is to understand the effectiveness of various social media techniques used to alert users to false news. False news postings spread more quickly than truthful postings. This study will add to the body of knowledge to help deter the spread of misinformation.

**What will I be doing if I agree to be in this research study?**

You will be taking a one-time, anonymous survey. The survey will take approximately 25 minutes to complete.

**Are there possible risks and discomforts to me?**

#SurveyAccess1 - If “I disagree” was chosen, the end of survey message appeared. If “I agree” was clicked, the control or treatment survey was randomly chosen. The user was given the random link to one of the surveys:

Click [\\*here\\*](#) to access the survey.

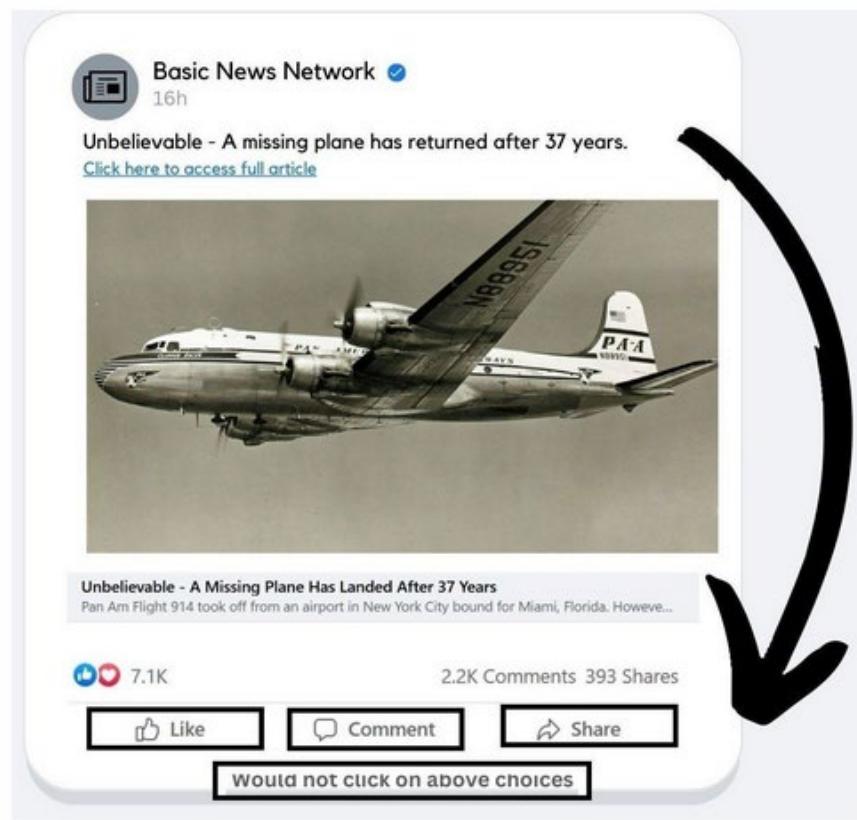
## Treatment ISS

#T-Instruction - Instruction screen:

Instructions: You will be presented with a series of news stories from 2022 and 2023 (24 in total.) We are interested in two things: (1) whether you think the articles are accurate or not, and (2) whether you would be willing to like, share, or write comments on the article on **any** social media platform, such as Facebook, Instagram, TikTok, LinkedIn, etc. Fact-checking has not been performed on all of the articles in this survey.

If you would like to read the full article, click on the [link](#) above the posting.

Below is an **example** of a posting. At the bottom of each social media posting, click on ONE of the buttons ("Like", "Comment", "Share", or "Would Not Click on Above Choices"):



The image shows a screenshot of a Facebook post. At the top left is the profile picture and name "Basic News Network" with a verified badge and "16h" indicating the post's age. The main text of the post reads "Unbelievable - A missing plane has returned after 37 years." followed by a blue link "Click here to access full article". Below the text is a photograph of a Pan Am Lockheed Constellation aircraft in flight, with the tail number "N89951" and "PAA" visible. Underneath the photo is a caption: "Unbelievable - A Missing Plane Has Landed After 37 Years" and a truncated description: "Pan Am Flight 914 took off from an airport in New York City bound for Miami, Florida. Howeve...". The post shows engagement metrics: 7.1K reactions (represented by thumbs up and hearts icons), 2.2K Comments, and 393 Shares. At the bottom, there are four buttons: "Like", "Comment", "Share", and "Would not click on above choices". A large black arrow points from the right side of the post towards the "Would not click on above choices" button.

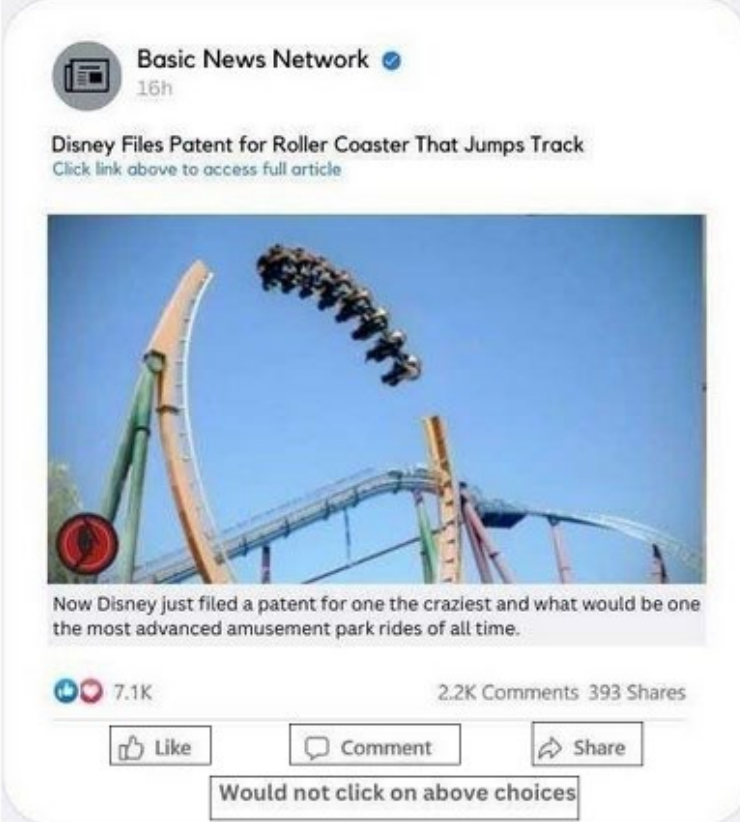
The 24 postings were randomly ordered; 12 were false and 12 were true. Only six of the false postings presented with a popup.

Scenario #1 (False posting with a message popup warning)

#1.1 - False posting (pre-treatment effectiveness question)

If you were to see this article on Facebook or another social media platform, would you consider liking, sharing, or commenting on it? Click your choice on the picture below.

To view the full article's text, click [here](#).



The screenshot shows a Facebook post from 'Basic News Network' (verified, 16h old). The post title is 'Disney Files Patent for Roller Coaster That Jumps Track' with a sub-link 'Click link above to access full article'. The image shows a roller coaster car in mid-air, having jumped its track. The caption reads: 'Now Disney just filed a patent for one the craziest and what would be one the most advanced amusement park rides of all time.' Engagement stats show 7.1K likes, 2.2K comments, and 393 shares. At the bottom, there are buttons for 'Like', 'Comment', and 'Share'. A text box below these buttons contains the text 'Would not click on above choices'.



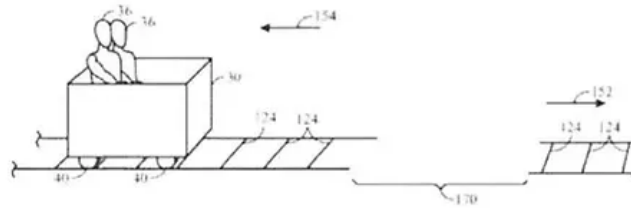
#1L - The hyperlink directed users to this webpage:

*Close this window to return to the survey.*

## Disney Files Patent for Roller Coaster That Jumps Track

Some of the most innovative theme park attractions can be found at the Disney Parks. Soarin' was the first of its kind. Disney is also home to Tower of Terror, one of the most impressive rides ever made. Now Disney just filed a patent for one of the craziest and what would be one of the most advanced amusement park rides of all time.

Walt Disney Imagineering (WDI) just filed a patent for a roller coaster that jumps the track. This means the rollercoaster will fly through the air during part of the ride, with no track under it. The classic video game Thrillville: Off the Rails had a roller coaster that jumps the track on the cover of the game. This was a crazy idea that no one actually thought could work, but now Disney is trying to make that dream a reality.



The roller coaster in the patent Disney filed completely jumps off the track and lands on another track an unspecified distance away. This is the truest attempt at making a jumping coaster. Universal Studios filed for a similar patent that uses the idea of a roller coaster jumping the track, but in their patent, the coaster is not actually flying through the air like Disney's patent describes.

The Universal roller coaster only appears to jump the track. The coaster runs on a track underneath that the guests can't see. This is similar to how Test Track cars run on wheels under the track. The actual car wheels that guests see are primarily for aesthetics, not the main driver. Despite that, Test Track has recently been closed due to failed emissions tests.

According to an internal source, the Disney patent does not guarantee the attraction will be built. They are merely protecting the idea in case they decide to build it. That said, Disney Imagineers are actively pursuing a roller coaster that jumps the track, which they have a working jump coaster model under strict confidentiality.

As you may imagine, a roller coaster of this caliber is not easy to create and requires extensive calculations and engineering. Our source mentioned that they rode the jump coaster model and were terrified, though they said it was an exhilarating experience. In order for this to become a reality however, Disney has rigorous testing to do and it's not expected to be ready for the public until 2030 at the earliest, if it even happens.

[Return to top of page](#)

#1.2 - After user clicked on “Like”, “Share”, “Comment”, or “Would not click on above choices”, the popup and post-treatment effectiveness question appeared.

If you were to see this article on Facebook or another social media platform, would you consider liking, sharing, or commenting on it? Click your choice on the picture below.

To view the full article's text, click [here](#).



#1.Q1 - After user clicked on “Like”, “Share”, “Comment”, or “Would not click on above choices”, the engagement question appeared:

Did you read the previous article?

No, not at all	Read little of article	Read somewhat	Yes, read fully
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

#1-Q2 - Followed by the credibility question:

To the best of your knowledge, how accurate is this article?

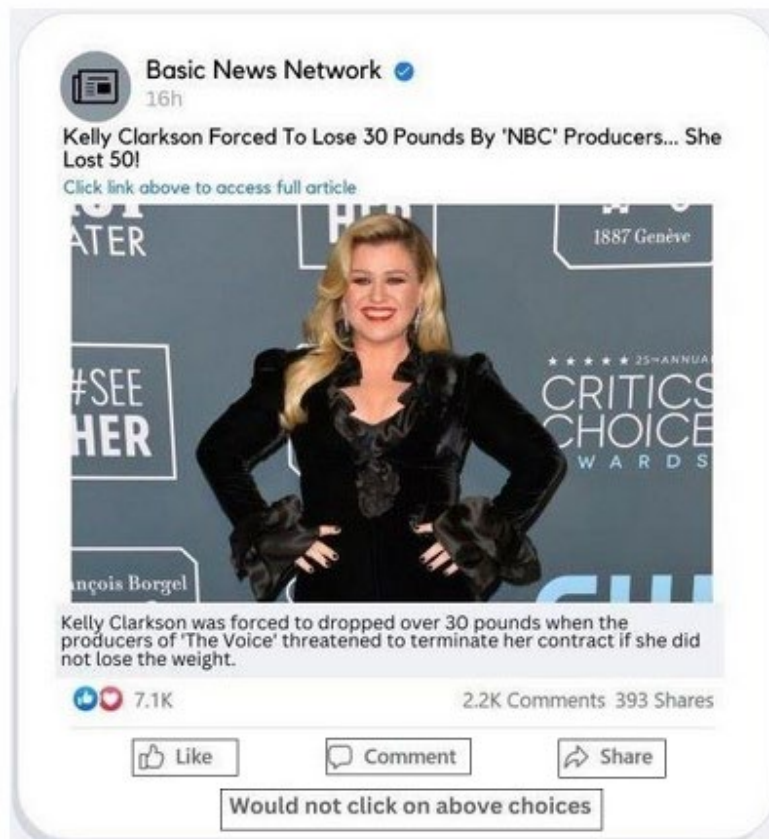
Not at all accurate	Not very accurate	Somewhat accurate	Very accurate
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Scenario #7

## #7.1 – False posting with no warning

If you were to see this article on Facebook or another social media platform, would you consider liking, sharing, or commenting on it? Click your choice on the picture below.

To view the full article's text, click [here](#).





#7L - The hyperlink directed users to this webpage:

*Close this window to return to the survey.*

## **Kelly Clarkson Forced To Lose 30 Pounds By 'NBC' Producers... She Lost 50! (Her Diabetes Finally In Control)**



*Best known as a coach for *The Voice*, hosting the recent *Billboard Music Awards* and a mom of four, Kelly Clarkson was forced to dropped over 30 pounds when the producers of *'The Voice'* threatened to terminate her contract if she did not lose the weight.*

Already being approved for the 21th season producers of 'The Voice' want Kelly Clarkson, to lose 30 pounds to accommodate her new role as head coach. As awful as it is to force a woman to lose weight, Kelly is actually contractually obligated to follow these requirements. The show's producers said they would support her through the time but if she did not meet the goal her contract would be terminated.

Frustrated with such an insane goal to meet in such little time Kelly actually got her lawyer involved but to no avail. Luckily for Kelly, the always generous Dr. Oz had caught wind of the ordeal and offered an easy solution for weight loss..

She believed she hit "rock bottom" in March and decided she had to change her ways. "I was tired, puffy, and desperate ... and I knew I shouldn't be feeling that way," she wrote. "Though I had gained considerable weight over the past three years and I definitely wanted to slim down for Alex's wedding in May, what motivated me the most was just wanting to feel better and have more energy." "When Dr. Oz called and gave me advice on how to lose the weight and stick it to those producers I was ecstatic." said Kelly. "He pulled some strings and gave me a product that burned the fat in the 22 days as promised. I couldn't believe it, I owe him my life!"

#7.Q1 - After user clicked on “Like”, “Share”, “Comment”, or “Would not click on above choices”, the engagement question appeared:

Did you read the previous article?


No, not at all <input type="radio"/>	Read little of article <input type="radio"/>	Read somewhat <input type="radio"/>	Yes, read fully <input type="radio"/>
---	---	--	--



#7.Q2 - Followed by the credibility question:

To the best of your knowledge, how accurate is this article?

Not at all accurate <input type="radio"/>	Not very accurate <input type="radio"/>	Somewhat accurate <input type="radio"/>	Very accurate <input type="radio"/>
--	--	--	--



## Scenario #13

## #13.1 - True posting

If you were to see this article on Facebook or another social media platform, would you consider liking, sharing, or commenting on it? Click your choice on the picture below.

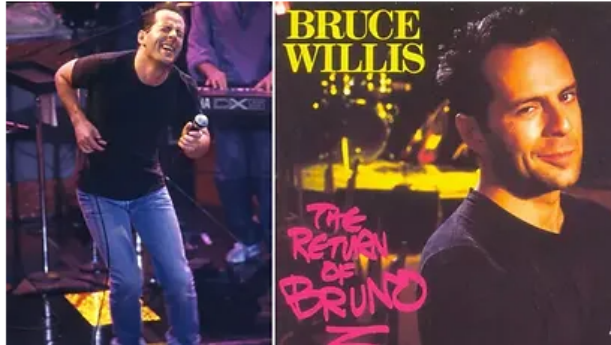
To view the full article's text, click [here](#).



#13L - The hyperlink directed users to this webpage:

*Close this window to return to the survey.*

## Bruce Willis Released an R&B Album Under the Name Bruno Radolini



In the late 1980s Bruce Willis started his career with a rockstar alter ego called Bruno Radolini.

We all know Bruce Willis for his roles in huge Hollywood blockbusters like Die Hard and The Sixth Sense. But you might not remember he actually tried his hand at singing back in the late 1980s.

In fact, the 67-year-old even had his own rockstar alter ego who went by the name of Bruno Radolini.

Yep, in 1987 Bruce released an album called The Return of Bruno which had 10-tracks and featured famous musicians like the Pointer Sisters and the Temptations.

It was an R&B album and peaked at an impressive number 14 in the Billboard 200 album chart thanks to his popular single Respect Yourself, which reached number five in the charts.

The song was originally performed by The Staples Singers in 1971, with Bruce remaking it with June Pointer.

He later released follow-up single Young Blood, but this didn't fare as well, reaching number 68 in the US charts. However, his track Under the Boardwalk was very popular in the UK and managed to get to number two.

As well as releasing his first album, Bruce also played the character of Bruno in an hour-long mockumentary which aired the same year.

The movie featured music superstars like [Elton John](#), [Ringo Starr](#), and [Brian Wilson](#), all of whom talk about Bruno's influence on their music.

It also featured footage of Bruno playing various concerts throughout the 1960s, 1970s, and 1980s, as well as interviews with his former bandmates.

Around the same time The Return of Bruno came out, Willis signed a deal with Seagram's Golden Wine Coolers, which would result in a series of adverts where he can also be heard singing.

#13.Q1 - After user clicked on “Like”, “Share”, “Comment”, or “Would not click on above choices”, the engagement question appeared:

Did you read the previous article?

No, not at all	Read little of article	Read somewhat	Yes, read fully
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

#13.Q2 - Followed by the credibility question:

To the best of your knowledge, how accurate is this article?

Not at all accurate	Not very accurate	Somewhat accurate	Very accurate
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## #SUS – SUS was presented

We would like to evaluate the usability of the disputed warning messages seen in this survey. Please choose the best answer below:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I think that I would like to use this tool frequently.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the tool unnecessarily complex.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought the tool was easy to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think that I would need the support of a technical person to be able to use this tool.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the various functions in this tool were well integrated.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought there was too much inconsistency in this tool.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would imagine that most people would learn to use this tool very quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the tool very cumbersome to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



#PUMS – PUMS was presented

We would like to evaluate the usefulness of the disputed warning messages seen in this survey. Please choose the best answer to the statements below:

	Extremely likely	Moderately likely	Slightly likely	Neither likely nor unlikely	Slightly unlikely	Moderately unlikely	Extremely unlikely
Using message popup warning on social media would enable me to detect false news more quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using message popup warnings would improve my performance in detecting false news.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using message popup warnings would increase my productivity on social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using message popup warnings would enhance my effectiveness on social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using message popup warnings would make it easier to use social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would find message popup warnings useful on social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[→](#)

## Demographics:

### #Gender

To which gender do you most identify?

Female

Male


Transgender female

Transgender male

Gender variant/Non-conforming

Not listed

Prefer not to answer



### #Race

Choose one or more races that you consider yourself to be

White or Caucasian

Black or African American


American Indian/Native American or Alaska Native

Asian

Native Hawaiian or Other Pacific Islander

Other

Prefer not to say





## #Age

How old are you?

Under 18

18-24 years old


25-34 years old

35-44 years old

45-54 years old

55-64 years old

65+ years old



## #SMFreq

How often do you use social media?

Less than monthly

1-2 times a month

Once a week

2-3 times a week

Daily

## #EngageFreq

How often do you click on like, share, or comment on a news posting in social media?

Less than monthly

1-2 times a month


Once a week

2-3 times a week

Daily

#Email – To send compensation to participants, the following screen displays:

If you want to be receive a \$10 Amazon gift card for completing the survey, type your email address below:



#End2 – End of survey screen

Misinformation is tricky to detect. This survey aims to find effective methods for people to detect false news. All actions (clicking on like, share, or comment) will make the posting appear in your contacts' newsfeeds. If the posting is untrue, it is best to not like, share, or comment.

The following postings were false:

- Disney Files Patent for Roller Coaster That Jumps Track
- Nostradamus predicted that Charles will abdicate and 'mystery king' could replace him
- Obvious Mistakes In Movies Which You Might Have Missed
- Elon Musk Shocks the World with Plans for Twitter's New HQ on Alcatraz Island
- The View Settles With The Rittenhouses For \$22 Million and a Formal Apology
- Unbelievable News: Dr. Oz Wins Nobel Prize in Medicine
- Kelly Clarkson Forced To Lose 30 Pounds By 'NBC' Producers... She Lost 50! (Her Diabetes Finally In Control)
- Bill Gates Arrest Warrant Issued in Philippines For 'Premeditated Murder' Linked To Vaccine Roll Out
- Disney World Leaves Florida For Georgia
- Budweiser Brands Won't Be at Oktoberfest for the First Time in 75 Years
- Chevy Corvairs Have a 'Dash Baby Cradle'
- Kelly Ripa is Leaving 'Live with Kelly and Ryan' to Sell Skincare Products

The remainder of articles were true.

***Thank you for taking the time to fill out this survey!***


## Control Group ISS

#C-Instruction - Instruction screen:

Instructions: You will be presented with a series of news stories from 2022 and 2023 (24 in total.) We are interested in two things: (1) whether you think the articles are accurate or not, and (2) whether you would be willing to like, share, or write comments on the article on **any** social media platform, such as Facebook, Instagram, TikTok, LinkedIn, etc. Fact-checking has not been performed on all of the articles in this survey.

If you would like to read the full article, click on the [link](#) above the posting.

Below is an **example** of a posting. At the bottom of each social media posting, click on ONE of the buttons ("Like", "Comment", "Share", or "Would Not Click on Above Choices"):



The image shows a screenshot of a social media post from "Basic News Network" (verified account, 16h old). The post features a black and white photograph of a Pan Am Lockheed Constellation aircraft (N98951) in flight. The text of the post reads: "Unbelievable - A missing plane has returned after 37 years. [Click here to access full article](#)". Below the photo, the post is truncated: "Unbelievable - A Missing Plane Has Landed After 37 Years Pan Am Flight 914 took off from an airport in New York City bound for Miami, Florida. Howeve...". The post has 7.1K likes, 2.2K comments, and 393 shares. At the bottom, there are four buttons: "Like", "Comment", "Share", and "Would not click on above choices". A large black arrow points from the text above to the "Would not click on above choices" button. A blue button with a minus sign is visible in the bottom right corner of the screenshot area.

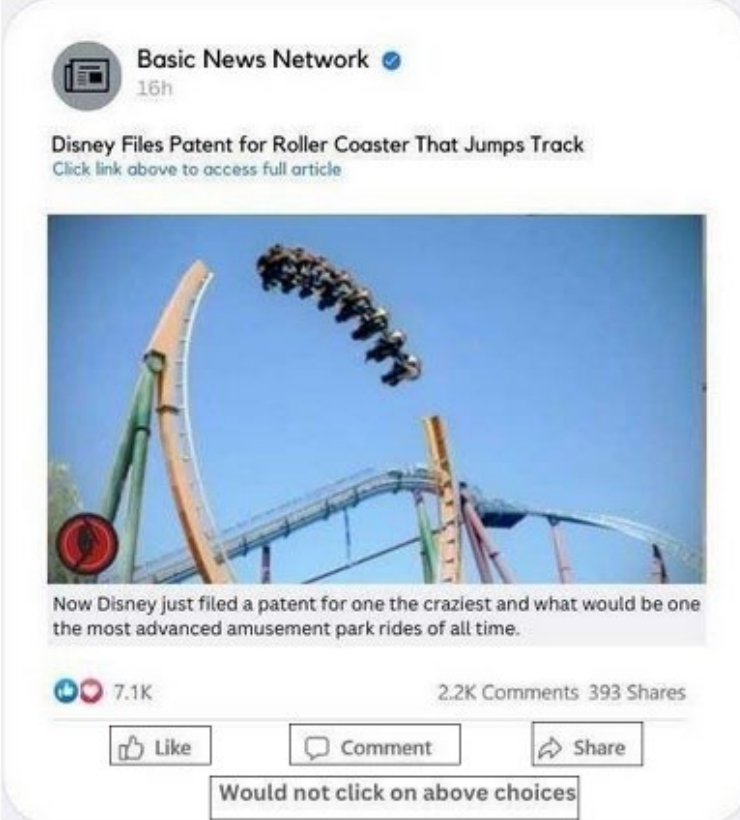
The 24 articles were randomly presented; 12 were false and 12 were true. None of the false postings were presented with a popup warning.

### Scenario #1

#### #1.1 - False posting without a message popup warning

If you were to see this article on Facebook or another social media platform, would you consider liking, sharing, or commenting on it? Click your choice on the picture below.

To view the full article's text, click [here](#).



The screenshot shows a Facebook post from 'Basic News Network' (verified, 16h old). The post title is 'Disney Files Patent for Roller Coaster That Jumps Track' with a link to the full article. The image shows a roller coaster car in mid-air, having jumped off its track. The caption reads: 'Now Disney just filed a patent for one the craziest and what would be one the most advanced amusement park rides of all time.' The post has 7.1K reactions, 2.2K comments, and 393 shares. At the bottom, there are buttons for 'Like', 'Comment', and 'Share'. A text box below these buttons contains the text 'Would not click on above choices'.



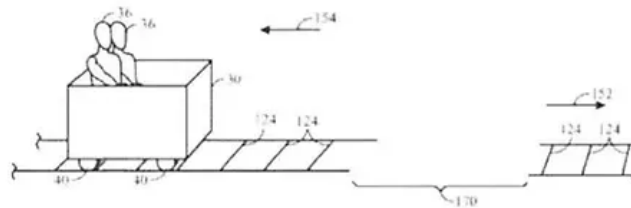
#1L - The hyperlink directed users to this webpage:

*Close this window to return to the survey.*

## Disney Files Patent for Roller Coaster That Jumps Track

Some of the most innovative theme park attractions can be found at the Disney Parks. Soarin' was the first of its kind. Disney is also home to Tower of Terror, one of the most impressive rides ever made. Now Disney just filed a patent for one of the craziest and what would be one of the most advanced amusement park rides of all time.

Walt Disney Imagineering (WDI) just filed a patent for a roller coaster that jumps the track. This means the rollercoaster will fly through the air during part of the ride, with no track under it. The classic video game Thrillville: Off the Rails had a roller coaster that jumps the track on the cover of the game. This was a crazy idea that no one actually thought could work, but now Disney is trying to make that dream a reality.



The roller coaster in the patent Disney filed completely jumps off the track and lands on another track an unspecified distance away. This is the truest attempt at making a jumping coaster. Universal Studios filed for a similar patent that uses the idea of a roller coaster jumping the track, but in their patent, the coaster is not actually flying through the air like Disney's patent describes.

The Universal roller coaster only appears to jump the track. The coaster runs on a track underneath that the guests can't see. This is similar to how Test Track cars run on wheels under the track. The actual car wheels that guests see are primarily for aesthetics, not the main driver. Despite that, Test Track has recently been closed due to failed emissions tests.

According to an internal source, the Disney patent does not guarantee the attraction will be built. They are merely protecting the idea in case they decide to build it. That said, Disney Imagineers are actively pursuing a roller coaster that jumps the track, which they have a working jump coaster model under strict confidentiality.

As you may imagine, a roller coaster of this caliber is not easy to create and requires extensive calculations and engineering. Our source mentioned that they rode the jump coaster model and were terrified, though they said it was an exhilarating experience. In order for this to become a reality however, Disney has rigorous testing to do and it's not expected to be ready for the public until 2030 at the earliest, if it even happens.

[Return to top of page](#)

#1.Q1 - After user clicked on “Like”, “Share”, “Comment”, or “Would not click on above choices”, the engagement question appeared:



Did you read the previous article?

No, not at all

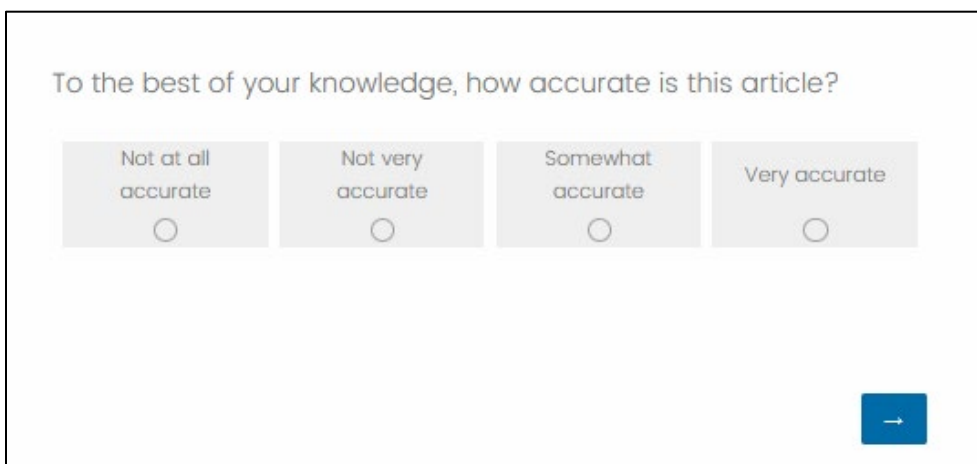
Read little of article

Read somewhat

Yes, read fully

→

#1.Q2 - Followed by the credibility question:



To the best of your knowledge, how accurate is this article?

Not at all accurate

Not very accurate

Somewhat accurate

Very accurate

→

Scenarios #7-24 were identical postings and questions as treatment group (See Treatment ISS above for examples of scenarios 7 and 13.)

Demographic questions (Gender, Age, Race, Frequency of Social Media usage, Frequency of Engagement), Email question, and End of Survey message same as treatment group (See Treatment ISS above.)

## Appendix B– Informed Consent Form



**INSTITUTIONAL REVIEW BOARD**  
3301 College Avenue  
Fort Lauderdale, Florida 33314-7796  
PHONE: (954) 262-5369

**Participant Letter for Anonymous Surveys  
NSU Consent to be in a Research Study Entitled**

*Heed the Warning Signs: The Effectiveness of Message Popup Warnings for Detering the Spread of Misinformation*

**Who is doing this research study?**

The person doing this study is Hollis Greenberg with the College of Computing and Engineering at NSU. She will be helped by Dr. Laurie Dringus.

**Why are you asking me to be in this research study?**

You are being asked to take part in this research study because your opinions will help shape how social media users are informed of false postings.

**Why is this research being done?**

The purpose of this study is to understand the effectiveness of various social media techniques used to alert users to false news. False news postings spread more quickly than truthful postings. This study will add to the body of knowledge to help deter the spread of misinformation.

**What will I be doing if I agree to be in this research study?**

You will be taking a one-time, anonymous survey. The survey will take approximately 25 minutes to complete.

**Are there possible risks and discomforts to me?**

This research study involves minimal risk to you. To the best of our knowledge, the things you will be doing have no more risk of harm than you would have in everyday life.

**What happens if I do not want to be in this research study?**

You can decide not to participate in this research and it will not be held against you. You can exit the survey at any time.

**Will it cost me anything? Will I get paid for being in the study?**

There is no cost for participation in this study. Participation is voluntary. If you provide your email address and complete the survey, you will be emailed a \$10 Amazon gift card.



**INSTITUTIONAL REVIEW BOARD**  
3301 College Avenue  
Fort Lauderdale, Florida 33314-7796  
PHONE: (954) 262-5369

**How will you keep my information private?**

Your responses are anonymous. Information we learn about you in this research study will be handled in a confidential manner, within the limits of the law. There will be no information requested that will link you to your survey answers. This data will be available to the researcher, the Institutional Review Board and other representatives of this institution, and any granting agencies (if applicable). All confidential data will be kept securely on the researcher's hard drive and storage backup. All data will be kept for 36 months from the end of the study and destroyed after that time by deleting all data files.

**Who can I talk to about the study?**

If you have questions, you can contact Hollis Greenberg at 617-989-4305. If Professor Greenberg is unavailable, contact Dr. Laurie Dringus at 954-262-2073.

If you have questions about the study but want to talk to someone else who is not a part of the study, you can call the Nova Southeastern University Institutional Review Board (IRB) at (954) 262-5369 or toll free at 1-866-499-0790 or email at [IRB@nova.edu](mailto:IRB@nova.edu).

**Do you understand and do you want to be in the study?**

If you have read the above information and voluntarily wish to participate in this research study, please click on arrow button below and you will be directed to the survey link.



## Appendix C – ISS Image Sources for Photographs and Drawings

#1.1, 1.2 From video “Rollercoaster Tycoons” by RocketJump (2011), RocketJump (<https://www.youtube.com/user/freddiew>).

#1L From “Disney Files Patent for Roller Coaster That Jumps Track” by Mouse Trap News (30 October 2022), The Mouse Trap (<https://mousetrapnews.com/disney-files-patent-for-roller-coaster-that-jumps-track/>), Copyright 2023 by Apex News. Reprinted with permission.

#2.1, 2.2 , 2L Image: Michael Nostradamus. Line engraving by G. W. Knor Wellcome V0004337.jpg via Openverse. Licensed under CC BY 4.0.

#3.1, 3L, 3.2 From “Obvious Mistakes in Movies Which You might Have Missed” by Raunak Samaiya (19 December 2017), CineTales (<https://www.cine-tales.com/obvious-mistakes-movies-might-missed/>), Copyright 2023 by CineTales.

#4.1, 4L, 4.2 From “Elon Musk Shocks the World With Plans for New Twitter HW on Alcatraz Island” by MJ (n.d.), Architizer, Inc. (<https://architizer.com/blog/inspiration/stories/elon-musk-twitter-hq-alcatraz-island/>). Copyright 2023 by Architizer. Reprinted with permission.

#5.1, 5L, 5.2 From “‘The View’ Hasn’t Settled a Lawsuit with Kyle Rittenhouse” by Ross D. Franklin (27 May 2022) by The Associated Press (<https://apnews.com/article/fact-check-kyle-rittenhouse-the-view-abc-677910767107>). Copyright 2023 by The Associated Press. Reprinted with permission.

#6.1, 6L, 6.2 Image: ID 143307530, Laurence Agron via Dreamstime. Reprinted with permission.

#7.1, 7L Image: ID 169145209, Featureflash via Dreamstime. Reprinted with permission.

#8.1, 8L Image: ID 179611700, F. Legrand via Dreamstime. Reprinted with permission.

#9.1, 9L Image: ID 115827236, Mehmet Guvenc via Dreamstime Reprinted with permission.

#10.1, 10L Image: ID 53703511, Pflugler via Dreamstime. Reprinted with permission.

#11.1, 11L Image: Shasta Mustang Supply via Facebook.

#12.1, 12L Image: ID 194366612, Hutchinsphoto via Dreamstime. Reprinted with permission.

- #13.1, 13L;Image: ID 364629205, RLFE Pix via Alamy; ID: 49085800, David Lichtneker via Alamy. Reprinted with permission
- #14.1, 14L From “Stephen Hawking Pictured on Jeffrey Epstein’s ‘sex slave’ Caribbean Island” by Tim Steward News Limited (13 January 2015) Independent (<https://www.independent.co.uk/news/people/stephen-hawking-pictured-on-jeffrey-epstein-s-sex-slave-caribbean-island-9974955.html>); Copyright 2023.
- #15.1, 15L Image: Jonathan Harford via Flickr. CC BY-NC 2.0, Reprinted with permission.
- #16.1, 16L Image: Honolulu Star Advertiser via Reddit
- #17.1 Image: ID 27157382, Carrielson1 via Dreamstime. Reprinted with permission.
- #17L Screenshot image from “You Slept With My Stripper Sister!” by The Jerry Springer Show (2 Nov 2011). Fair Use.
- #18.1, 18L From “Polish Forces During the Second World War” by Imperial War Museum (1942) United Kingdom Government. (<https://www.iwm.org.uk/collections/item/object/205026420>). Public domain.
- #19.1 Image: ID 159215132, Yannick Morelli via Dreamstime. Reprinted with permission.
- #19L Image: ID 208477334, Wachiwit via Dreamstime. Reprinted with permission.
- #20.1, 20L Image: ID 273197753, Rokas Tenys via Dreamstime. Reprinted with permission.
- #21.1, 21L From “30 February” by Sendker (1 March 1712) Swedish Almanac (<https://tondering.dk/main/calendar-information/15-30-february>). Public Domain.
- #22.1 From “Ford Applies to Patent Self-Repossessing Cars That Can Drive Themselves Away” by Peter Holderith (2 March 2023) The Drive. (<https://www.thedrive.com/news/future-fords-could-repossess-themselves-and-drive-away-if-you-miss-payments>). Reprinted with permission.
- #22L From “Systems and Methods to Repossess a Vehicle” by Maraldo et al., United States Patent Application Publication (23 February 2023), Pub. No. US 2023/0055959 A1
- #23.1, 23L Image: ID 102487439, Denis Makarenko via Dreamstime. Reprinted with permission.
- #24.1, 24L Image: Darryl Brooks via Dreamstime. Reprinted with permission.

## Appendix D– Institutional Review Board Documentation (NSU)



INSTITUTIONAL REVIEW BOARD  
3301 College Avenue  
Fort Lauderdale, Florida 33314-7756  
PHONE: (954) 262-5369

### MEMORANDUM

**To:** Hollis Greenberg, PhD, Information Systems  
College of Engineering and Computing

**From:** Ling Wang, Ph.D.  
College Representative, College of Engineering and Computing

**Date:** July 21, 2023

**Subject:** IRB Exempt Initial Approval Memo

**TITLE:** Heed the Warning Signs: The Effectiveness of Message Popup Warnings for Detering the Spread of Misinformation– NSU IRB Protocol Number 2023-350

Dear Principal Investigator,

Your submission has been reviewed and Exempted by your IRB College Representative or their Alternate on **July 21, 2023**. You may proceed with your study.

*NOTE: Exempt studies do not require approval stamped documents. If your study site requires stamped copies of consent forms, recruiting materials, etc., contact the IRB Office.*

**Level of Review:** Exempt

**Type of Approval:** Initial Approval

**Exempt Review Category:** Exempt 2: Interviews, surveys, focus groups, observations of public behavior, and other similar methodologies

**Annual Status of Research Update:** You are required to notify the IRB Office annually if your research study is still ongoing via the *Exempt Research Status Update xForm*.

**Changes:** Any changes in the study (e.g., procedures, consent forms, investigators, etc.) must be approved by the IRB prior to implementation using the *Amendment xForm*.



**INSTITUTIONAL REVIEW BOARD**  
3301 College Avenue  
Fort Lauderdale, Florida 33314-7756  
PHONE (954) 262-5369

**Post-Approval Monitoring:** The IRB Office conducts post-approval review and monitoring of all studies involving human participants under the purview of the NSU IRB. The Post-Approval Monitor may randomly select any active study for a Not-for-Cause Evaluation.

**Final Report:** You are required to notify the IRB Office within 30 days of the conclusion of the research that the study has ended using the *Exempt Research Status Update xForm*.

**Translated Documents:** No

*Retain this document in your IRB correspondence file.*

CC: Ling Wang, Ph.D.

Laurie Dringus, Ph.D.

## Appendix E– Institutional Review Board Documentation (WIT)



Office of the Provost

August 30, 2023

Hollis Greenberg  
 Associate Professor, Business Management  
 School of Management  
[Greenbergh1@wit.edu](mailto:Greenbergh1@wit.edu)

IRB #: 2023-47

**Project Title:** The Effectiveness of Message Popup Warnings for Deterring the Spread of Misinformation

Dear Professor Greenberg,

This letter officially informs you that the Institutional Review Board has completed its review and approved your project. It is the Board's opinion that you have provided adequate safeguards for the rights and welfare of the participants in this study based on the information provided.

We want to remind you that this project must be conducted in full accordance with all applicable sections of Wentworth's IRB guidelines and DHHS Regulations for the Protection of Human Subjects (45 CFR 46). Additionally, you should notify the IRB immediately of any proposed changes that may affect the project. You should also report any unanticipated problems involving risk to the participants or others to the IRB. Further, if the project continues for more than one year, you must gain approval for continuing review before the end of the first year and each additional year.

Regards,

A handwritten signature in cursive script that reads "Paul Firenze".

Paul Firenze, Ph.D.  
 Chair, Institutional Review Board

650 HUNTINGTON AVENUE  
 BOSTON, MA 02115-6998

617.989.4000  
 WIT.EDU



## Appendix F– Recruitment Flyer

**STUDENT PARTICIPANTS NEEDED**

*For a research study to STOP THE SPREAD OF MISINFORMATION!*

**THE PURPOSE** of this study is to find effective methods to combat the spread of misinformation.

**LOCATION:** remotely through this link  
[https://wentworth.az1.qualtrics.com/jfe/form/SV\\_1GMBdht3V3Lowho](https://wentworth.az1.qualtrics.com/jfe/form/SV_1GMBdht3V3Lowho) or this QR code





**QUESTIONS?**  
 Email Hollis Greenberg (PI) at [greenbergh1@wit.edu](mailto:greenbergh1@wit.edu)

*This is a Nova Southeastern University (NSU) dissertation research study.*

**PARTICIPANTS MUST BE** 18 or older and a student at WIT

**WHAT'S INVOLVED?**  
 Participants will fill out a survey online that will take approximately 25 minutes

**PARTICIPANTS WILL RECEIVE:**  
 Upon completion of the survey, you will receive a \$10 Amazon gift card.

Participation in this survey is completely voluntary. You have the right to leave this research study at any time, or not be in it. If you decide to not complete the study, there will be no penalties and you will not lose any services you have a right to get.

## Appendix G – Think-Aloud Checklist and Form

### Checklist for Think-Aloud Protocol – Pilot Testing

- Welcome and thanks for participating
- Explain how the testing room will work. *(Put user at end of the table facing me with his/her back to the large screen, as to see facial expressions and the enlarged screen at the same time.)*
- Explain purpose of study: *“I am studying misinformation/fake news and how to deter/stop its spread.”*
- Explain: *The Pilot Test is to test the survey itself, not the person taking the survey. You cannot do anything wrong.*
- Explain the testing process and Think Aloud protocol: *All participants will take the survey and will “think out loud”. Let me explain what this means. I realize that is not common to think out loud while working but doing so will help me get insight into your experience. Here’s some examples of things you might say:*
  - *This is not at all what I expected to see when I clicked on that link.*
  - *I have no idea what that word means.*
  - *I sure wish this survey would let me do X here.*

*If you happen to go quiet, which is a perfectly normal thing to do while you’re concentrating, I may ask a question, such as “What are you doing or thinking about now?”*

- Ask: *Do you have any concerns? Remember, you can stop the testing at any time.*
- At END: *Thank you. I will send you a link to the \$20 Amazon gift card today.*

Tester: \_\_\_\_\_

Date: \_\_\_\_\_

Device Used: \_\_\_\_\_

<b>Scenario</b>	<b>Negative/Positive Feedback</b>	<b>Comments</b>



## References

- Abdulin, E., & Billman, D. (2016). Confirmation responses: In-context, visible, predictable design versus popup windows. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2969–2975. <https://doi.org/10.1145/2851581.2892403>
- Ahmed, A. (2020). Twitter is expanding its warning pop-ups for the liking of disputed tweets now. *Digital Information World*. Retrieved Oct 20, 2020, from <https://www.digitalinformationworld.com/2020/11/twitter-is-expanding-its-warning-pop.html#>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2), 211-236.
- Alonzo, I. (2020). Twitter vs fake news: Warning now pops-up when liking tweets with 'misleading' labels—Is this acceptable or a nuisance? *Tech Times*. <https://www.techtimes.com/articles/254432/20201123/twitter-vs-fake-news-warning-now-pops-up-when-liking.htm>
- Anitha, K., Saraswathy, V., Gayathri, K., & Priya, S. S. (2019). An approach to comment analysis in online social media. *Proceedings of the 3rd International Conference on Computing and Communications Technologies (ICCCT)*, 332-335.
- Architizer. (2023). *Elon Musk Shocks the World With Plans for New Twitter HQ on Alcatraz Island*. Architizer, Inc. Retrieved June 10, 2023 from <https://architizer.com/blog/inspiration/stories/elon-musk-twitter-hq-alcatraz-island/>
- Ardèvol-Abreu, A., Delponti, P., & Rodríguez-Wangüemert, C. (2020). Intentional or inadvertent fake news sharing? Fact-checking warnings and users' interaction with social media content. *Profesional de la Información*, 29(5).
- Argo, J. J., & Main, K. J. (2004). Meta-analyses of the effectiveness of warning labels [research-article]. *Journal of Public Policy & Marketing*, 23(2), 193-208. <https://ezproxywit.flo.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.30000760&site=eds-live&scope=site>
- Avram, M., Micallef, N., Patil, S., & Menczer, F. (2020). Exposure to social engagement metrics increases vulnerability to misinformation. *arXiv preprint arXiv:2005.04682*.
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114-123.

- Barnum, C. (2021). *Usability Testing Essentials: Ready, Set...Test!* (2nd ed.). Elsevier Inc.
- Bartram, N. (2022). *The Amazing Story of Bruce Willis' Popstar Alter Ego Bruno Radolini*. Smooth Radio. Retrieved June 12, 2023 from <https://www.smoothradio.com/features/bruce-willis-popstar-alter-ego-bruno-radolini/>
- Beatty, S. E., & Smith, S. M. (1987). External search effort: An investigation across several product categories. *Journal of Consumer Research*, 14(1), 83-95. <https://doi.org/10.1086/209095>
- Bentley, F., Quehl, K., Wirfs-Brock, J., & Bica, M. (2019). Understanding online news behaviors. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Paper 590. <https://doi.org/10.1145/3290605.3300820>
- Bode, L., & Vraga, E. K. (2017). See something, say something: Correction of global health misinformation on social media. *Health communication*, 33(9), 1131-1140.
- Boehm, L. E. (1994). The validity effect: A search for mediating variables. *Personality and Social Psychology Bulletin*, 20(3), 285-293.
- Brandtzaeg, P. B., Følstad, A., & Chaparro Domínguez, M. Á. (2018). How journalists and social media users perceive online fact-checking and verification services. *Journalism practice*, 12(9), 1109-1129. <https://doi.org/10.1080/17512786.2017.1363657>
- Brooke, J. (1995). SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189.
- Brooke, J. (2013). SUS: a retrospective. *Journal of Usability Studies*, 8(2), 29-40.
- Brzozowski, A. (2014). *1960 Chevy Corvair Offered First-in-Class Infant Storage Solution: A Touch of Innovation*. GMAuthority.com. Retrieved June 11, 2023 from <https://gmauthority.com/blog/2014/12/1960-chevy-corvair-offered-first-in-class-infant-storage-solution-a-touch-of-innovation/>
- Chan, M.-P., Jones, S., Hall, C. R., Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531-1546.
- Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., & Morgan, E. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, 42, 1073-1095.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

- Colliander, J. (2019). "This is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media. *Computers in Human Behavior*, 97, 202-215.  
<https://doi.org/10.1016/j.chb.2019.03.032>
- Davis, C. (2023). *Kelly Clarkson Forced to Lose 30 Pounds by 'NBC' Producers... She Lost 50! (Her Diabetes Finally in Control)*. Exclusiveinsiderdeals.com. Retrieved June 9, 2023 from <https://archive.ph/MYUdv>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319-340.
- DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6), 753-773.
- Disney Files Patent for Roller Coaster That Jumps Track*. (2022). Mouse Trap News. Retrieved June 10, 2023 from <https://mousetrapnews.com/disney-files-patent-for-roller-coaster-that-jumps-track/>
- Dixon, S. J. (2023a). *Most Popular Social Networks Worldwide as of October 2023, Ranked by Number of Monthly Active Users*. Statista. Retrieved November 7, 2023 from <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Dixon, S. J. (2023b). *Share of Facebook Users in the United States as of August 2023, by Age Group*. Statista. Retrieved November 7, 2023 from <https://www.statista.com/statistics/187549/facebook-distribution-of-users-age-group-usa/>
- Dmitry, B. (2023). *Bill Gates Arrest Warrant Issued in Philippines for 'Premeditated Murder' Linked to Vaccine Roll Out*. The People's Voice. Retrieved June 12, 2023 from <https://thepeoplesvoice.tv/bill-gates-arrest-warrant-issued-in-philippines-for-premeditated-murder-linked-to-vaccine-roll-out>
- Doherty, K., & Doherty, G. (2018). Engagement in HCI: Conception, theory and measurement. *ACM Computing Surveys*, 51(5), Article 99.  
<https://doi.org/10.1145/3234149>
- Doll, W. J., Hendrickson, A., & Deng, X. (1998). Using Davis's perceived usefulness and ease-of-use instruments for decision making: A confirmatory and multigroup invariance analysis. *Decision Sciences*, 29(4), 839-869.

- Dunhill, J. (2023). *1900s "One Night Cough Syrup" Scares the Internet with Wild Ingredients*. Retrieved June 11, 2023 from <https://www.iflscience.com/1900s-one-night-cough-syrup-scaries-the-internet-with-wild-ingredients-68056>
- Eagleton-Patriot, F. (n.d.). *The View Settles with the Rittenhouses for \$22 Million and a Formal Apology*. USA Taters. Retrieved June 8 from <https://usataters.com/the-view-settles-with-the-rittenhouses-for-22-million-and-a-formal-apology/>
- Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: the presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 813.
- Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, 38(8), 1087-1100.
- Fan, J., Shao, M., Li, Y., & Wang, J. (2018). The impact of different design characteristics of warning message on users' perceived usefulness and perceived effectiveness of health information searching. *Proceedings of the 15th International Conference on Service Systems and Service Management (ICSSSM)*, 1-6.
- Gao, M., Xiao, Z., Karahalios, K., & Fu, W.-T. (2018). To label or not to label: The effect of stance and credibility labels on readers' selection and perception of news articles. *Proceedings of the ACM Human-Computer Interaction*, 2(CSCW), Article 55. <https://doi.org/10.1145/3274324>
- Geeng, C., Yee, S., & Roesner, F. (2020). Fake news on Facebook and Twitter: Investigating how people (don't) investigate. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-14.
- Ginsca, A. L., Popescu, A., & Lupu, M. (2015). Credibility in information retrieval. *Found. Trends Inf. Retr.*, 9(5), 355-475. <https://doi.org/10.1561/15000000046>
- Gu, R., & Hong, Y. K. (2019). Addressing health misinformation dissemination on mobile social media. *ICIS 2019 Proceedings*, 29.
- Hendrickson, A. R., Massey, P. D., & Cronan, T. P. (1993). On the test-retest reliability of perceived usefulness and perceived ease of use scales. *MIS Quarterly*, 227-230.
- Henney, K. (2020). *The Times They are a-Changin'*. Medium. Retrieved June 11, 2023 from <https://kevinhenney.medium.com/the-times-they-are-a-changin-37f9d7c43c8f>

- Hills, T. T. (2019). The dark side of information proliferation. *Perspectives on Psychological Science*, 14(3), 323-330.
- Holderith, P. (2023). *Ford Applies to Patent Self-Repossessing Cars That Can Drive Themselves Away*. The Drive. Retrieved June 12, 2023 from <https://www.thedrive.com/news/future-fords-could-repossess-themselves-and-drive-away-if-you-miss-payments>
- Jahanbakhsh, F., Zhang, A. X., & Karger, D. R. (2022). Leveraging structured trusted-peer assessments to combat misinformation. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-40.
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1420.
- Jun, Y., Meng, R., & Johar, G. V. (2017). Perceived social presence reduces fact-checking. *Proceedings of the National Academy of Sciences*, 114(23), 5976-5981. <https://doi.org/doi:10.1073/pnas.1700175114>
- Kim, C., & Yang, S.-U. (2017). Like, comment, and share on Facebook: How each behavior differs from the other. *Public Relations Review*, 43(2), 441-449.
- Kirchner, J., & Reuter, C. (2020). Countering fake news: A comparison of possible solutions regarding user acceptance and effectiveness. *Proceedings of ACM Human-Computer Interaction*, 4(CSCW2), Article 140. <https://doi.org/10.1145/3415211>
- Kluger, B. (2021). *Unbelievable News: Dr. Oz Wins Nobel Prize in Medicine*. Retrieved June 10, 2023 from <https://www.benzikluger.com/2021/04/01/unbelievable-news-dr-oz-wins-nobel-prize-in-medicine>
- Koch, T., Frischlich, L., & Lerner, E. (2021). *The effects of warning labels and social endorsement cues on credibility perceptions of and engagement intentions with fake news [Preprint]*. PsyArXiv.
- LaMagdeleine, I. S. (2023a). *Budweiser Brands Won't Be at Oktoberfest for the First Time in 75 Years?* Snopes.com. Retrieved June 15 from <https://www.snopes.com/fact-check/budweiser-oktoberfest/>
- LaMagdeleine, I. S. (2023b). *Is Disney World Moving to Georgia?* Snopes.com. Retrieved June 12 from <https://www.snopes.com/fact-check/disney-world-georgia-florida/>
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research Methods in Human-Computer Interaction* (2nd ed.). Elsevier Inc.

- Lewis, J. R. (2019). Comparison of Four TAM Item Formats: Effect of Response Option Labels and Order. *Journal of Usability Studies*, 14(4).
- Liles, J. (2023). *No, Kelly Ripa Isn't Leaving 'Live with Kelly and Ryan' to Sell Skincare Products*. Snopes.com. Retrieved June 15 from <https://www.snopes.com/fact-check/kelly-ripa-leaving-live-show-skincare-products/>
- Ma, X., Cheng, J., Iyer, S., & Naaman, M. (2019). When do people trust their social groups? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Paper 67. <https://doi.org/10.1145/3290605.3300297>
- MacDonald, C. M., & Atwood, M. E. (2014). What does it mean for a system to be useful? An exploratory study of usefulness. *Proceedings of the 2014 Conference on Designing Interactive Systems*, 885-894.
- The Mammoth*. (n.d.). Colossal Laboratories and Biosciences. Retrieved June 12, 2023 from <https://colossal.com/mammoth/>
- Mason, A. (2023). *10 Crazy Facts You Didn't Know About Jerry Springer*. TV Overmind. Retrieved June 11 from <https://tvovermind.com/10-crazy-facts-didnt-know-jerry-springer/>
- McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communications Monographs*, 66(1), 90-103.
- Meta. (n.d.). *Transparency Center policy details*. Facebook. Retrieved June 19, 2022 from <https://transparency.fb.com/policies/community-standards/false-news/>
- Moran, M. (2022). *Nostradamus predicted that Charles will abdicate and 'mystery king' could replace him*. Daily Star. Retrieved June 11, 2023 from <https://www.dailystar.co.uk/news/weird-news/nostradamus-predicted-charles-abdicate-mystery-27973773>
- Moravec, P. L., Kim, A., & Dennis, A. R. (2020). Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media. *Information Systems Research*, 31(3), 987-1006.
- Müller, H., Sedley, A., & Ferrall-Nunge, E. (2014). Survey research in HCI. In *Ways of Knowing in HCI* (pp. 229-266). [https://doi.org/10.1007/978-1-4939-0378-8\\_10](https://doi.org/10.1007/978-1-4939-0378-8_10)
- Nianias, H. (2015). *Stephen Hawking Pictured on Jeffrey Epstein's 'Sex Slave' Caribbean island*. Independent. Retrieved June 12, 2023 from <https://www.independent.co.uk/news/people/stephen-hawking-pictured-on-jeffrey-epstein-s-sex-slave-caribbean-island-9974955.html>

- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Nielsen, J. (2012). *Thinking Aloud: The #1 Usability Tool*. Nielson Norman Group. Retrieved June 16 from <https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>
- Number of monthly active Facebook users worldwide as of 3rd quarter 2022*. (2022). Statista. Retrieved November 16, 2022 from <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303-330.
- O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938-955.
- O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112, 28-39. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2018.01.004>
- Papakyriakopoulos, O., & Goodman, E. (2022). The impact of Twitter labels on misinformation spread and user engagement: Lessons from Trump's election Tweets. *Proceedings of the ACM Web Conference 2022*, 2541–2551. <https://doi.org/10.1145/3485447.3512126>
- Parker, L. (1988). *Roof Ripped Off Jet in Apparent Structural Failure*. The Washington Post. Retrieved June 8, 2023 from <https://www.washingtonpost.com/archive/politics/1988/04/30/roof-ripped-off-jet-in-apparent-structural-failure/7c101220-e10d-4fcb-9176-3e0ae4762336/>
- Pasquetto, I. V., Jahani, E., Atreya, S., & Baum, M. (2022). Social debunking of misinformation on WhatsApp: The case for strong and in-group ties. *Proceedings of ACM Human-Computer Interaction*, 6(CSCW1), Article 117. <https://doi.org/10.1145/3512964>
- Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, 66(11), 4944-4957.
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590-595. <https://doi.org/10.1038/s41586-021-03344-2>

- Quesenbery, W. (2003). Dimensions of usability: Defining the conversation, driving the process. *Proceedings of the Usability Professionals' Association 2003 Conference*, 23-27.
- Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide*. John Wiley & Sons.
- Rogers, W. A., Lamson, N., & Rousseau, G. K. (2000). Warning research: An integrative perspective. *Human Factors*, 42(1), 102-139.
- Salkind, N. (2017). *Exploring Research* (9th ed.). Pearson Education, Inc.
- Samaiya, R. (2017). *Obvious Mistakes In Movies Which You Might Have Missed*. CineTales. Retrieved June 8, 2023 from <https://www.cine-tales.com/obvious-mistakes-movies-might-missed/>
- Sauro, J. (2011). *Measuring usability with the System Usability Scale (SUS)*. Retrieved December 23, 2023 from <https://measuringu.com/sus/>
- Sekaran, U., & Bougie, R. (2016). *Research Methods for Business* (7th ed.). John Wiley and Sons, Ltd.
- Shneiderman, B. (1997). Direct manipulation for comprehensible, predictable and controllable user interfaces. *Proceedings of the 2nd International Conference on Intelligent User Interfaces*, 33-39.
- Sittig, A., & Zuckerberg, M. (2013). *Managing information about relationships in a social network via a social timeline* (USA Patent No. US-8583690-B2). <https://portal.unifiedpatents.com/patents/patent/US-8583690-B2>
- Snopes is the internet's definitive fact-checking resources*. (n.d.). Snopes Media Group. Retrieved March 10, 2023 from <https://www.snopes.com/about/>
- Son, J. (2023). Quick-and-Wide Propagation of Disaster Tweets: Why It Matters and How to Measure It. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, 54(4), 11-30.
- Spezzano, F., Shrestha, A., Fails, J. A., & Stone, B. W. (2021). That's fake news! Reliability of news when provided title, image, source bias. *Proceedings of ACM Human-Computer Interaction*, 5(CSCW1), Article 109. <https://doi.org/10.1145/3449183>
- Spiro, E., Irvine, C., DuBois, C., & Butts, C. (2012). Waiting for a retweet: Modeling waiting times in information propagation. *Proceedings of 2012 Neural Information Processing Systems Conference (NIPS)*, 12.



- SPSS Tutorials: Chi-Square Test of Independence. (2023). Retrieved October 22, 2023, from <https://libguides.library.kent.edu/spss/chisquare>
- Stephenson, C. (2023). *Pokemon Card Collection Worth \$10,000 Appears on Antiques Roadshow*. Dexerto. <https://www.dexerto.com/pokemon/pokemon-north-america-international-championships-2023-how-to-watch-dates-games-2194558/>
- Stetson, J. N., & Tullis, T. S. (2004). A comparison of questionnaires for assessing website usability. *Usability Professionals' Association Presentation*.
- Tahir, R., Batool, B., Jamshed, H., Jameel, M., Anwar, M., Ahmed, F., Zaffar, M. A., & Zaffar, M. F. (2021). Seeing is believing: Exploring perceptual differences in DeepFake videos. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Article 174. <https://doi.org/10.1145/3411764.3445699>
- TikTok CEO Was an Intern for Rival Facebook: 'Working for a Startup...'*. (2023). Money Control. Retrieved June 10, 2023 from <https://www.moneycontrol.com/news/trends/tiktok-ceo-shou-zi-chew-facebook-internship-10312121.html>
- Treisman, R. (2022). *Martin Luther King Jr. Paid the Bill for Julia Roberts' Birth. Here's the Backstory*. NPR. Retrieved June 10, 2023 from <https://www.npr.org/2022/11/01/1133121228/julia-roberts-mlk-birth>
- Twitter. (2021, October). *Civic integrity policy*. Retrieved June 19, 2022 from <https://help.twitter.com/en/rules-and-policies/election-integrity-policy>
- van der Nat, D. J., Huiskes, V. J., Taks, M., Pouls, B. P., van den Bemt, B. J., & van Onzenoort, H. A. (2022). Usability and perceived usefulness of patient-centered medication reconciliation using a personalized health record: a multicenter cross-sectional study. *BMC Health Services Research*, 22(1), 1-17.
- Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, 27(3), 451-481.
- Vraga, E. K., & Bode, L. (2018). I do not believe you: How providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, 21(10), 1337-1353.
- Wash, R., & Lampe, C. (2012). The power of the ask in social media. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 1187-1190. <https://doi.org/10.1145/2145204.2145381>

- Waxman, O. B. (2017). *The Bear Who Became a Cigarette-Smoking, Beer-Drinking World War II Hero*. Time USA. Retrieved June 12, 2023 from <https://time.com/4731787/wojtek-the-bear-history/>
- Wix. (n.d.). *Top Reasons for Choosing Wix*. Retrieved April 22, 2023 from <https://support.wix.com/en/article/top-reasons-for-choosing-wix>
- Wogalter, M. S., & Mayhorn, C. B. (2005). Providing cognitive support with technology-based warning systems. *Ergonomics*, 48(5), 522-533.
- Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, 41, 135-163.
- Zeng, L., Starbird, K., & Spiro, E. S. (2016). Rumors at the speed of light? Modeling the rate of rumor transmission during crisis. *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)*, 1969-1978.