Article

new media & society
1–32

DOI: 10.1177/14614448241227841
journals.sagepub.com/home/nms
Sage

Building resilience to misinformation in communities of color: Results from two studies of tailored digital media literacy interventions

Angela Y Lee*
Stanford University, USA

Ryan C Moore*
Stanford University, USA

Jeffrey T Hancock
Stanford University, USA

Abstract

Interventions to build resilience to misinformation should consider the needs of communities of color, who experience (mis)information in unique ways. We evaluated digital media literacy interventions to improve misinformation resilience among four communities of color in the United States (Black, Latino, Asian American/Pacific Islander, Native American), which were designed by the nonprofit PEN America and community partner organizations. We assessed intervention efficacy in two studies: (1) a quasi-experimental field study among diverse participants recruited via community outreach and (2) a randomized controlled trial among Latinos recruited via a survey company (total $N = 370$). Results indicated that participants in both studies improved their comprehension of digital media literacy skills after taking the intervention. However, only those recruited via community outreach improved their ability to accurately identify true and false online news in a behavioral detection task. Our findings highlight the need to consider heterogeneous treatment effects in misinformation interventions, particularly across different communities and intervention contexts.

*shared first-authorship.

Corresponding author:
Angela Y Lee, Department of Communication, Stanford University, Stanford, CA, 94305, USA.
Email: angela8@stanford.edu

Keywords
AAPI, African American, digital literacy, diversity, intervention, Latino, media literacy, misinformation, motivation, Native American

Introduction

The consequences of the misinformation crisis are clear and concerning. False claims on the Internet can undermine personal and public health by promoting pseudoscientific remedies, undermining clinical guidance, and reducing trust in treatments, screenings, and vaccines (Loomba et al., 2021; Peng et al., 2023). Misinformation can also threaten core civic processes. Exposure to false claims about politics discourages voter turnout (Green et al., 2022), produces a less-informed citizenry (Morgan et al., 2021), and erodes confidence in democracy itself (Ognyanova et al., 2020; Dahlke & Hancock, 2023).

Although misinformation affects people of all backgrounds, little research has focused on the experiences of communities of color (Camargo and Simon, 2022). This is concerning as recent investigations reveal that Black (Dodson et al., 2021), Latino (Cortina and Rottinghouse, 2022), Asian American/Pacific Islander (AAPI) (Nguyễn et al., 2022), and Native American communities (Getahun, 2021) are uniquely targeted by misinformation. In addition to translating falsehoods into different languages to accelerate their spread and avoid fact-checking (Uribe, 2022), purveyors of misinformation disseminate false claims across social media networks used frequently by communities of color (e.g. WeChat, Pang, 2022; WhatsApp, Garimella and Eckles, 2020). These focused attacks can compound ongoing disparities in Black, Indigenous, and people of color (BIPOC)'s access to health resources (Hill and Artiga, 2022) and opportunities for civic participation (Igielnik and Budiman, 2020).

Therefore, interventions are needed to support communities of color in building resilience against misinformation. While interventions have successfully improved people's ability to detect false claims by teaching digital media literacy skills to interrogate information encountered online (e.g., McGrew, 2020), few have focused on serving individuals from ethnically diverse communities (Camargo and Simon, 2022). As such, it is important to test whether such approaches are effective in improving misinformation detection and response among participants from diverse backgrounds—particularly given that BIPOC communities experience misinformation in unique ways (Lee et al., 2023). Indeed, efforts to teach digital media literacy skills should be responsive to diverse digital media ecologies, including multilingual news sources (e.g. Univision), ethnic news outlets (e.g. NextShark, Indian Country Today), and peer-to-peer messaging platforms (e.g. WeChat) that are essential sources of information for many.

In this article, we report on the efficacy of a novel digital media literacy intervention developed with and for BIPOC communities by the nonprofit organization PEN America. The educational intervention taught individuals a set of digital literacy skills to help sort fact from fiction online across diverse media ecologies. Following best practices for developing interventions for diverse communities (Griffin et al., 2013), the content of the intervention was informed by the perspectives of Black, Latino, AAPI, and Native American community leaders (see Lee et al., 2023).

To assess the impact of the tailored intervention on misinformation resilience, we conducted two studies: a community-based study (Study 1) and a survey panel-based

study (Study 2). First, we launched a field study where the nonprofit organizations recruited participants from the communities they served (Study 1). Next, we conducted a randomized control trial of the intervention with a nationally representative sample of Latino adult participants recruited via the survey company YouGov (Study 2). While the intervention improved people's digital media literacy skills in both studies, it only improved the ability to identify true and false online information among participants recruited via community outreach. Specifically, those recruited through community outreach were more likely to apply the digital media literacy skills they learned from the intervention when evaluating online news, and became more likely to accurately identify true and false news on a behavioral task.

We interpret these findings as suggesting that why and how people participate in media literacy interventions can have important impacts on intervention efficacy. Specifically, our results indicate it may be valuable to consider participants' *motivations* for completing the intervention and the *social affordances* of the intervention context, in addition to centering the experiences of *diverse communities*. We discuss the implications of these three considerations for future misinformation research and real-world interventions. Our goal in this article was to examine the effects of an intervention tailored for diverse communities. In doing so across multiple studies, our findings reveal that tailored interventions can be effective, but also that it may be valuable to consider the potential roles of motivation and social affordances in future research.

Literature review

Combating misinformation with digital media literacy

Although digital platforms make it easier for people to share information, they also enable bad actors to publish and propagate false claims at scale. To do so, originators of misinformation often attempt to establish a "patina of credibility" around their claims (Cooke, 2017). In addition to generating factitious headlines, they also fabricate information sources, manipulate photos and videos, and take statistics and quotes out of context (Bastos and Farkas, 2019). Furthermore, purveyors try to increase the uptake of their false claims by using language capitalizing on people's tendency to engage with highly emotional content (e.g. fear-inducing, morally outrageous; Rathje et al., 2021), and by infiltrating online communities with fake accounts (Wilson and Starbird, 2020).

Fortunately, digital media literacy interventions can combat misinformation by equipping people with skills to see through these deceptive strategies. Lateral reading, for example, teaches individuals to investigate the veracity of claims by comparing what multiple sources have to say about the same story, reducing the chances that they will be duped by disingenuous sources (Wineburg and McGrew, 2019). Similarly, teaching people to use reverse image searches can improve their ability to evaluate the credibility of visual content, such as whether a photo was altered or taken out of context (Cao et al., 2020). Across multiple studies, interventions to improve individuals' digital media literacy skills have effectively improved their ability to discern true from false information online (Axelsson et al., 2021; Guess et al., 2020; McGrew, 2020; Nygren et al., 2021; Wineburg et al., 2022; Moore & Hancock, 2022).

These approaches can be understood through the framework of digital resilience, which aims to provide people with the skills, self-efficacy, and support systems they need to respond adaptively to challenges they may encounter online (Lee & Hancock, 2023). Building on psychological research that defines resilience as when individuals and communities "fare better than expected given adversity" (Troy et al., 2023), efforts to promote digital resilience work to equip people with the knowledge they need to recognize and respond to online threats, such as misinformation, should they encounter them. Indeed, prior studies summarized by Roozenbeek et al. (2021) have framed digital media literacy interventions as "a means to preemptively build resilience against anticipated exposure to misinformation," instead of responding to false claims after they have already taken hold. In line with this prior work, we argue that being able to accurately identify true from false news in an online headline classification task is a measure of resilience to misinformation and can be defined as the ability to respond adaptively, given exposure to the threat of false claims encountered online.

Studies of resilience emphasize the importance of considering how individual attributes, such as skill acquisition, interact with structural access to resources (Masten & Barnes, 2018). In the context of interventions to build resilience to misinformation, such as digital media literacy interventions, scholars have noted that the majority have been developed and tested among predominantly English-speaking communities in the United States (Blair et al., 2023), which may limit their ability to support individuals from minoritized backgrounds, particularly if they are already misinformation-susceptible (Amazeen et al., 2024). Therefore, though these existing results are encouraging, more work is needed to make interventions responsive to the needs and strengths of diverse communities. Ideally, digital media literacy interventions should help individuals identify misinformation by giving them the tools to investigate claims wherever on the Internet they may encounter them. In existing interventions, however, skills tend to be taught in the context of news sources that may not be representative of the diverse ways people encounter news (Camargo and Simon, 2022). For example, people are often taught to apply lateral reading to claims encountered on mainstream digital news outlets (i.e. LA Times) and widely used social media platforms (i.e. Facebook, Twitter). However, we know that people can obtain news from a variety of sources, including ethnic or international news outlets (e.g. Indian Country Today, China Global Television Network), multilingual websites (e.g. Univision), identity-based accounts on social media (e.g. NextShark), and peer-to-peer messenger apps (e.g. WhatsApp), each of which has unique affordances to consider when examining information veracity. Integrating these diverse media ecologies into digital media literacy interventions is critical to protecting all individuals from misinformation and its harms.

Accounting for diverse media ecologies in misinformation interventions

In making media literacy initiatives more inclusive and responsive to diverse communities, it is essential to integrate the perspectives of individuals from those communities (Bogart and Uyeda, 2009; Griffin et al., 2013). Interviews and focus groups can shed light on how BIPOC experiences and respond to (mis)information, surfacing strategies for tailoring digital literacy education efforts.

Results from a series of listening sessions with members of nonprofit organizations serving Black, Latino, AAPI, and Native American communities identified several ways that media literacy interventions can better account for diverse media ecologies (Lee et al., 2023). First, efforts to build misinformation resilience should be made accessible in languages other than English. In addition to reaching underserved communities for whom language is a barrier to participation, providing multilingual options for interventions is essential to combating multilingual misinformation. Already, purveyors have posted Spanish-language misinformation to social media platforms to sway the vote of the increasingly influential Latino voting bloc (Mochkofsky, 2022; Navia, 2021). Similarly, false claims in Vietnamese have exploited the concerns of Vietnamese American immigrants and refugees to influence support for certain politicians and policies (Nguyễn et al., 2022; Tran, 2020).

In addition to addressing language, interventions need to be responsive to misinformation purveyors' reference to salient concerns of historically marginalized groups. For example, Spanish-language misinformation targeted toward underserved Latino communities (e.g. low-income individuals, first-generation immigrants) falsely claimed that proof of residence, insurance, or government ID was required to get COVID-19 treatments when they were often free to obtain (Longoria et al., 2021; Nguyen and Catalan-Matamoros, 2020). Similar reference to historical injustices can be seen in misinformation targeted toward African American and Native American tribal communities. Indeed, misinformation has compared contemporary health measures (e.g. vaccinations, preventative screenings) to historical wrongs like the Tuskegee Syphilis Crisis and disease epidemics tied to colonialism (Ostler, 2020) to discourage individuals from obtaining care and to sow doubt in healthcare institutions (Dodson et al., 2021). In addition to undermining public health responses to crises like COVID-19, these misinformation narratives can exacerbate existing health inequities. Therefore, Lee et al., 2023 argue that media literacy interventions need to understand how misinformation interacts with communities' specific political, cultural, and historical contexts.

It is also important to consider the characteristics of the diverse communication channels through which people are exposed to (mis)information. For example, relative to white Americans, a greater proportion of Black, Latino, and Asian Americans report regularly obtaining news from social media platforms (Abrajano et al., 2022; Forman-Katz and Matsa, 2022). While Facebook is a popular news source for all, BIPOC are more likely to report using platforms like Instagram and WhatsApp to find news than white individuals. Indeed, group-messaging apps like WhatsApp and WeChat can be important sources of information for many communities of color. Examinations of the flows of (mis)information across these platforms indicate that their unique affordances influence how people encounter, detect, and respond to claims (De Freitas Melo et al., 2019; Garimella and Eckles, 2020). For example, the use of "forwarding" or "sharing" news stories to large group chats is central to WhatsApp, whereas platforms more commonly covered in misinformation interventions (e.g. Twitter) deliver information via personalized streams of content from followed or recommended sources. Therefore, interventions may be more effective for diverse communities if they consider these unique platforms and their affordances.

In discussing interventions designed to improve misinformation resilience in communities of color, we note that we are not claiming that individuals from these

communities are more vulnerable, or susceptible to believing misinformation. Rather, we emphasize that misinformation purveyors may differentially target people from different backgrounds. Although some evidence suggests that individuals are worse at detecting misinformation in their nonnative languages (Muda et al., 2023), nationally representative studies of misinformation susceptibility indicate that participants' ethnicity is not significantly associated with misinformation sharing (Guess et al., 2020). However, there is an important discrepancy in the resources that are provided to communities of color in recognizing and responding to misinformation (Amazeen et al., 2024; Nguyễn et al., 2022). For instance, many digital media literacy interventions have been developed with, tested on, and shared with predominantly white, English-speaking communities in the United States (see Blair et al., 2023 for a review).

The present study evaluates the efficacy of digital media literacy interventions designed with the needs of diverse communities and the nuances of their media ecologies in mind. To be clear, some studies have tested the effects of media literacy interventions among diverse participants (e.g. a sizable number of participants were non-white; McGrew, 2020; Wineburg et al., 2022). However, these studies did not examine interventions tailored to specific communities. While some studies have evaluated interventions for non-English speakers (e.g. Ali and Qazi, 2023; Badrinathan, 2021; Basol et al., 2021; Nygren et al., 2021), this work has been predominantly conducted outside the United States. To our knowledge, interventions tailored for diverse communities within the United States have not been developed and evaluated, which is essential to better capture the unique and diverse experiences many Americans have with media (Ghai, 2021). Therefore, we pose the following research question:

> Are digital media literacy interventions effective when tailored for diverse American communities?

When we evaluate the impact of misinformation interventions, it is essential to identify interventions' effects on people's detection of both false and true news *separately* (Moore & Hancock, 2022). Interventions can improve overall detection abilities by increasing people's ability to detect false news but decreasing their ability to detect true news, if the intervention simply makes people more suspicious and thus judges all news as false more often (see Guess et al., 2020). This outcome is concerning given that people encounter more true news than false news in their everyday media consumption (Allen et al., 2020). Thus, interventions should ideally improve people's ability to accurately identify both true *and* false news, an ability referred to as *discriminant trust* (Moore & Hancock, 2022).

Study 1: Community-based intervention evaluation

Study 1 examined the effects of a tailored intervention through a field study. The design of the intervention was informed by the results of a series of in-depth focus groups conducted with representatives from non-profit organizations serving the Black, Latino, AAPI, and Native American communities (see Lee et al., 2023 for more information about the focus group findings). In this article, we do not report any qualitative data on

intervention development, but instead the results of a quantitative field study examining the effects of the developed intervention on participants' misinformation resilience.

We evaluated the interventions' effects on digital media literacy skills and misinformation detection among BIPOC using a one-shot, pre-post design that was optimized for ecological validity (Murphy, 2003; Orne and Holland, 1968). In an effort to most closely mirror how people from diverse communities would participate in a digital media literacy workshop in the real world, we implemented the intervention in the form of public-facing workshops organized by members of the four nonprofit organizations, who recruited participants through community-based means, including social media posts, emails, and peer conversations.

Methods

Intervention development and materials. Intervention content was developed by PEN America based on listening sessions conducted with community members and leaders from (1) *Mi Familia Vota*, a civic engagement organization dedicated to engaging Latino communities; (2) the *National Congress of American Indians*, the oldest and largest American Indian and Alaska Native organization serving the interests of tribal governments and communities; (3) *Asian Americans Advancing Justice*, an organization advocating for Asian Americans' civil and human rights; and (4) the *National Action Network*, a leading civil rights organization promoting a modern civil rights agenda in the spirit of Dr. Martin Luther King, Jr (see the Supplemental Materials for more information about these four partner organizations). The listening sessions shed light on the dynamics of online (mis)information within the different communities and allowed community members to explain community needs, strengths, and challenges (see Lee et al., 2023 for more information about the listening sessions). Interventions ensured language access for participants in several ways. Native speakers translated all intervention materials, as well as surveys, into Spanish. Live translations were provided for the synchronous sessions with Mi Familia Vota, by using the Zoom translation feature, where participants could opt to listen in English or Spanish. The intervention also acknowledged the need for multilingual fact-checking resources and shared links to specific organizations (e.g. Univision's El Detector). The interventions considered diverse media ecologies by highlighting a variety of sources of trusted information, including ethnic media organizations, and discussed how to address misinformation across a variety of platforms. For example, our resource packet for Native American communities included links to Indian Country Today's coverage of the COVID-19 crisis, as well as the NCAI's list of COVID-19 resources.

The final interventions were 30-minute online webinars,[1] which taught five primary digital media literacy skills: lateral reading, click restraint, reverse image search, monitoring emotional reactions to headlines, and using fact-checking resources. Skills were chosen based on prior research indicating their utility in helping people discern between true and false online information (see Table S1), and were presented alongside application examples (e.g. demonstrating how to use Google to conduct a reverse image search).

Study design. Our choice of study design for Study 1 was informed by a request from our community partners to make the intervention publicly available to all individuals who were interested in participating, in line with their values of accessibility and inclusion. Following this request precluded us from using study designs that involved random assignment to treatment and control groups, which would have provided greater internal validity to our study. Therefore, the next-strongest design given this limitation would be to conduct a one-group pretest posttest design using a nonequivalent dependent variable, a quasi-experimental study design that is frequently used in contexts when random assignment to a treatment or control groups is not ethically or pragmatically feasible (see Harris et al., 2006). While all quasi-experimental studies face internal validity concerns, we included a dependent variable that we did not expect to change as a result of the intervention, as recommended by Harris et al. (2006). In addition to completing a measure of five digital media literacy skills that were taught in the intervention (e.g. lateral reading, reverse image search), we also included another measure about their "ability to use search engine queries"—a comparable digital media literacy skill that was not taught in the intervention and should not be expected to change.

While Study 1's quasi-experimental approach is limited in internal validity, other scholars have noted (e.g. Schmuckler, 2001) that it is important to complement the results of laboratory studies that are optimized for internal rather than external validity with more ecologically valid designs, as participants often respond to stimuli in rigorously-controlled environments in different ways than how they would engage with that stimuli in a more naturalistic context (i.e. in a workshop in the real world). We therefore conceptualized our Study 1 as a "proof of concept" that would allow us to examine the potential efficacy of the tailored interventions, as informed by the design suggestions learned from focus groups with BIPOC community leaders (see Lee et al., 2023)."

Participants and procedure. From October to December 2021, 70 participants were recruited to complete the intervention, which was embedded within four free, publicly available workshops organized by the nonprofit organization PEN America. To promote the study, PEN America and their partner nonprofits distributed informational materials through email lists, social media posts, and word of mouth. Interested individuals could register to join synchronously on Zoom, or participate through live-streams on YouTube and Facebook Live. Information distributed before and during the workshop emphasized that people did not have to complete the research study to attend the workshop. People who consented to participate in the study completed a brief online survey before and after the intervention. Those who completed both the pre- and post-intervention surveys were entered into a raffle to win one of five $100 gift cards. The pre-intervention survey, which participants took immediately prior to the intervention, took approximately 10 minutes to complete, the intervention lasted 40 minutes, and the post-intervention survey, which participants took immediately after the intervention concluded, took approximately 10 minutes to complete. All study procedures were approved by the Stanford University IRB.

We excluded 25 participants who did not complete both the pre- and postintervention surveys. Our final sample of 45 participants was ethnically diverse, with 18.4% identifying as American Indians/Alaskan Natives, 12.2% as Hispanic/Latino, 20.4% as Asian/

Asian American, 6.1% as Black/African/African American, and 38.8% as White. Data were analyzed in line with a quantitative analytic framework used to evaluate field studies.

Positionality statement. Identities represented in our research team included woman of color, immigrant, and first-generation college student. Our team included experts in media psychology, misinformation, and educational interventions.

Measures

Digital media literacy skills. We assessed participants' digital literacy skills by asking participants to complete a 6-item measure rating their understanding of lateral reading, click restraint, monitoring emotional reactions to headlines, reverse image search, and using fact-checking resources. Following prior work, participants rated their comprehension of each skill on a 5-point scale, ranging from 1: no understanding to 5: full understanding (Moore & Hancock, 2022). This survey-based measure is modeled after Hargittai and Hsieh's (2011) measure of Internet skills, which was validated against people's actual skills during scale development (for details, see Hargittai, 2002, 2005) and more recently was shown to be highly predictive of people's performance on online information retrieval tasks (e.g. answering the question "What is the only US National Park that begins with the letter T?") by Guess and Munger (2023). Following Guess and Munger (2023), we also included a measure of a digital skill that was not taught in the intervention (using search engines) to assess the specificity of treatment effects, as our nonequivalent dependent variable. Internal reliability was assessed with Cronbach's alpha before and after the intervention, indicating good internal consistency ($a_{t1} = .85$, $a_{t2} = .93$).

News headline classification task. As in most studies of misinformation interventions, we evaluated participants' ability to identify true and false claims online by asking participants to judge the veracity of a series of headlines (Pennycook et al., 2021). Participants in Study 1 saw different series of four headlines in the pre- and postsurveys, viewing a total of eight headlines over the course of the study (see Table S2). All headlines were determined to be either true or false by reputable fact-checking organizations such as AP News, Reuters, or Snopes. Headlines presented in each survey consisted of two true and two false headlines. In each survey, two of the four headlines (one true and one false) were health-related and the other two were about political events. The headlines were also balanced with respect to the extent to which they aligned with conventionally Republican or Democrat beliefs. Participants were asked about different headlines in the pre- and post-surveys to eliminate the possibility that they "learned" the headlines between the survey waves (Roozenbeek et al., 2021).

For each headline, participants reported how true or false they thought each headline was (scale from 1: definitely false to 7: definitely true). The accuracy of each judgment was calculated by coding headline evaluations as correct if individuals rated false news as definitely to somewhat false (1–3 on a 7-point scale), and true news as definitely to somewhat true (5–7 on a 7-point scale); responses of 4 were coded as incorrect. After completing veracity judgments for each set of four headlines, we asked participants

whether they had informed their veracity judgment by doing research online to investigate the headline ("Did you do any research on this headline before you provided a judgment as to how likely you believed it to be true?"; yes, no) (Nygren et al., 2021; Moore & Hancock, 2022). We interpreted this as a behavioral measure of the use of digital media literacy skills. Afterwards, participants were provided with an English or Spanish fact-sheet informing them which headlines they viewed were true and false, with links to fact-checking articles.

Results

Digital media literacy skills. We first examined whether the intervention was effective in improving community-recruited participants' digital media literacy skills. To test this, we estimated a series of separate mixed-effects linear regression models where Time was the independent variable (IV; 1 = post-intervention, 0 = pre-intervention) and people's ratings of their digital media literacy skills were the dependent variables (DVs). Participants were modeled as random effects to account for repeated observations per participant.

Overall, participants' digital media literacy skills increased after the intervention (see Figure 1). Specifically, we observed significant improvements in participants' comprehension of lateral reading ($\beta = .94$, $SE = .17$, $p < .001$), reverse image search ($\beta = .57$, $SE = .16$, $p < .001$), and click restraint ($\beta = .49$, $SE = .14$, $p < .001$). We observed a marginally significant improvement in monitoring emotional reactions to headlines ($\beta = .20$, $SE = .12$, $p < .1$).

Ability to identify true and false news. Next, we examined whether the intervention improved participants' ability to accurately identify true and false news. Mixed-effects logistic regression models were estimated to evaluate whether participation in the intervention (1 = post-intervention, 0 = pre-intervention) changed participants' ability to accurately identify headline veracity (1 = judgment of a headline was accurate, 0 = judgment of a headline was inaccurate). Participants were modeled as random effects.

Participants' overall ability to correctly identify true and false news increased significantly after the intervention ($\beta = .56$, $SE = .21$, $p < .01$). As shown in Figure 2, participants' likelihood of correctly judging the veracity of headlines rose from 47% (95% CI = 39–56%) before the intervention to 61% (95% CI = 52–69%) after the intervention. Closer examination revealed that participants improved in their ability to discern between both false ($\beta = .54$, $SE = .30$, $p < .1$) and true ($\beta = .68$, $SE = .33$, $p < .05$) headlines, indicating that the intervention improved participants' ability to engage in *discriminant trust*.

Application of digital media literacy skills. Why might the intervention improve people's ability to identify true and false news? One reason may be that people were more likely to investigate the headlines after the intervention by applying their digital media literacy skills. After evaluating the headlines, participants were asked to indicate (at the headline

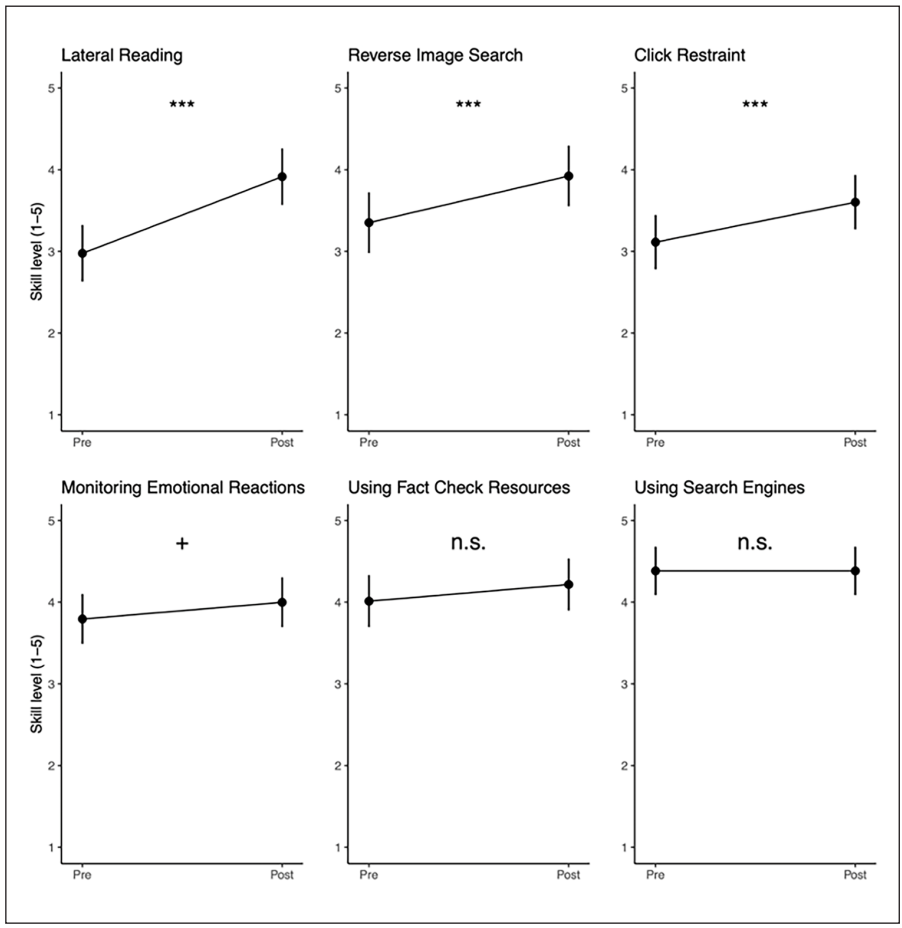

Figure 1. Change in digital media literacy skills (Study 1).
The y-axis represents ratings of individuals' comprehension of digital literacy skills taught in the intervention (1 = no understanding, 5 = full understanding). The x-axis indicates whether comprehension ratings came from pre-intervention or post-intervention. These predicted value estimates come from regression models described in the main text. Using search engines was not taught in the intervention and was included as a measure to assess the specificity of the intervention's effect on skills. Bars are 95% confidence intervals. ***$p < .001$; +$p < .1$; n.s. $p > .1$.

level) whether they conducted additional research online to inform their judgment. To evaluate the impact of the intervention on this research behavior, we estimated another logistic regression model where the DV was participants' likelihood of conducting research on a given headline and the IV was time. The result was significant ($\beta = 2.05$, $SE = .35$, $p < .001$), with the likelihood of researching a headline increasing from 6% (95% CI = 3–10%) before the intervention to 32% (95% CI = 26–38%) afterward (see Figure 3).

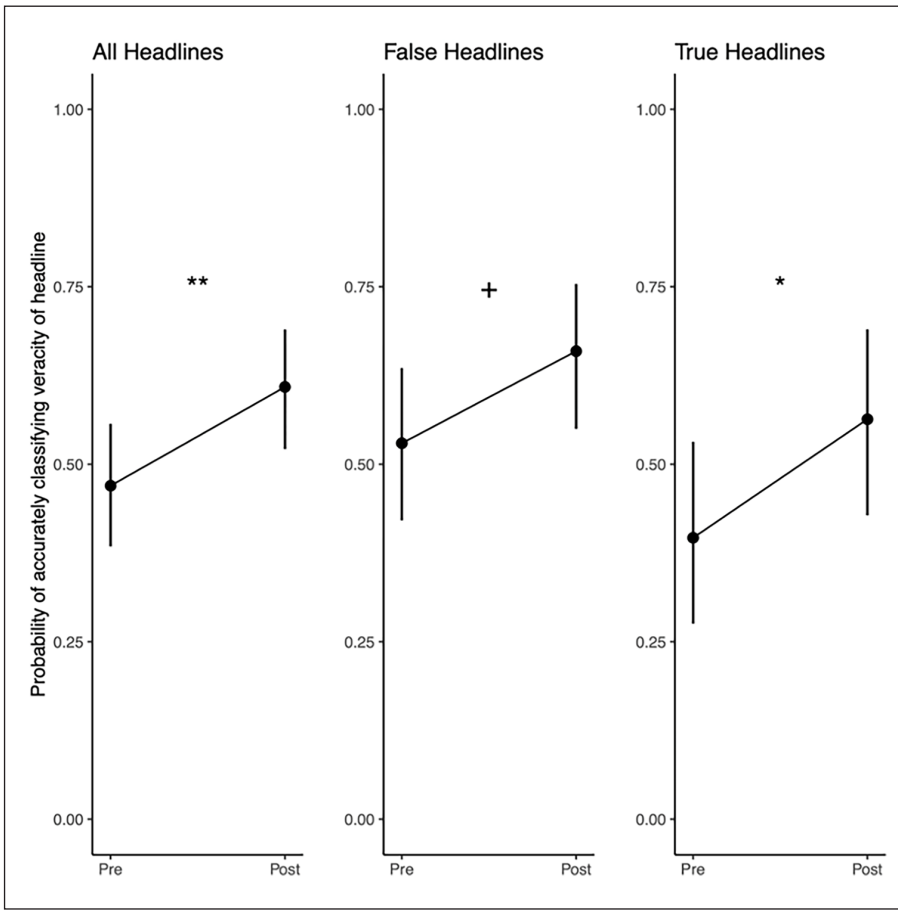

Figure 2. Change in news veracity detection abilities (Study 1).
The y-axis indicates predicted probabilities of accurately identifying the veracity of a given headline. Bars indicate 95% confidence intervals.
*$p < .05$; **$p. < .01$; + $p < .1$.

Discussion

Overall, the results of Study 1 indicate that the tailored intervention was effective at improving digital media literacy skills and misinformation resilience within a diverse sample recruited through community outreach. In addition to improving participants' digital literacy skills, they were also more likely to put these skills into practice by doing research on news stories to inform their judgments of story veracity. Crucially, they demonstrated improvements in *discriminant trust*, improving their ability to identify both false news as false and true news as true.

To our knowledge, our study is one of the first to study digital media literacy interventions tailored for ethnically diverse communities in the United States. In Study 1, we optimized for ecological validity by collaborating with community partners from nonprofit

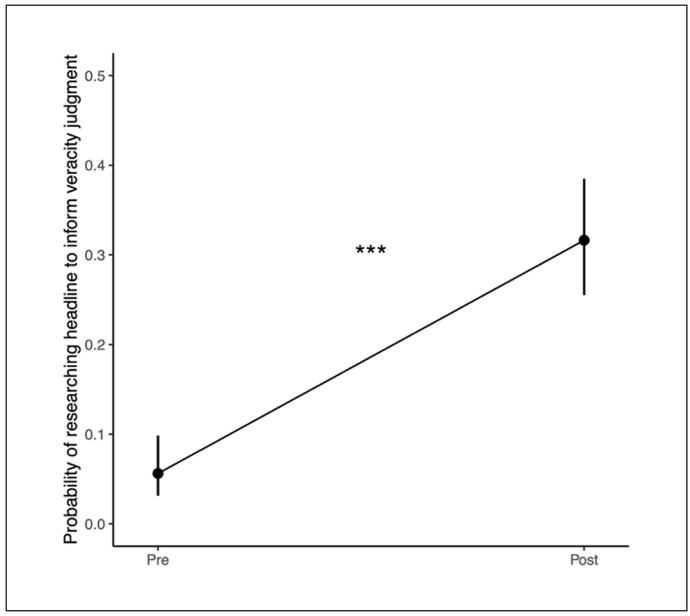

Figure 3. Change in likelihood of researching headlines to inform veracity judgments (Study 1). The y-axis indicates probabilities of individuals reporting they researched a headline online to inform their judgment of its veracity. Bars indicate 95% confidence intervals.
***$p < .001$.

organizations that represented and served the interests of local communities of color. Specifically, we followed procedures for participant recruitment that real-world organizations employ when disseminating misinformation prevention interventions. This allowed us to examine how people responded to the intervention as it would be taken in the real world, unlike the settings more common in academic research, such as in a laboratory.

However, these choices ultimately limited the internal validity of our study. Because participants volunteered to take part in the workshop, we were not able to ensure that all participants completed both the pre- and post-intervention surveys. In addition, as with other field studies, we also needed to consider the role of self-selection on the part of participants who chose to do the intervention. Because we studied the participants in a community organization's event, we did not have a control group, which lessened our ability to draw conclusions about the causal effects of the intervention and limited internal validity. We were also limited in the number of headlines utilized as part of the stimuli for the behavioral deception detection task due to the time constraints of the workshop. To address these limitations, we conducted a second study using a research design that afforded greater internal validity.

Study 2: Panel-based randomized control trial

We specifically designed the next study to address the limitations of Study 1 and strengthen the internal validity of our research. Study 2 assessed the impact of the

intervention on digital media literacy skills and misinformation resilience in a randomized controlled trial, revealing the causal effects of the intervention.

To improve our ability to generalize our inferences, we focused on the effect of one particular tailored intervention by studying the effect of the intervention designed for the Latino community. We worked with a survey company to recruit a sample of Latino participants that was representative of sociodemographic characteristics of the Latino adult population in the United States.

Methods

Participants and procedure. Between March and April 2022, 300 participants who reported identifying as Hispanic or Latino were recruited by YouGov. These 300 participants were diverse in gender (173 women and 127 men), age (M=45.5 years old, range: 19–81 years), and political affiliation (126 Democrats, 60 Republicans, 73 Independents, 41 Other/Not Sure). YouGov weighted participants to approximate the demographic attributes of the Latino adult population in the United States, and we use these weights in our analyses when applicable.

Half of the participants were randomly assigned to the treatment condition, completing the digital media literacy intervention. The other half were randomly assigned to a control condition where they viewed a video about Zoom fatigue (Fauville et al., 2023) hosted by the same facilitator as the digital literacy intervention. The Zoom fatigue control video and the digital media literacy intervention video were similarly structured such that they followed the same order of the facilitator (1) introducing himself and the subject he was speaking on (misinformation or Zoom fatigue), (2) discussing the reason these topics were important to understand, (3) presenting empirically based findings from recent research on the subject, and (4) presenting recommendations on how participants could apply the skills they had learned (i.e. how to recognize misinformation, how to mitigate Zoom fatigue). Specifically, both videos taught participants digital skills. The misinformation intervention taught participants digital media literacy skills such as how to engage in lateral reading and how to conduct a reverse image search, while the Zoom fatigue control video taught participants digital skills such as how to use features like turning their self-view window on and off.

In addition, to avoid any potential confounding based on experimenter effects (i.e. the identity or tone of the speaker in the videos), we asked the same presenter from the digital media literacy intervention to record the control video on Zoom fatigue. He was instructed to maintain as similar a demeanor and presentational style to his initial recording of the misinformation intervention as possible. Finally, we worked to match the modality of the control video to the intervention by having the same speaker present slides that were organized using the same graphic design elements as the misinformation intervention video. We did not include another control condition, such as a no-video control where participants did not watch any video at all, because strongly matched controls are often considered to be a more rigorous test of control vs experimental group differences and allow for stronger identification of the causal role of the construct of interest (see Mohr et al., 2009). Participants under both conditions took pre-surveys immediately prior to exposure to either the digital literacy intervention or the Zoom fatigue video and post-surveys immediately after exposure.

Regarding the time interval for the pre-post tests in Study 1 and Study 2, we note that data collection for Study 1 lasted from October—December 2021 and from March and April 2022 for Study 2. Participants were asked to complete the pre-test measure before being assigned to watch either the intervention video or the control video. They completed the post-test measure after completing the video.

Measures. Participants completed the same survey measures as Study 1. The measure of digital literacy skills demonstrated good internal reliability at both time points (Cronbach's $a_{t1} = .84$, $a_{t2} = .88$). Compared with Study 1, participants in Study 2 were tested on a greater number of headlines (20) in the headline classification task. As in Study 1, all headlines were counterbalanced to account for potential differences in the political slant of the claims (e.g. supporting conventionally liberal views, supporting conventionally conservative views, apolitical; Pennycook et al., 2021) (see Table S3).

Results and discussion

First, we examined whether the intervention improved the digital media literacy skills of the treatment group relative to the control group. We estimated a series of mixed effects linear regression models where the IVs included condition (1 = treatment group, 0 = control group), time (1 = post-intervention, 0 = pre-intervention), and the interaction between condition and time, which was our key variable of interest. DVs were participants' ratings for each skill. Participants were modeled as random effects.

As shown in Figure 4, people who completed the intervention demonstrated significantly improved understanding of lateral reading ($\beta = .79$, $SE = .11$, $p < .001$), reverse image search ($\beta = .60$, $SE = .10$, $p < .001$), click restraint ($\beta = .69$, $SE = .12$, $p < .001$), monitoring emotional reactions to headlines ($\beta = .69$, $SE = .12$, $p < .001$), and using fact-checking resources ($\beta = .34$, $SE = .12$, $p < .01$) relative to the control group. As in Study 1, these results show the intervention was effective in bolstering participants' digital media literacy skills.

While people's self-reported comprehension of digital literacy skills improved, the likelihood that people investigated headlines to inform their veracity judgments did not change after the intervention (see Figure 5). We estimated a logistic regression model where the DV was whether a participant reported doing research on a given headline (1) or not (0). IVs were time, condition, and their interaction. This interaction was not statistically significant ($\beta = .31$, $SE = .20$, $p = .12$), suggesting that while the intervention improved Study 2's participants' knowledge of core digital media literacy skills, it did not make them more likely to apply those skills when judging the veracity of headlines.

There are several potential reasons why participants in Study 2 did not become more likely to investigate the claims in the detection task. One explanation could be that they were insufficiently motivated to fully engage with the task because they were recruited to participate in the study in exchange for financial compensation (Frey and Jegen, 2001). Relative to the community-based sample where participants may have been incentivized by a desire to protect their community, participants in the RCT may not have felt as strong a need to employ their digital media literacy skills in their own lives. Alternately, participants may have also felt less motivated to investigate the claims

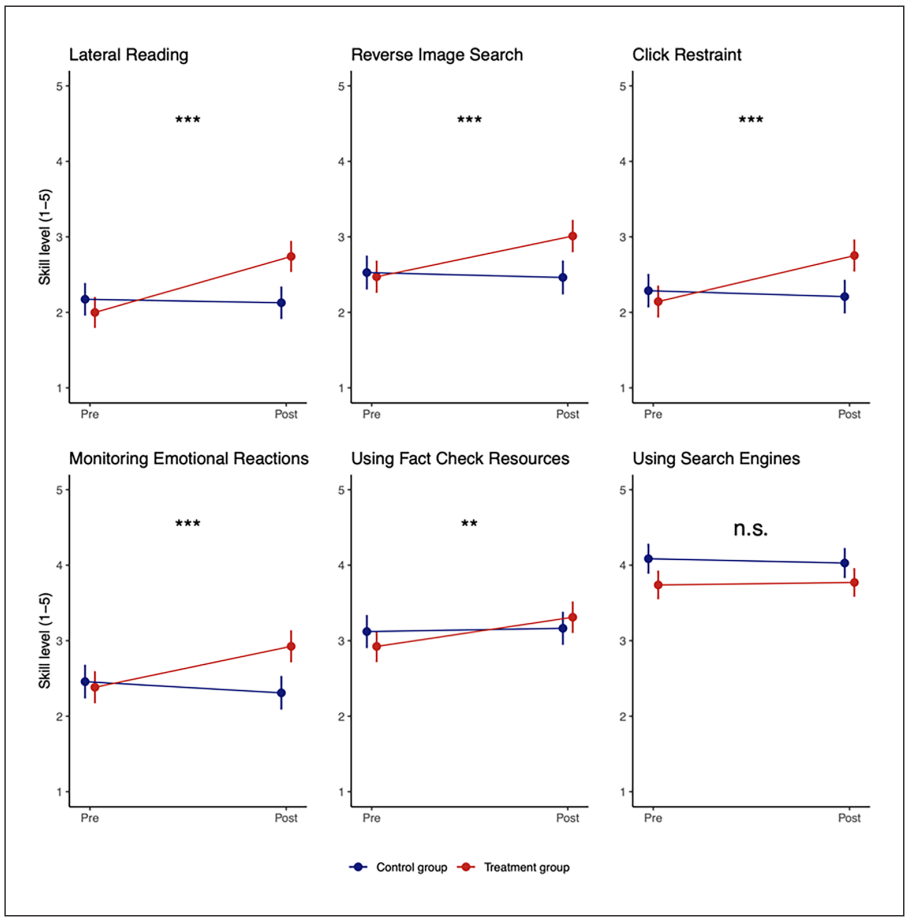

Figure 4. Comparisons of the change in digital media literacy skills between treatment and control groups (Study 2).
The y-axis represents ratings of individuals' comprehension of digital literacy skills taught in the intervention. Red points are ratings from the treatment group and blue points are ratings from the control group. Using search engines was not taught in the intervention and was included as a measure to assess the specificity of the intervention's effect on skills. Bars indicate 95% confidence intervals. Significance markers represent the statistical significance of the interaction between time and condition.
$^{**}p < .01$; $^{***}p < .001$; n.s. $p > .1$.

because Study 2 was conducted in an individualized manner (i.e. each participant watched video recordings of the intervention and completed it on their own). In contrast, the community-based nature of Study 1 involved multiple people participating in workshops synchronously, which may have increased participants' willingness to put their digital media literacy skills to use. The General Discussion section contains further discussion of these factors.

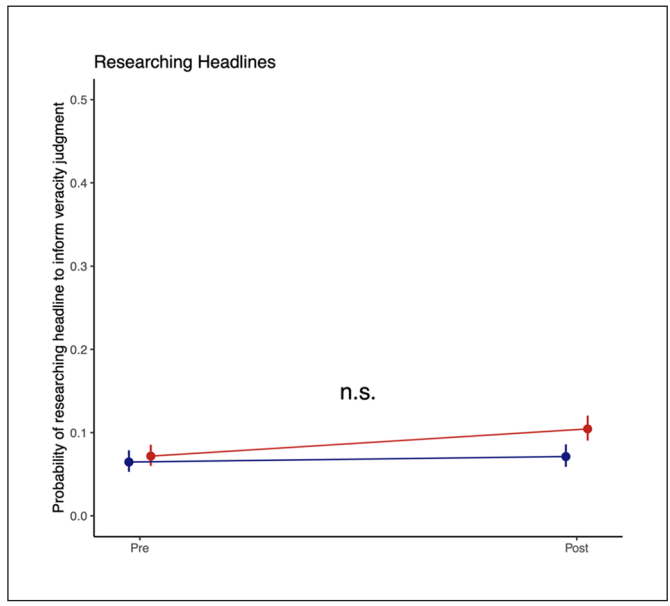

Figure 5. Comparisons of the change in likelihood of researching headlines to inform veracity judgments between treatment and control groups (Study 2).
The y-axis indicates probabilities of individuals reporting they researched a headline online to inform their judgment of its veracity. Red points are likelihoods from the treatment group and blue points are likelihoods from the control group. Bars indicate 95% confidence intervals. The significance marker represents the statistical significance of the interaction between time and condition.
n.s. $p > .1$.

In Study 2, the intervention did not improve treated participants' ability to detect true and false news relative to the control group (see Figure 6). To determine that, we estimated a series of mixed effects logistic regression models in which the DV was whether a headline veracity judgment was accurate (1 = accurate, 0 = inaccurate). IVs were time, condition, and their interaction and participants were modeled as random effects. Overall, there was no significant interaction between time and condition ($\beta = -.07$, $SE = .11$, $p = .55$), suggesting that the intervention did not improve people's ability to identify true and false headlines over the control group. The interaction was also non-significant when analyzing false headlines ($\beta = -.11$, $SE = .18$, $p = .52$) and true headlines ($\beta = -.05$, $SE = .17$, $p = .78$) separately.

Overall, these results indicate that digital media literacy interventions administered through panel recruitment services to members of marginalized communities can be successful in improving their knowledge of approaches to addressing misinformation but require strategies to support participants in applying these skills when evaluating the veracity of online news. While we found the intervention improved digital media literacy skills in Study 2, it did not improve participants' application of these skills to researching headlines, or their ability to discern between true and false news in a behavioral task.

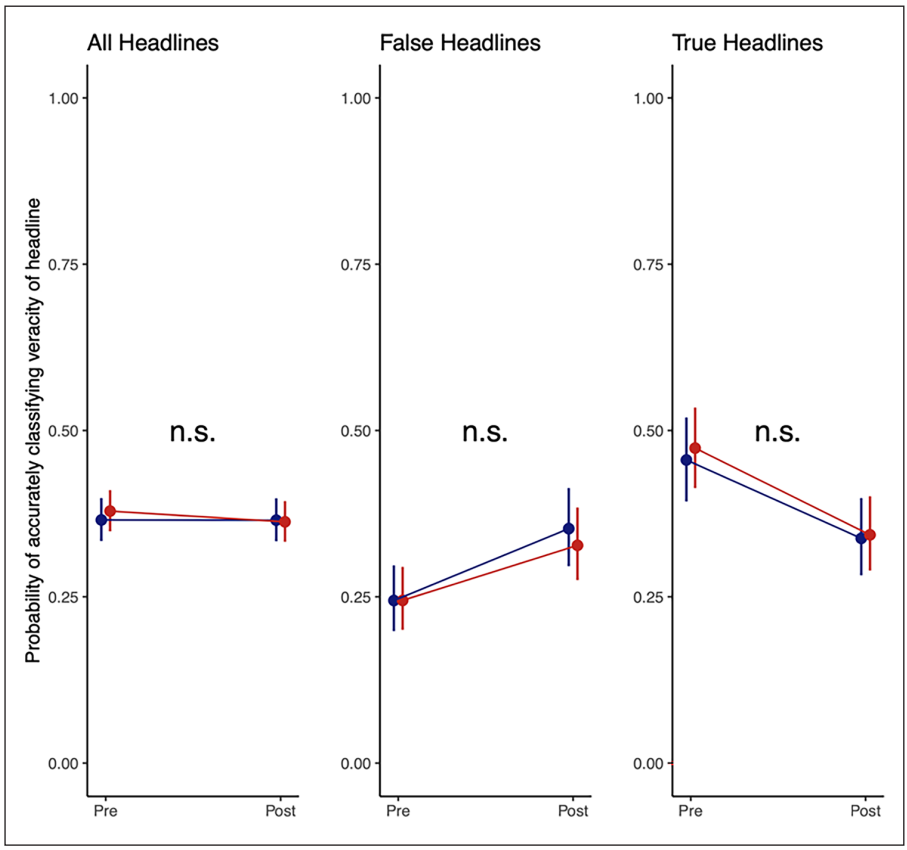

Figure 6. Comparisons of the change in news veracity detection abilities between treatment and control groups (Study 2).
The y-axis indicates predicted probabilities of accurately identifying the veracity of a given headline. Red points are likelihoods from the treatment group and blue points are likelihoods from the control group. Bars indicate 95% confidence intervals. Significance markers represent the statistical significance of the interaction between time and condition.
n.s. $p > .1$.

These results were at odds with the findings from our community-based study, where skills, research intentions, and veracity judgments all improved after the intervention. To better understand this pattern of findings, we conducted a post hoc analysis between the groups that received the intervention in our two studies (i.e. the full sample from Study 1 and the treatment group from Study 2).

Comparing intervention effects between community-recruited and panel-recruited samples

First, we examine the intervention's effect on treated participants in both studies' digital media literacy skills. Like in the analyses for Study 1 and Study 2, we fit mixed-effects

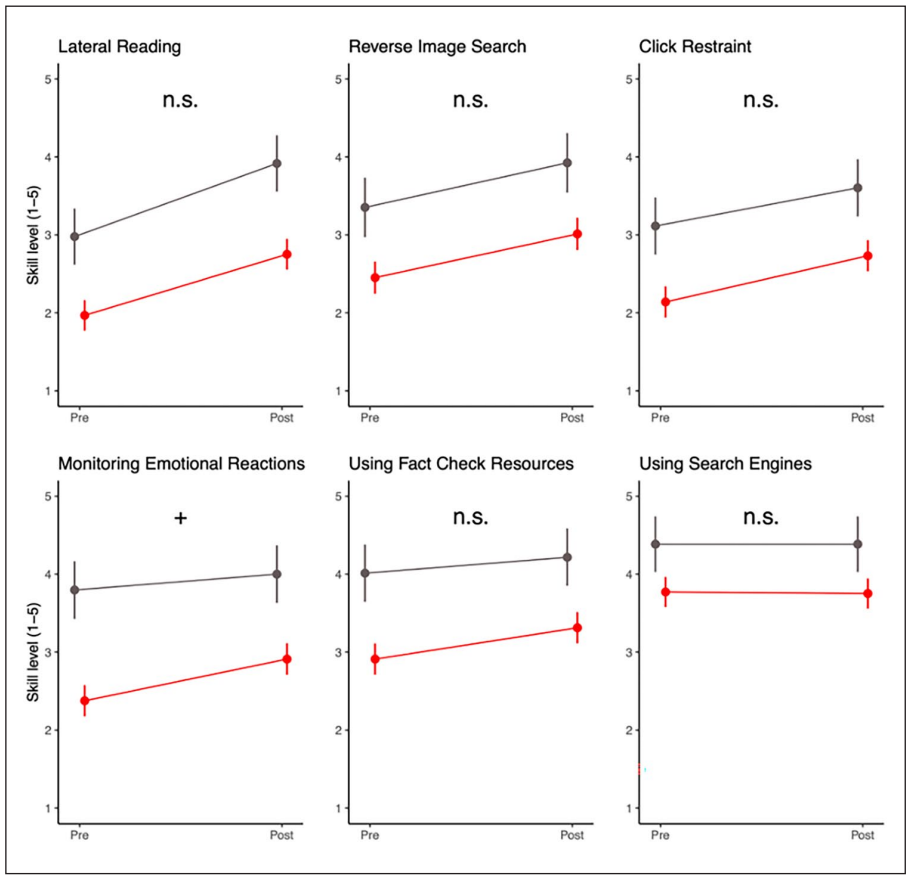

Figure 7. Comparisons of digital media literacy skill improvement between community- and panel-recruited samples.
The y-axis represents ratings of individuals' comprehension of digital literacy skills taught in the intervention. Black points represent participants from Study 1 (community-based study), and red points represent participants from the treatment group of Study 2 (panel-based study). Bars indicate 95% confidence intervals. Significance markers represent the statistical significance of the interaction between time and sample. $+ p < .1$; n.s. $p > .1$.

linear regressions in which the DVs were participants' self-reported comprehension level for a given digital literacy skill. IVs were time (1 = post-intervention, 0 = pre-intervention), sample (1 = Study 2 treatment group, 0 = Study 1 sample), and the interaction between time and sample. Participants were modeled as random effects. As can be seen in Figure 7, there was no significant difference between the two groups in improvement on any of the digital literacy skills taught in the intervention, including lateral reading ($\beta = -.16$, $SE = .20$, $p = .43$), reverse image searching ($\beta = -.01$, $SE = .17$, $p = .95$), click restraint ($\beta = .10$, $SE = .19$, $p = .60$), monitoring emotional reactions to headlines ($\beta = .33$, $SE = .20$, $p > .05$), and using fact-checking resources ($\beta = .20$, $SE = .18$, $p = .28$). This indicates that the

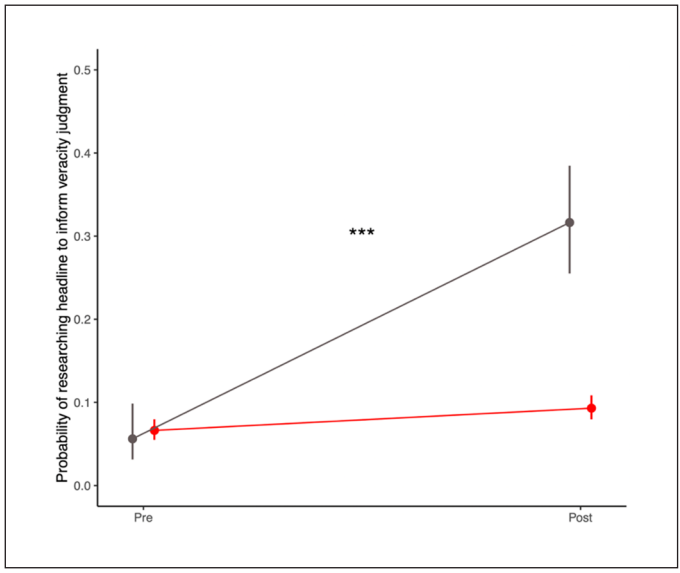

Figure 8. Comparisons of change in researching headlines between community- and panel-recruited samples.
The y-axis indicates probabilities of individuals reporting they researched a headline online to inform their judgment of its veracity. Black points represent likelihoods from participants in Study 1 and red points represent likelihoods of participants in the treatment group in Study 2. Bars indicate 95% confidence intervals. Significance markers represent the statistical significance of the interaction between the time and sample.
***$p < .001$.

intervention appeared to improve skills to a similar extent across both samples. Interestingly, at baseline, participants recruited through community outreach reported substantially higher digital media literacy skill levels than those recruited via YouGov.

Next, we tested whether there were differences in the application of digital media literacy skills between the samples by examining participants' likelihood of doing research on headlines. We estimated a logistic regression in which the DV was whether (1) or not (0) an individual reported doing research on a headline to inform their judgment of its veracity. IVs were time, sample, and their interaction. Although neither sample was very likely to report doing research on headlines at baseline, the community-recruited sample became significantly more likely to do research on headlines after the intervention compared to the panel-recruited sample ($\beta = -1.68$, $SE = 0.37$, $p < .001$). After the intervention, the community-recruited sample reported doing research 32% (95% CI = 26–38%) of the time, while the panel-recruited sample reported doing research only 9% of the time (95% CI = 8–11%) (see Figure 8).

Finally, we examined whether the intervention's effect on people's likelihood of accurately identifying the veracity of online news differed between the two samples. We estimated mixed effects logistic regressions in which the DV was whether a headline judgment

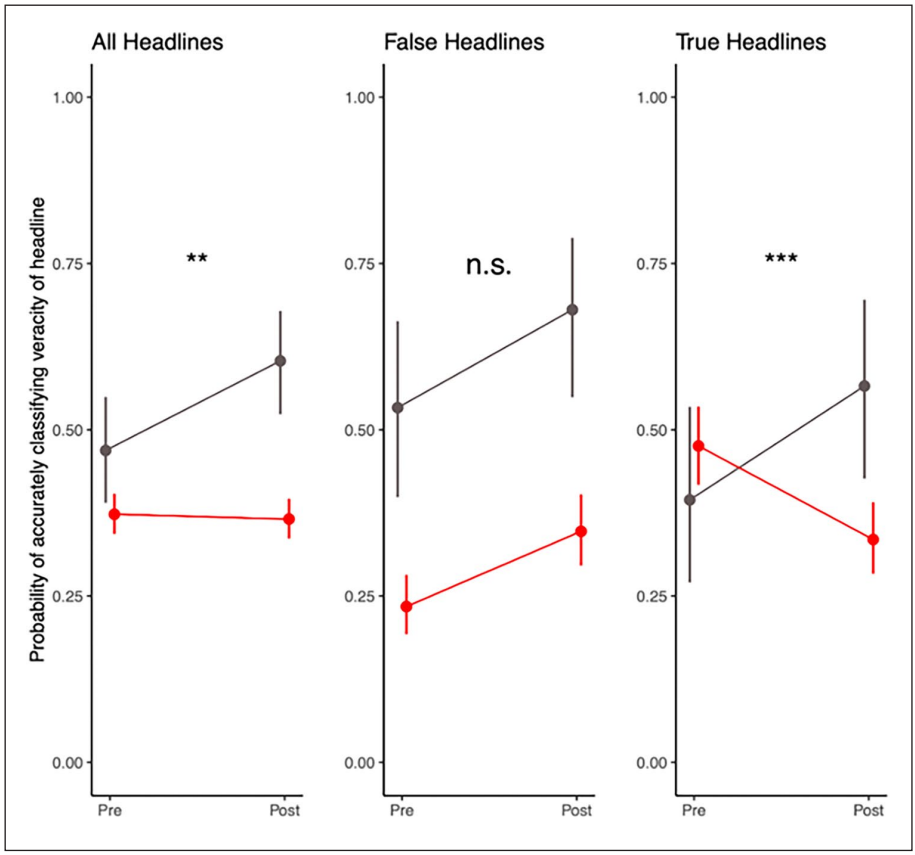

Figure 9. Comparisons of change in news veracity detection abilities between community- and panel-recruited samples.
The y-axis indicates predicted probabilities of accurately identifying the veracity of a given headline. Black points represent likelihoods from participants in Study 1 and red points represent likelihoods of participants in the treatment group in Study 2. Bars indicate 95% confidence intervals. Significance markers represent the statistical significance of the interaction between time and sample.
$^{**}p < .01$; $^{***}p < .001$; n.s. $p > .1$.

was accurate (1) or inaccurate (0). IVs were sample, time, and their interaction. Participants were modeled as random effects. Results revealed significant differences in the effect of the intervention on likelihoods of accurately identifying true and false headlines ($\beta = -.58$, $SE = .22$, $p < .01$), such that the community sample became more accurate in judging headlines after the intervention than the YouGov sample. Furthermore, while the intervention's effect on the groups' abilities to identify false headlines did not differ ($\beta = -.07$, $SE = .34$, $p = .84$), the intervention significantly improved the community-recruited sample's ability to identify true headlines compared to the YouGov sample ($\beta = -1.28$, $SE = .35$, $p < .001$) (see Figure 9).

General discussion

Through two studies, we examined the impact of a tailored digital media literacy intervention on misinformation resilience in communities of color. In Study 1, we found that the intervention was effective among a community-recruited sample at improving people's digital media literacy skills, their ability to accurately identify true and false online news, and their likelihood of doing research online to inform their judgments about news stories' veracity. However, Study 2 revealed that, among a sample recruited by YouGov, the intervention only improved participants' digital literacy skills; there was no improvement in their ability to discern true and false news nor their likelihood to apply their skills by doing research to inform their veracity judgments.

Studying the effects of the intervention in different contexts allowed us to identify three considerations that we argue are crucial to building interventions that improve misinformation resilience in the real world. First, our study highlights the need to *engage with diverse communities.* In addition to diversifying the populations targeted by interventions (i.e. non-WEIRD samples, or those beyond those that are White, Educated, Industrialized, Rich, and Democratic), our work demonstrates the value obtained from collaborating with community organizations in designing and implementing the intervention itself. Second, the discrepancy in effects between our community-based and panel-based studies emphasizes the need to consider the *social affordances* of the intervention context: the ways in which the social elements of the intervention context interact with individuals' experiences. Experiencing the intervention as part of a synchronous workshop organized by a trusted organization (e.g. in the community-based study) rather than as an independent online activity (e.g. in the panel-recruited study) may influence the perception, uptake, and application of intervention content. Finally, our findings speak to the need for research to consider the *motivations of individuals* who choose to participate in the interventions. Whereas panel-recruited participants may have been motivated by a financial incentive, community-recruited participants may have been compelled by a desire to protect their communities from misinformation, or to develop their own skills. Below, we discuss the implications of these three considerations for research on misinformation and interventions in greater detail.

Engaging diverse communities in (mis)information interventions

Our findings underscore the need for (mis)information studies to engage with individuals from diverse communities. To our knowledge, our studies are the first to evaluate a digital media literacy intervention with BIPOC communities in the United States. At a fundamental level, interventions designed to bolster misinformation resilience should be shared with populations that are ethnically and culturally diverse. From a methodological perspective, testing interventions in diverse contexts is essential to evaluating their generalizability and identifying the causes of heterogeneous effects (Walton and Yeager, 2020). Indeed, scholars and practitioners have highlighted the need for efforts combating the misinformation crisis to increase their scope (Nguyễn et al., 2022). In a recent commentary, Camargo and Simon (2022) emphasize that research should "consider in more

detail the role of race, gender, and class" as well as the ways in which (mis)information is experienced within diverse communities' sociocultural contexts. To date, most work has developed and administered interventions for predominantly WEIRD populations—a problem across the behavioral sciences (Rad et al., 2018). While a growing body of work has examined the impact of digital media literacy interventions in non-Western countries (e.g. India; Guess et al., 2020), it is also essential to consider sample diversity within diverse American contexts (Ghai, 2021). Perceptions and responses to online (mis)information vary significantly within and across American communities of color (Dodson et al., 2021; Nguyễn et al., 2022). For instance, Abrajano et al. (2022) found that Latinos in the United States who use Spanish-language social media tend to be exposed to more misinformation than those who instead frequent English-based platforms.

Furthermore, our findings speak to the value of engaging diverse communities in the development and dissemination of interventions. By integrating the perspectives of BIPOC community leaders from partner non-profit organizations (see Lee et al., 2023), the digital media literacy interventions were tailored to the needs of communities of color. Specifically, the intervention we study in this article (1) accounted for the different media ecologies used by diverse communities (i.e. some groups use certain platforms more than others), (2) included multilingual options to increase accessibility, and (3) acknowledged the ways in which misinformation narratives can reference historical injustices and community concerns to target marginalized groups.

Considering social context

Our findings also speak to the importance of considering misinformation interventions in broader social contexts. Efforts to combat misinformation are often discussed at the individual level, such as encouraging individuals to apply digital media literacy skills to verify online information (e.g. McGrew et al., 2018) or to pause and consider the accuracy of a claim before sharing it (e.g. Pennycook et al., 2021). However, as Camargo and Simon (2022) note, individuals' exposure and responses to misinformation are shaped by the nuances of their social, political, and cultural contexts. For instance, qualitative interviews with AAPI individuals found that they are targeted by purveyors who seek to exploit cultural beliefs (e.g. in traditional Eastern medicine) (Nguyễn et al., 2022; Lee et al., 2023). Furthermore, they often encountered misinformation in the context of their preexisting social networks, such as when friends forwarded fabricated news articles to WhatsApp or WeChat group chats (Lee et al., 2023). Therefore, it is important for interventions to teach individuals how to navigate false claims in context.

Considering the interaction between individual-level interventions and the features of participants' social context can also advance research by improving our understanding of the conditions under which interventions are most effective (Koppel et al., 2023). Though the content of the intervention was similar between Study 1 and Study 2, participants recruited through community-engaged methods (Study 1) improved their ability to identify true and false information whereas those recruited from a survey company (Study 2) did not. These different treatment effects suggest that features of the two study contexts may have influenced the intervention's efficacy. For instance, while intervention-takers

in both Studies 1 and 2 improved their understanding of digital media literacy skills, only those in Study 1 saw gains in their ability to accurately identify true and false information in the behavioral task. One possibility is that the social affordances of Study 1 better supported those participants' application of the skills they learned and thus their greater accuracy in the behavioral task.

In their work on psychological interventions, Walton and Yeager (2020) offer the metaphor of "seeds and soil" to explain heterogeneous treatment effects. For an intervention to yield a desired effect (e.g. improved ability to identify true and false news), the intervention must not only be high-quality (a "good seed") but also be placed into "fertile soil": a social context that supports individuals in maintaining the new skills or belief systems taught in the intervention. For example, in our community-recruited sample, participants were sent information about the intervention from organizations that they know and trust, who identified the intervention as relevant and suitable for their community. This endorsement may have affected community-recruited participants' engagement with the intervention content, and thus the intervention's efficacy. By contrast, those recruited via YouGov were not exposed to such strong endorsement cues. Relatedly, when the opportunity to participate in the intervention was presented to community-recruited individuals, the recruitment materials made salient misinformation's threat to their communities and the value of improving one's ability to identify it. In comparison, YouGov panelists who elected to participate were told they would simply watch a video and answer questions about identifying false information. In receiving a more intense prime about misinformation, the community-recruited participants may have gone into the intervention in a more attentive way, motivated to learn as much as possible.

Indeed, working with community-based organizations to distribute the intervention to diverse participants may yield additional advantages such as improving potential positive spillover effects of digital media literacy interventions. Specifically, collaborating with community organizations also allowed the intervention to tap into existing community networks by serving as trusted intermediaries. This distribution strategy may also help catalyze further intervention participation and benefits by convening participants through pre-existing relationships (e.g. mailing lists to their members, community-based events) who may then go on to share what they learned with their networks (e.g. friends and family).

Certain technological affordances of the intervention contexts may have also supported individuals in the community-based study in better achieving the goals of the intervention. Just as theories of psychological affordances consider whether the social context affords the way of thinking offered by the intervention (Walton and Yeager, 2020), theories of technological affordances consider whether the properties of a digital context afford certain ways of thinking about and engaging with digital content (Evans et al., 2017). For instance, the synchronous nature of the community-based intervention may have increased a sense of *co-presence*—an important feature of computer-mediated communication that can increase engagement in online tasks (Luo, 2021). Seeing peers completing the intervention with them and feeling that they are part of a communal educational experience may have supported Study 1 participants' internalization of the taught digital media literacy skills, unlike the panel-recruited participants who completed the intervention independently and asynchronously.

Recognizing diverse motivations for intervention participation

Finally, our findings emphasize the need to consider the *motivations* that individuals bring to misinformation interventions. In addition to the social affordances mentioned above, another potential reason the participants in our community-based sample demonstrated improvement in their ability to identify true and false news, as well as their knowledge of digital media literacy skills, may be because they were more personally motivated to engage with the intervention materials than those in the panel-based study. Future work should consider systematically investigating the effect of different motivational frames on educational intervention uptake and engagement.

After all, people can have many different reasons for why they might want to participate in a digital media literacy intervention, such as wanting to improve their own skills or wanting to protect members of their community, or simply because of financial incentives involved—each of which may influence their engagement with the intervention materials and, importantly, whether they actually apply the skills they learn after the intervention. Understanding such differences in personal motivations and addressing this "motivational challenge" has been essential to designing effective behavioral interventions in past research (Landry and Halperin, 2023). For instance, research on smoking cessation shows individuals are most successful in quitting smoking when strongly internally motivated (Schlagintweit et al., 2017; Wiseman et al., 2019). Similarly, efforts to build misinformation resilience may be most effective when they speak to and strengthen intrinsic motivations, like wanting to avoid manipulation by misinformation purveyors or to build a safer media environment for others.

Limitations

Our studies should be interpreted in light of their limitations. Detailed limitations of Study 1 are presented in the discussion of Study 1. Overall, Study 1 featured a high level of external validity in that participants in the interventions participated in a naturalistic setting, under the conditions they would have otherwise participated outside of our study, which traded off with internal validity. The lack of control condition and random assignment to the intervention means that Study 1 could not credibly identify the causal effects of the tailored interventions. Study 2, on the other hand, was designed to improve internal validity and allow us to estimate the causal effects of the intervention on a larger and representative sample. However, these methodological improvements came with important tradeoffs that limited Study 2's external validity—participants in Study 2 did not come to participate in the intervention under naturalistic conditions and were instead paid to participate. Furthermore, Study 2 participants focused on evaluating the effect of the intervention developed for the Latino community. Future research should conduct systematic randomized controlled trials on the effects of the interventions on each population to understand the nuances of heterogeneous intervention effects across diverse communities of color.

There are also global limitations that affect both studies. First, our studies relied exclusively on survey-based measures to determine intervention efficacy. For example, while the headline-identification task is commonplace in misinformation research

(Pennycook et al., 2021), it does not fully reflect the experience of encountering news in everyday life. For example, contextual information such as source and social endorsement cues were absent from the stimuli we used, despite being prevalent online (Luo et al., 2022). Future work should endeavor to collect naturalistic digital trace data from participants in media literacy interventions to better understand their effects on subsequent real-world media behaviors (for example, see Aslett et al., 2022).

Second, our evaluations consist of a pre- and postintervention survey, administered immediately before and after participants took the intervention. Although the intervention was highly effective among our community-recruited sample, those benefits may have decayed shortly after the intervention ended. To understand the durability of the interventions' effects, it is essential to follow up with people days and weeks afterward. Third, the intervention itself was multimodal and complex, teaching a number of different digital literacy skills through a combination of text, photographs, and videos. As a consequence, we cannot speak to the specific components of the intervention that were most or least effective. Future work should break apart longer-form interventions to better understand the returns to teaching about certain skills via certain modalities, useful for real-world organizations that may have an easier time scaling shorter-form content to a greater number of users (Fagan et al., 2019).

Finally, we note that our studies demonstrate that the tailored intervention was effective at improving digital media literacy skills among communities of color in certain contexts, but cannot attest to their comparative effectiveness relative to non-tailored or "generic" intervention approaches. To our knowledge, no prior studies have yet investigated how tailoring misinformation education interventions shapes participants' learning, engagement, and outcomes. While Ali and Qazi (2023) found that providing personalized feedback on individuals' ability to accurately detect true and false news can strengthen intervention efficacy, the intervention focuses on tailoring educational interventions to communities' perspectives and experiences rather than individual performance. However, research from health communication indicates that tailoring educational interventions can increase skill-building and behavioral change because participants perceive them as more personalized, enjoyable, and relevant (for a review, see Ryan and Lauver, 2002). Indeed, it may be that participants from underrepresented or historically underserved communities may be more motivated to participate in tailored interventions delivered by trusted messengers in ways that align with their interests and concerns. Therefore, an important priority for future research is to investigate the effectiveness of tailored misinformation interventions against generic interventions on not only digital media literacy skill comprehension and discriminant trust in the news but also people's perceptions of the intervention being relevant to their own experiences.

Conclusion

In conclusion, we provide evidence that a digital media literacy intervention tailored for BIPOC communities can be effective at building resilience to misinformation, but also that why and how people participate in such interventions matters and is an essential area for future research. We hope our work highlights the importance of recognizing the

different media and (mis)information experiences of members of diverse communities and incorporating those into media literacy interventions serving those populations. While we want to make clear that the communities we worked with in this paper are by no means monoliths and that there is more work to do to make interventions more relevant, responsive, and equitable for all, the ideas and practices that we highlight in this paper may be a useful step in diversifying media literacy education, and misinformation research more broadly.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: AYL and RCM are supported by the Stanford Interdisciplinary Graduate Fellowship and the Stanford Impact Labs PhD Fellowship. JTH received an unrestricted academic gift from PEN America, which supported data collection costs. PEN America had no role in data analysis, decision to publish, or preparation of the manuscript beyond minor grammatical changes.

ORCID iDs

Angela Y Lee https://orcid.org/0000-0002-9527-5730

Ryan C Moore https://orcid.org/0000-0002-5669-721X

Supplemental material

Supplemental material for this article is available online.

Note

1. Examples of the interventions are available online as recordings at https://www.youtube.com/watch?v=qVJQB29AG7o.

References

Abrajano M, Garcia M, Pope A, et al. (2022) Social media information politics: insights on Latinos in the US. *OSF*. Available at: https://bit.ly/CSIP-Paper1

Ali A and Qazi IA (2023) Countering misinformation on social media through educational interventions: evidence from a randomized experiment in Pakistan. *Journal of Development Economics* 163: 103108.

Allen J, Howland B, Mobius M, et al. (2020) Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* 6(14): eaay3539.

Amazeen MA, Vasquez RA, Krishna A, et al. (2024) Missing voices: examining how misinformation-susceptible individuals from underrepresented communities engage, perceive, and combat science misinformation. *Science Communication* 46(1): 3–35. DOI: 10.1177/10755470231217536.

Aslett K, Guess AM, Bonneau R, et al. (2022) News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions. *Science Advances* 8(18): eabl3844.

Axelsson CAW, Guath M and Nygren T (2021) Learning how to separate fake from real news: scalable digital tutorials promoting students' civic online reasoning. *Future Internet* 13(3): 60.

Badrinathan S (2021) Educative interventions to combat misinformation: evidence from a field experiment in India. *American Political Science Review* 115: 1325–1341.

Basol M, Roozenbeek J, Berriche M, et al. (2021) Towards psychological herd immunity: cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society* 8(1): 20539517211013868.

Bastos M and Farkas J (2019) "Donald Trump is my president!": the Internet research agency propaganda machine. *Social Media + Society* 5(3): 2056305119865466.

Blair RA, Gottlieb J, Nyhan B, et al. (2023) Interventions to counter misinformation: lessons from the global north and applications to the global south. *Current Opinion in Psychology* 101732.

Bogart LM and Uyeda K (2009) Community-based participatory research: partnering with communities for effective and sustainable behavioral health interventions. *Health Psychology* 28(4): 391–393.

Camargo CQ and Simon FM (2022) Mis-and disinformation studies are too big to fail: six suggestions for the field's future. *Harvard Kennedy School Misinformation Review*. Available at: https://misinforeview.hks.harvard.edu/article/mis-and-disinformation-studies-are-too-big-to-fail-six-suggestions-for-the-fields-future/

Cao J, Qi P, Sheng Q, et al. (2020) Exploring the role of visual content in fake news detection. In: Shu K, Wang S, Lee D, et al. (eds) *Disinformation, Misinformation, and Fake News in Social Media*. Cham: Springer, pp. 141–161.

Cooke NA (2017) Posttruth, truthiness, and alternative facts: information behavior and critical information consumption for a new age. *The Library Quarterly* 87(3): 211–221.

Cortina J and Rottinghaus B (2022) With the 2022 midterms ahead, expect another Latino misinformation crisis. *The Washington Post*, 25 February. Available at: https://www.washingtonpost.com/politics/2022/02/25/latino-misinformation-spanish-social-media/

Dahlke R and Hancock J (2023, November 17) The effect of online misinformation exposure on false election beliefs. Available at: https://doi.org/10.31219/osf.io/325tn

De Freitas Melo P, Vieira CC, Garimella K, et al. (2019) Can WhatsApp counter misinformation by limiting message forwarding? In: Cherifi H, Gaito S, Mendes J, et al. (eds) *International Conference on Complex Networks and Their Applications*. Cham: Springer, pp. 372–384.

Dodson K, Mason J, Smith R, et al. (2021) First draft news. Available at: https://firstdraftnews.org/long-form-article/covid-19-vaccine-misinformation-black-communities/

Evans SK, Pearce KE, Vitak J, et al. (2017) Explicating affordances: a conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication* 22(1): 35–52.

Fagan AA, Bumbarger BK, Barth RP, et al. (2019) Scaling up evidence-based interventions in US public systems to prevent behavioral health problems: challenges and opportunities. *Prevention Science* 20(8): 1147–1168.

Fauville G, Luo M, Queiroz ACM, et al. (2023) Video-conferencing usage dynamics and nonverbal mechanisms exacerbate Zoom Fatigue, particularly for women. *Computers in Human Behavior Reports* 10: 100271.

Forman-Katz N and Matsa K (2020) News platform fact sheet. Pew Research Center. Available at: https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/

Frey BS and Jegen R (2001) Motivation crowding theory. *Journal of Economic Surveys* 15(5): 589–611.

Garimella K and Eckles D (2020) Images and misinformation in political groups: evidence from WhatsApp in India. ArXiv [Preprint]. Available at: https://arxiv.org/abs/2005.09784

Getahun H (2021) COVID misinformation plagues California's indigenous speakers. *CalMatters*, 10 November. Available at: https://calmatters.org/health/coronavirus/2021/11/covid-indigenous-language-barriers/

Ghai S (2021) It's time to reimagine sample diversity and retire the WEIRD dichotomy. *Nature Human Behaviour* 5(8): 8.

Green J, Hobbs W, McCabe S, et al. (2022) Online engagement with 2020 election misinformation and turnout in the 2021 Georgia runoff election. *Proceedings of the National Academy of Sciences* 119(34): e2115900119.

Griffin JP Jr, Yancey E and Armstrong-Mensah E (2013) Building partnerships in community-based participatory research. In: Blumenthal DS and DiClemente RJ (eds) *Community-Based Participatory Health Research: Issues, Methods, and Translation to Practice*. New York: Springer, pp. 19–34.

Guess A and Munger K (2023) Digital literacy and online political behavior. *Political Science Research and Methods* 11(1): 110–128.

Guess AM, Lerner M, Lyons B, et al. (2020) A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences* 117(27): 15536–15545.

Hargittai E (2002) Beyond logs and surveys: in-depth measures of people's web use skills. *Journal of the American Society for Information Science and Technology* 53(14): 1239–1244.

Hargittai E (2005) Survey measures of web-oriented digital literacy. *Social Science Computer Review* 23(3): 371–379.

Hargittai E and Hsieh YP (2011) Succinct survey measures of web-use skills. *Social Science Computer Review* 29(3): 391–395.

Harris AD, McGregor JC, Perencevich EN, et al. (2006) The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical Informatics Association* 13(1): 16–23.

Hill L and Artiga S (2022) COVID-19 cases and deaths by race/ethnicity: current data and changes over time. *Kaiser Family Foundation*, 22 February. Available at: https://www.kff.org/coronavirus-covid-19/issue-brief/covid-19-cases-and-deaths-by-race-ethnicity-current-data-and-changes-over-time/

Igielnik R and Budiman A (2020) The changing racial and ethnic composition of the US electorate. Available at: https://www.pewresearch.org/2020/09/23/the-changing-racial-and-ethnic-composition-of-the-u-s-electorate/

Koppel L, Robertson C, Doell KC, et al. (2023) Individual-level solutions may support system-level change â if they are internalized as part of one's social identity. *Behavioral and Brain Sciences* 46: e165.

Landry A and Halperin E (2023) Intergroup psychological interventions: the motivational challenge. *American Psychologist*. Epub ahead of print 7 December. DOI: 10.1037/amp0001289.

Lee AY and Hancock JT (2023) Developing digital resilience: an educational intervention improves elementary students' response to digital challenges. *Computers and Education Open* 100144.

Lee AY, Moore RC and Hancock JT (2023) Designing misinformation interventions for all: perspectives from AAPI, Black, Latino, and Native American community leaders on misinformation educational efforts. *Harvard Kennedy School Misinformation Review*.

Longoria J, Acosta D, Urbani S, et al. (2021) A limiting lens: how vaccine misinformation has influenced Hispanic conversations online. Available at: https://firstdraftnews.org/long-form-article/covid19-vaccine-misinformation-hispanic-latinx-social-media/

Loomba S, de Figueiredo A, Piatek SJ, et al. (2021) Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour* 5(3): 337–348.

Luo M (2021) *What Makes Live Streaming Enjoyable? Understanding the Impacts of Live Streaming Affordances on Hedonic Media Experiences*. Stanford University.

Luo M, Hancock JT and Markowitz DM (2022) Credibility perceptions and detection accuracy of fake news headlines on social media: effects of truth-bias and endorsement cues. *Communication Research* 49(2): 171–195.

Masten AS and Barnes AJ (2018) Resilience in children: developmental perspectives. *Children* 5(7): 98.

McGrew S (2020) Learning to evaluate: an intervention in civic online reasoning. *Computers & Education* 145: 103711.

McGrew S, Breakstone J, Ortega T, et al. (2018) Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education* 46(2): 165–193.

Mochkofsky G (2022) The Latinx community and COVID-disinformation campaigns. *The New Yorker*, 14 January. Available at: https://www.newyorker.com/news/daily-comment/the-latinx-community-and-covid-disinformation-campaigns

Mohr DC, Spring B, Freedland KE, et al. (2009) The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychotherapy and Psychosomatics* 78(5): 275–284.

Moore RC and Hancock JT (2022) A digital media literacy intervention for older adults improves resilience to fake news. *Scientific Reports* 12(1): 6008.

Morgan V, Auskova A and Janoskova K (2021) Pervasive misinformation, COVID-19 vaccine hesitancy, and lack of trust in science. *Review of Contemporary Philosophy* 20: 128–138.

Muda R, Pennycook G, Hamerski D, et al. (2023) People are worse at detecting fake news in their foreign language. *Journal of Experimental Psychology: Applied* 29(4): 712–724. DOI: 10.1037/xap0000475.

Murphy GL (2003) Ecological validity and the study of concepts. *Psychology of Learning and Motivation*, 43, 1–42.

Navia T (2021) COVID misinformation is running rampant online, and it's worse in Spanish. *Vice*, 24 November. Available at: https://www.vice.com/en/article/jgm8z8/covid-misinformation-is-running-rampant-online-and-its-worse-in-spanish

Nguyen A and Catalan-Matamoros D (2020) Digital mis/disinformation and public engagement with health and science controversies: fresh perspectives from Covid-19. *Media and Communication* 8(2): 323–328.

Nguyễn S, Kuo R, Reddi M, et al. (2022) Studying mis-and disinformation in Asian diasporic communities: the need for critical transnational research beyond Anglocentrism. *Harvard Kennedy School Misinformation Review*. Available at: https://misinforeview.hks.harvard.edu/article/studying-mis-and-disinformation-in-asian-diasporic-communities-the-need-for-critical-transnational-research-beyond-anglocentrism/

Nygren T, Guath M, Axelsson CAW, et al. (2021) Combatting visual fake news with a professional fact-checking tool in education in France, Romania, Spain and Sweden. *Information* 12(5): 201.

Ognyanova K, Lazer D, Robertson RE, et al. (2020) Misinformation in action: fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*. Available at: https://misinforeview.hks.harvard.edu/article/misinformation-in-action-fake-news-exposure-is-linked-to-lower-trust-in-media-higher-trust-in-government-when-your-side-is-in-power/

Orne MT and Holland CH (1968) On the ecological validity of laboratory deceptions. *International Journal of Psychiatry* 6(4): 282–293.

Ostler J (2020) Disease has never been just disease for Native Americans. *The Atlantic*, 19 April. Available at: https://www.theatlantic.com/ideas/archive/2020/04/disease-has-never-been-just-disease-native-americans/610852/

Pang H (2022) Connecting mobile social media with psychosocial well-being: understanding relationship between WeChat involvement, network characteristics, online capital and life satisfaction. *Social Networks* 68: 256–263.

Peng W, Lim S and Meng J (2023) Persuasive strategies in online health misinformation: a systematic review. *Information, Communication & Society* 26: 2131–2148.
Pennycook G, Binnendyk J, Newton C, et al. (2021) A practical guide to doing behavioral research on fake news and misinformation. *Collabra: Psychology* 7(1): 25293.
Rad MS, Martingano AJ and Ginges J (2018) Toward a psychology of Homo sapiens: making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences* 115(45): 11401–11405.
Rathje S, Van Bavel JJ and Van der Linden S (2021) Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences* 118(26): e2024292118.
Roozenbeek J, Maertens R, McClanahan W, et al. (2021) Disentangling item and testing effects in inoculation research on online misinformation: Solomon revisited. *Educational and Psychological Measurement* 81(2): 340–362.
Ryan P and Lauver DR (2002) The efficacy of tailored interventions. *Journal of Nursing Scholarship* 34(4): 331–337.
Schlagintweit HE, Campbell NK and Barrett SP (2017) Quit intentions moderate subjective and physiological responses to acute nicotine replacement therapy administration in dependent smokers. *Nicotine & Tobacco Research* 19(8): 922–929.
Schmuckler MA (2001) What is ecological validity? A dimensional analysis. *Infancy* 2(4): 419–436.
Tran D (2020) Is Trump really anti-socialist and anti-communist? *Viet Fact Check*, 18 September. Available at: https://vietfactcheck.org/2020/09/18/is-trump-really-anti-socialist-and-anti-communist/
Troy AS, Willroth EC, Shallcross AJ, et al. (2023) Psychological resilience: an affect-regulation framework. *Annual Review of Psychology* 74: 547–576.
Uribe M (2022) Spanish-language misinformation often goes unchecked, but individuals can help combat it. *Poynter*. Available at: https://www.poynter.org/fact-checking/2022/spanish-language-misinformation-often-goes-unchecked-but-individuals-can-help-combat-it/
Walton GM and Yeager DS (2020) Seed and soil: psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science* 29(3): 219–226.
Wilson T and Starbird K (2020) Cross-platform disinformation campaigns: lessons learned and next steps. *Harvard Kennedy School Misinformation Review*. Available at: https://misinforeview.hks.harvard.edu/article/cross-platform-disinformation-campaigns/#:~:text=Based%20on%20what%20we%20learned,detect%20and%20characterize%20disinformation%20campaigns.
Wineburg S and McGrew S (2019) Lateral reading and the nature of expertise: reading less and learning more when evaluating digital information. *Teachers College Record* 121(11): 1–40.
Wineburg S, Breakstone J, McGrew S, et al. (2022) Lateral reading on the open Internet: a district-wide field study in high school government classes. *Journal of Educational Psychology* 114: 893–909.
Wiseman KP, Coa KI and Prutzman YM (2019) Predictors of retention in an adult text messaging smoking cessation intervention program: cohort study. *JMIR mHealth and uHealth* 7(8): e13712.

Author biographies

Angela Y Lee is a PhD candidate at Stanford University in the Department of Communication. She studies the effects of social technology use on individuals' psychological and informational well-being.

Ryan C Moore is a PhD candidate in the Department of Communication at Stanford University. His research focuses on older adults' digital media use and also explores how people encounter, process, and engage with (mis)information online.

Jeffrey T Hancock is the founding director of the Stanford Social Media Lab and is Harry and Norman Chandler Professor of Communication at Stanford University. Dr. Hancock and his group work on understanding psychological and interpersonal processes in social media.