Pilot study suggests online media literacy programming reduces belief in false news in Indonesia
by Pamela Bilo Thomas, Clark Hogan-Taylor, Michael Yankoski, and Tim Weninger

Abstract
Amidst the threat of digital misinformation, we offer a pilot study regarding the efficacy of an online social media literacy campaign aimed at empowering individuals in Indonesia with skills to help them identify misinformation. We found that users who engaged with our online training materials and educational videos were more likely to identify misinformation than those in our control group (total N=1,000). Given the promising results of our preliminary study, we plan to expand efforts in this area, and build upon lessons learned from this pilot study.

Contents
Introduction
Why Indonesia?
Assessing the online and social media landscape in Indonesia
Methods
Findings
Conclusions

Introduction
While the use of targeted misinformation campaigns by state actors is well documented (Keller, et al., 2020; Zannettou, et al., 2019; Howard and Bradshaw, 2018), ordinary citizens often both consume and spread (mis)information through their well-meaning online activities, which can inadvertently help to spread falsehoods to their friends and followers. Many users of online and social media systems are not aware of how misinformation is generated or spread, and thus they may unwittingly participate in the acceleration or distribution of these materials. Furthermore, the more exposed to misinformation a person becomes, the more likely they are to spread it, for repeated exposure leads to an increased belief of this information (Guess, et al., 2020). Ultimately, healthy democracies depend on a well-informed public, the very possibility of which is undermined by this new deluge of digital misinformation (Lewandowsky, et al., 2017).

In response to these challenges, we present the results of a small pilot study of a media literacy campaign in...
Pilot study suggests online media literacy programming reduces belief in false news in Indonesia.

Indonesia. This pilot study is the first step in a long-running program to understand the effects of digital misinformation on individuals in developing digital economies — those who may not be sufficiently equipped to understand and navigate perils that proliferate on social media platforms. Recent studies have suggested that those without advanced digital skills are more susceptible to the effects of misinformation (Angeline, et al., 2020).

There are many active projects seeking effective ways to stop this spread of misinformation, from new AI-based misinformation detection warning systems (Yankoski, et al., 2020), to “inoculation” style misinformation games such as Harmony Square (Roozenbeek and van der Linden, 2020), to fact checking organizations that meticulously comb through available evidence to affirm or debunk claims, such as Snopes.com in the U.S. or Mafindo in Indonesia. But there is no silver bullet for this accelerating problem: the adversarial nature of misinformation content creation processes means that AI-based misinformation detection systems will likely encourage the development of more sophisticated misinformation campaigns designed to elude the latest generation of AI-based detection systems (Anderson and Rainie, 2017). Furthermore, some argue that fact-checking efforts may backfire by actually increasing the visibility of the very falsehoods they are seeking to debunk, though this effect has been recently contested (Swire-Thompson, et al., 2020). For even if exposure to misinformation comes by seeing a piece of misinformation proven false, once a person has been exposed it is very difficult to turn them back to a state where they have never heard of it (Lewandowsky, et al., 2012).

Many view media literacy and fact checking as the obvious solution(s) to misinformation (Kim and Walker, 2020). For example, watching others be corrected on social media can reduce beliefs in misperceptions (Bode, et al., 2020), and techniques such as gamification (Chang, et al., 2020), inoculation theory, and prebunking can reduce a user’s susceptibility to misinformation (Roozenbeek, et al., 2020; Lee, 2018). Indeed, social media platforms like Facebook and Twitter are rolling out fact checking systems (Fowler, 2020) and Google is promoting its Be Internet Awesome media literacy program (Seale and Schoenberger, 2018). However, others view these efforts with warranted suspicion. boyd (2017) argues that “thorny problems of fake news and the spread of conspiracy theories have, in part, origins in efforts to educate people against misinformation” because media literacy campaigns naturally ask the savvy information consumer to question the narratives that are presented to them. Inoculation-style media literacy projects like this one have much promise, but there is limited research on the effect of implementation (Fard and Lingeswaran, 2020). Studies looking at inoculation campaigns against climate change show that it is helpful to mention scientific consensus, along with reminding audiences that companies or political groups have economic interests in deceiving the public (van der Linden, et al., 2017). A study by Craft, et al. (2017) showed that higher levels of news literacy predicted a lower endorsement of conspiracy theories. In contrast, Jones-Jang, et al. (2019) found that only information literacy, which was measured by a skills test, predicted a user’s ability to recognize fake news stories; whereas a user’s media, news, and digital literacy, which are self-reported metrics, were not predictive of a user’s ability to recognize fake news stories (Jones-Jang, et al., 2021).

In this article, we offer preliminary results from our pilot study in combating the growing phenomenon of online misinformation via a social media literacy education campaign strategy that seeks to empower new digital arrivals in the Republic of Indonesia with increased ability to identify misinformation. After an assessment of the media landscape in Indonesia, we created six animated videos and two live-action videos demonstrating various lessons in online social media literacy. These lessons were presented as advertisements on YouTube, Google, Facebook, and Twitter and backed by http://literasimediasosial.id (which has been recently rebranded to http://literata.id).

We ask the following research question: Does online media literacy content lead to an increased ability to identify misinformation? In other words, can we teach people to do their own fact checking, and do such interventions lead to a scalable treatment against the global problem of misinformation? We measure the effectiveness of our campaign by asking visitors to our website via a phone survey to rate the accuracy of true, misleading, and false headlines. The results from this pilot study suggest that there is a modest increase in the ability to determine whether a story is real or not after engaging with our media literacy lessons.
Why Indonesia?

Indonesia is a large, diverse, and young democracy with a rapidly growing Internet user base and social media penetration. As of 2019, approximately 68 percent of Indonesia’s 270 million citizens were online, a figure that represents a dramatic increase from the approximately 43 percent of the population just four years prior in 2015 (Statista, 2020). By 2025 an estimated 89 percent of Indonesians will be online (Statista, 2020). As of 2019, 88 percent, 84 percent, 82 percent, and 79 percent of Indonesian Internet users self-reported use of YouTube, WhatsApp, Facebook, and Instagram, respectively (Statista, 2020). Because of this, many Indonesian citizens are new to the Internet. We position our work in this country to measure the effectiveness of this media literacy approach to a population which has had less exposure to the Internet, as compared to Western countries. In many countries, increased Internet penetration is correlated with increased violence (Oliveira, 2021), so finding a solution to this problem is important. This work provides important insights into the problems of online misinformation and propaganda, specifically within southeast Asia, which remains an understudied region. This particular context allows us to perform our research in a geographic area which is both in the Muslim world and in a country that is subjected to the Chinese sphere of influence. While similar previous work, such as Harmony Square and Bad News Game, was done targeting a Western audience, this study specifically targets another cultural context.

Polarization in Indonesia tends to exist between pluralists and Islamists, which is a divide that has a lengthy and complex history (Aspinall and Mietzner, 2019). Many political events are frequently met with substantial coordinated misinformation campaigns (known as “hoaxes” in the Indonesian context). For example, in the 2018 Jakarta mayoral election, recent reports have indicated that the election campaigns paid as much as US$280 per month to individuals who would promote messages from a particular candidate on social media (Lamb, 2018). Additionally, digital misinformation played a significant and highly divisive role in the national election in April of 2019 (Thiesen, et al., 2021). Protests against the results of the election resulted in six deaths and the temporary suspension of access to social media platforms by the Indonesian government (BBC, 2019).

As more Indonesians gain reliable access to the Internet and begin to use online social media platforms, it is possible that those who are new to the Internet have not yet fully developed the media literacy skills needed to distinguish between trustworthy and false news sources, and may therefore be vulnerable to manipulation through misinformation. Media literacy has been identified as something that is desperately needed to combat misinformation in the Indonesian context (Angeline, et al., 2020). Because Indonesia is a relatively young democracy, traditional democratic institutions like the press may not be as robust as those institutions in more established democracies (Bennett and Livingston, 2018). Furthermore, the susceptibility of the voting population to misinformation campaigns may also pose a threat to the stability of the democratic institutions of the nation itself.

Assessing the online and social media landscape in Indonesia

This initial work focused on identifying legitimate news stories, propaganda, and disinformation, and the popular narratives and hashtags that were used in these stories and across these domains. This methodology uses a combination of desk research, consultation and workshops with subject-matter experts, open-source intelligence methods, and computer systems to identify words, phrases, tropes, slogans, memes, slang, and other indicators of engagement, in multiple languages and across search, social media, image boards, forums, apps, and other discursive online spaces as necessary. Exact sources vary depending on the subject matter and for this deployment in February 2019 we focused on words and phrases indicative of intent to engage with disinformation on Google Search, YouTube, Facebook, and Twitter.

From this methodology, we were able to develop a clear understanding of how individuals — once they become preoccupied with certain patterns of ideas — continue to engage with misinformation online.
Pilot study suggests online media literacy programming reduces belief in false news in Indonesia

Depending on the exact methods of collection, this can be broken down by time, platform, aggregate user location, age and gender, subcategories of harm and different levels of risk. This understanding is then used to inform the creation of campaigns designed to reduce the impact of that harm, as was the case for this project.

Analysis of the resulting dataset revealed four common themes of disinformation specific to the Indonesian context: (1) anti-Chinese, (2) anti-Communist, (3) Islamic chauvinism, and (4) political smears. This assessment of the social media and search engine landscape in Indonesia was used to inform the specific content used in the social media literacy campaign, based upon the common themes of disinformation that were identified.

Methods

Social media literacy intervention

We developed a preliminary social media literacy education Web site (http://literasimediapasial.id) consisting of six informative lessons with several short educational videos and context-specific slogans that encourage social media behavior characterized by an increased ability to identify misinformation for widespread online delivery. This Learn To Discern (L2D) (Murrock, et al., 2017) approach, developed by IREX, builds communities’ resilience to state-sponsored misinformation, inoculates communities against public health misinformation, promotes inclusive communities by empowering its members to recognize and reject divisive narratives and hate speech, improves young people’s ability to navigate increasingly polluted online spaces, and enables leaders to shape decisions based on facts and quality information (Vogt, 2021). L2D is a purely demand-driven approach to media literacy, encouraging participants to increase self-awareness of their own media environments and the forces or factors that affect the news and information that they consume. L2D is traditionally presented as a training program and it typically includes a package of in-person activities, online games, distance-learning courses, public service announcements, and other methods that are tailored to the needs of social media users. Examples of L2D lessons are shown in Figure 1. For this campaign several short explainer videos were created. The full-length videos are hosted on YouTube (https://www.youtube.com/channel/UCU7jNIA-4gH3cxsFV6_xkjw) and direct users to http://literasimediapasial.id. Shortened versions of these animated and live action videos were also created to capture various aspects of the media literacy curriculum in short commercial-length snippets. At the conclusion of this preliminary study, this site was rebranded to http://literata.id and additional studies are ongoing.

Delivering media literacy content to social media users

Whereas previous media literacy efforts have been shown to be effective when conducted via in-person classroom settings (Hobbs and Frost, 2003), our goal was to use the digitized lessons and explainer videos to directly reach users — not just students — within the social media platform itself. The Redirect Method (Helmus and Klein, 2018) is a methodology created by Moonshot and Jigsaw in 2016, originally to fight against violent extremism. It is now used by Moonshot to counter a range of online harms by reaching individuals who perform searches which indicate intent to engage in, or that they are affected by, harmful content on social media or on search engines, with positive, alternative content.

Just as other companies use advertisements on social media and search engines to sell material products to an audience defined by, in whole or in part, the keywords they are searching for (for instance, a vacuum cleaner company might bid for advertising space from users who are searching for cleaning products), the Redirect Method places ads in the search results and social media feeds of users who are searching for pre-identified terms that we have associated with disinformation.

The Redirect Method can be extensively tailored to platform requirements and campaign goals, but at its core are three fundamental components: the indicators of risk (e.g., keywords); the advertisements triggered by the indicators; and the content to which users are redirected by the advertisements.
Pilot study suggests online media literacy programming reduces belief in false news in Indonesia

The indicators of risk

We curated an extensive gazette of keywords and phrases indicative of a desire to engage with harmful content — in this case, disinformation. The keywords for this campaign were created through a combination of desk research, consultation, and workshops with independent and local subject-matter experts and fact-checking organizations (some of whom shared their own keywords), and continuous mining for new keywords throughout the course of the project. The database currently contains more than 1,000 individual keyword phrases in Bahasa Indonesia, the main language of the country.

Of the keywords identified for the campaign as indicative of intent to engage with disinformation, the three most-searched for were as follows:

1. “jokowi pki,” which translates to, “Jokowi [President Joko Widodo] is a member of PKI [the communist party of Indonesia].”
2. “PDIP adalah topeng pki,” which translates to, “PDIP [the Indonesian Democratic Party of Struggle and the party to which the President belongs] is a mask of the PKI.”
3. “9 naga pendukung jokowi,” which translates to “the nine dragons support Jokowi.” The nine dragons refers to an age-old conspiracy that the capital city of Indonesia and its politicians are controlled by an underworld of nine Chinese mafia bosses. It is a myth based on the racist trope of “ethnic-Chinese control,” used to instill fear and distrust.

The database used for the campaign contained keywords indicative of support for all sides of the political spectrum in Indonesia, from the liberal-leaning base of Joko Widodo through to the more nationalist policies of Prabowo Subianto. It is important to note that the Indonesian political spectrum is best understood as a gradient between liberal and nationalist-authoritarian, rather than between left-wing and right-wing, as there is no political bloc analogous to the Western ‘left-wing’ archetype. Likewise, the demonization of ‘communist’ Chinese people living in Indonesia has historically come from across the Indonesian political spectrum. In this context, our ‘Anti-Communist’ keyword category is inherently apolitical.

The advertisements

The fundamental purpose of these advertisements is the same as those used in the commercial sector, i.e., to entice people to click on them. The difference in the case of this deployment of the Redirect Method is that social gain takes the place of commercial gain, with users who were initially seeking disinformation being taken instead to content designed to improve their media literacy.

In the end, a post-experimental questionnaire was administered via Google Forms, which dealt with aspects of participant preferences and opinions about the two user interfaces. At the end of the experiment, participants were thanked for their participation.
Using Google Ads to display advertisements, e.g., Figure 2 (A), in response to disinformation-related keyword searches means we can gather data on the number of times our advertisements were triggered (‘impressions’), thereby gaining an insight into the volume of searches for disinformation. When plotted by time, as in Figure 2 (B), we can also check for correlations in the search data with relevant and notable off-line events.

**The content**

Our social media literacy and search engine campaign was launched in August 2019 on Facebook, Twitter, Instagram, Google, and YouTube and concluded in April 2020. In total, the campaign generated 3.4 million impressions approximately equally distributed over seven different video lessons. Those impressions resulted in 72,976 unique page views of the campaign Web site. Because of the wide reach of the advertisement campaign, we cannot be certain that individuals in our control group did not see our content. However, because 72,976 unique user sessions page views represents approximately 0.03 percent of Indonesia’s estimated 183
A pilot study suggests online media literacy programming reduces belief in false news in Indonesia. With over 100 million online users, we feel fairly confident that most of the control group users did not see our media literacy content, though we do not rule out this possibility. A sample of the media literacy lessons is illustrated in Figure 2.

![Figure 2:](A): An English-language version of one of the advertisements used in this deployment of the Redirect Method, visible only to users who were otherwise searching for disinformation (actual ad text was in Bahasa Indonesia); (B): Ad impressions on Google Search for disinformation content. Disinformation-related searches in Indonesia peaked during significant political and social events, such as the 2019 student riots, and the beginning of the COVID-19 pandemic.

<table>
<thead>
<tr>
<th>Source</th>
<th>Clicks</th>
<th>Bounce rate</th>
<th>Session duration (s)</th>
<th>CTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>11,878</td>
<td>96.61%</td>
<td>6.00</td>
<td>2.08%</td>
</tr>
<tr>
<td>Instagram</td>
<td>1,312</td>
<td>89.46%</td>
<td>28.04</td>
<td>0.54%</td>
</tr>
<tr>
<td>Twitter</td>
<td>609</td>
<td>85.26%</td>
<td>48.63</td>
<td>0.45%</td>
</tr>
<tr>
<td>Google</td>
<td>104,952</td>
<td>25.91%</td>
<td>29.00</td>
<td>4.79%</td>
</tr>
</tbody>
</table>

Overall, we received 3,444,398.2 impressions (as calculated by each platform). Of those, 72,976 users clicked through to the Web site. In Table 1, we show how long each user spent on our site based upon the platform that the user was coming from. Additional campaign statistics are described in the Table A1 in the Appendix.

**Phone surveys**

After the conclusion of our social media literacy campaign, a team of interviewers based in Jakarta, Indonesia, conducted a computer-assisted telephone interview (CATI) for 1,000 successful interviews. The surveys were
Pilot study suggests online media literacy programming reduces belief in false news in Indonesia
nationally representative proportionate to 2019 census data estimates at the province-level of location, age
group, and gender. The main interview language of the survey was Bahasa Indonesia, but the interviewer team
was able to switch to other languages such as Balinese and Javanese if requested by the interviewee.

The research protocol and questionnaire were approved by the internal ethics committee at the University of
Notre Dame (#18-11-5009). Data is anonymized, but demographic information is included. Data and full cross
tabulations can be found at https://www.geopoll.com/misinformation-indonesia/.

Visitors to the media literacy education Web site were asked to be included in a phone survey in exchange for
approximately US$0.50 equivalent in local currency in phone credit. Altogether, 331 individuals agreed and
provided their phone number, and of those, 94 completed the CATI survey. This constitutes the treatment
group. The control group was composed of 996 respondents to verified random digit dialing phone surveys.

We acknowledge that the treatment group is likely to have self-selection bias. Our control group was selected
through random digit dialing, but the treatment group were redirected to the website because of their
propensity to search for misinformation subjects. Therefore, we might expect to see that our treatment group is
more Internet savvy because of how they were reached (via Internet instead of random digit dialing), or more
likely to believe in misinformation because they were actively seeking out false stories or searching for known
misinformation topics online. However, we believe that as a pilot study, this data provides important
information about the efficacy of media literacy campaigns which invites further study. We asked participants
about general demographic information, inquired about their media consumption habits, and also queried their
ability to correctly judge the veracity of news headlines that we presented to them. The treatment group was
both older and skewed more male as compared to the control, as shown in Tables A2 and A3 in the Appendix,
which might introduce more bias.

During data collection the survey team made approximately 4,500 phone calls that resulted in a 21 percent
response rate. The margin of error for the survey at the 95% confidence level is ±3.10 percent.

![Graph showing Real, Misleading, and Misinformation judgments for Treatment and Control groups]
**Figure 3**: Overall, 58.3 percent of our treatment group was able to identify the misinformation as either very or somewhat inaccurate, compared with 42.4 percent of the control group. We supply the number of respondents that gave an accuracy rating for each headline. For this plot, we averaged together the number of individuals who replied to our real headlines, since each person was given two real headlines to rate.

---

**Findings**

To judge the effectiveness of the social media literacy campaigns, each survey participant was asked if they recognized two true, one misleading, and one false headline. If so, we asked them to rate the accuracy of that headline on a Likert scale from very accurate to very inaccurate.

Our goal was to examine the differences between how the control and treatment groups rated the accuracy of headlines. An improvement in the treatment group’s ability to correctly identify false, misleading, and real stories when compared to the control group would demonstrate that our campaign had an effect on a user’s ability to analyze the news headlines they encountered.

<table>
<thead>
<tr>
<th>Headline</th>
<th>Group</th>
<th>N</th>
<th>Mean rank</th>
<th>p</th>
<th>Significance</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>Control</td>
<td>1,113</td>
<td>3.36</td>
<td>0.470</td>
<td>0.871</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>120</td>
<td>3.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misleading</td>
<td>Control</td>
<td>321</td>
<td>3.00</td>
<td>0.497</td>
<td>0.520</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>48</td>
<td>3.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misinformation</td>
<td>Control</td>
<td>273</td>
<td>2.71</td>
<td>0.569</td>
<td>0.081</td>
<td>Significant at 90% confidence</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>36</td>
<td>2.44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Finding 1**: Users who encountered the media literacy content were more likely to identify false news than those that did not.

Our preliminary findings from this pilot study suggest that our media literacy intervention was positively correlated to a respondent’s ability to identify misinformation. Treatment group users in our pilot study were 15.9 percent more likely to identify a misinformation headline as either very or somewhat inaccurate. These results are significant at 90 percent confidence (Mann-Whitney ρ = 0.569, N = 309, p = 0.081 one-tailed). **Figure 3** shows the results of the Likert-scale responses and **Table 2** shows the statistical breakdown of the responses. Because the treatment group consisted of individuals who were actively searching for misinformation, these results are even more encouraging because they suggest that our campaign had a measurable effect even on those who are familiar with and are actively searching for misinformation content. We plan to expand upon this pilot study’s encouraging results in our future research, as we acknowledge the
limitations of such a small sample size. Across both the control and treatment groups, accuracy ratings were very similar for the different headline types.

Interestingly, as shown in Figure 3, it appears that relatively few individuals in the treatment group had a neutral opinion of the misinformation story, as compared to the other groups and headline types. This suggests that our media literacy campaign might have persuaded those who were not sure about the validity of the story that the narrative was false, via the new skills they acquired from the website. This is supported by the observation that the truthfulness rating of the misinformation headlines was similar between the treatment and control groups. More research is needed to understand the effect of media literacy campaigns on these fence-sitters — those who have neutral feelings about the headlines that they see, and are unable to tell if these stories are true or false.

In future studies, we plan to study the effects of educating citizens about misinformation on their skepticism of media content more generally. Some research has shown that discussion about misinformation for “fake news” by elite sources results in less ability to identify real news stories (Van Duyn and Collier, 2019). The question becomes how to prevent media literacy education from becoming “media nihilism” (Van Duyn and Collier, 2019), lest a decreasing trust in all journalism actually lead to higher levels of acceptance from other illegitimate, polarizing sources (Egelhofer and Lecheler, 2019).

In summary, we found that our social media literacy intervention is associated with an increase in the ability of users to correctly identify misinformation. However, despite these findings, participants in our social media literacy campaign did not show an increased ability to accurately identify either misleading or real news stories. This phenomenon presents an intriguing area for future research.

### Table 3: Awareness of news stories is positively correlated with its veracity $X^2 (2, N = 3,610) = 327.60 p < 0.001$.

<table>
<thead>
<tr>
<th>Heard of story?</th>
<th>Real ($N = 1,865$)</th>
<th>Misleading ($N = 875$)</th>
<th>Misinformation ($N = 870$)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>1,332 (71.4%)</td>
<td>401 (45.8%)</td>
<td>333 (38.2%)</td>
<td>$N = 2,066$</td>
</tr>
<tr>
<td>No</td>
<td>533 (28.5%)</td>
<td>474 (54.1%)</td>
<td>537 (61.7%)</td>
<td>$N = 1,544$</td>
</tr>
</tbody>
</table>

**Finding 2:** False news did not spread as broadly in Indonesia as real news.

Interestingly and rather unexpectedly, we found that the headline accuracy was positively correlated with its reach. These results, shown in Table 2, suggest that misinformation does not spread quite as easily as real news stories in Indonesia — however, further research is needed to understand this phenomenon. In future work, we plan to expand the number of headlines that we supply to our survey respondents.

Our results suggest that our media literacy campaign had an effect on the ability of individuals to accurately label hoaxes as misinformation. However, further research is needed to answer some of the questions that our findings raise. For instance, we see that large numbers of respondents have neutral feelings about news stories. While our campaign focused on identifying misinformation, perhaps future work can promote trust in legitimate news sources, which would create statistically significant differences in the ability of media literacy campaign participants to identify real headlines as such, not simply to identify misinformation.
Conclusions

In this work, we present findings from our early pilot study on the efficacy of a media literacy campaign in Indonesia in helping users identify misinformation headlines. These results indicate that 58.3 percent of users who visited our media literacy Web site and engaged with the education content were able to correctly identify misinformation headlines as inaccurate, compared to our 42.4 percent of the control group. These results suggest that it is possible to use a media literacy approach to help citizens learn to identify misinformation.

This media literacy approach presents an alternative to other proposed solutions to the misinformation problem such as censorship. Questions about the constitutional right to free speech on various platforms remain very complex, particularly amidst varied legal and normative frameworks in different contexts. There are even more questions concerning the implications for social media content. As social media companies grapple with how to deal with the ramifications of their algorithms on society, it is important to understand the effects of censorship on other platforms and in other countries (Thomas, Saldanha, et al., 2021), so that policy-level decisions can be developed which strike the right balance when answering these questions. Some social media companies, like Reddit, have historically had a laissez-faire attitude towards removing hate speech, but they have begun to censor as well. The effect of Reddit’s censorship has also been studied (Thomas, Riehm, et al., 2021). While governments and social media companies have trended towards censoring content, this work proposes an alternative solution.

This work presented a small, limited study to understand the misinformation landscape in Indonesia, and demonstrates the feasibility of conducting a larger research project of this nature. Since our results were positive, we plan on expanding this study in several ways. First, we will introduce a gamification approach - analogous to Harmony Square and Bad News game — into our media literacy education materials. Second, we will enlarge the sample size and test the significance of the results on larger groups. Third, we will track information such as how long a user stayed on the site and engaged with the material. Fourth, we will ask more questions about Internet usage, such as how long a user has been using the Internet. Fifth, we will ask users about more disinformation headlines, since currently we only ask them about the veracity of one headline. With these additional considerations, we will create and deploy a broader study which will give us even more insights into the efficacy of media literacy campaigns.

We present these preliminary results to others that might be considering larger, more comprehensive studies in this area. Researchers have found that inoculation campaigns appear to be more successful than corrections, but few large-scale studies have been done (Levy, et al., 2021). We hope that our work will help those who are designing such studies, and implement the lessons learned from our study to further study media literacy, especially in developing countries or for new digital arrivals.

About the authors

Pamela Bilo Thomas is an assistant professor in the Speed School of Engineering at the University of Louisville, Kentucky. She received her Ph.D. from the University of Notre Dame in 2021. She received her B.A. in mathematics and political science at Indiana University Bloomington in 2011, and her Master’s in computer science, also from Indiana University Bloomington, in 2013. Her research interests include machine learning and data science.
E-mail: pamela [dot] thomas [dot] 1 [at] louisville [dot] edu

Clark Hogan-Taylor is a Manager at Moonshot, specializing in methodological solutions to online harms with a focus on southeast Asia. He has led projects across the region both online and on the ground, covering ultranationalism and gender-based violence, disinformation, civil society capacity-building, and counter-narrative campaigns.
E-mail: clark [dot] hogantaylor [at] mshot [dot] com
Pilot study suggests online media literacy programming reduces belief in false news in Indonesia

**Michael Yankoski** is a postdoctoral research associate at the University of Notre Dame. He is a scholar of ethics and peace studies, and earned his Ph.D. from the Kroc Institute for International Peace Studies at the University of Notre Dame.  
E-mail: myankosk [at] nd [dot] edu

**Tim Weninger** is the Frank M. Friemann Collegiate Associate Professor of Engineering in the Department of Computer Science and Engineering at the University of Notre Dame. His current research interests include the intersection of social media, data mining, and network science in which he studies how humans create and consume networks of information.  
E-mail: tweninger [at] nd [dot] edu

**Acknowledgements**

We would like to acknowledge and thank Carolina Rocha da Silva, Rachel Fielden, Joel Turner, Tavian MacKinnon, and Walter Scheirer, Joshua Macleder, and Anders Mantius for their assistance on this project. The funding for this research study was supported by USAID Cooperative Agreement #7200AA18CA00059.

**References**


doi: [https://doi.org/10.1109/ICIMTech50083.2020.9211124](https://doi.org/10.1109/ICIMTech50083.2020.9211124), accessed 5 January 2022.


Pilot study suggests online media literacy programming reduces belief in false news in Indonesia


Andrew M. Guess, Dominique Lockett, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, and Jason Reifler, 2020. “‘Fake news’ may have limited effects beyond increasing beliefs in false claims,” *Harvard Kennedy School Misinformation Review* (14 January).


Pilot study suggests online media literacy programming reduces belief in false news in Indonesia


Pamela Bilo Thomas, Daniel Richm, Maria Glenski, and Tim Weninger, 2021. “Behavior change in response to Subreddit bans and external events,” IEEE Transactions on Computational Social Systems, volume 8,
Appendix

Table A1: Reach and viewership statistics of media literacy campaign.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Impressions</th>
<th>Clicks</th>
<th>Average duration (s)</th>
<th>Click through rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>252,511.0</td>
<td>6,877</td>
<td>—</td>
<td>4.0%</td>
</tr>
<tr>
<td>Facebook</td>
<td>568,452.3</td>
<td>22,747</td>
<td>14.17</td>
<td>4.0%</td>
</tr>
<tr>
<td>Instagram</td>
<td>525,292.8</td>
<td>2,315</td>
<td>27.54</td>
<td>0.4%</td>
</tr>
<tr>
<td>Twitter</td>
<td>294,793.4</td>
<td>2,265</td>
<td>35.20</td>
<td>0.8%</td>
</tr>
<tr>
<td>YouTube</td>
<td>1,803,348.8</td>
<td>38,772</td>
<td>0.62</td>
<td>2.2%</td>
</tr>
</tbody>
</table>

Table A2: Treatment group is older than our control group $X^2 (5, N = 1,000) = 29.66 p < 0.001$.

<table>
<thead>
<tr>
<th>Age</th>
<th>Treatment ($n = 94$)</th>
<th>Control ($n = 906$)</th>
<th>Total ($n = 1,000$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15–24</td>
<td>13 (13.8%)</td>
<td>201 (22.1%)</td>
<td>N = 214</td>
</tr>
</tbody>
</table>
Pilot study suggests online media literacy programming reduces belief in false news in Indonesia

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Treatment</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>25–35</td>
<td>23 (24.4%)</td>
<td>305 (33.6%)</td>
<td>N = 328</td>
</tr>
<tr>
<td>36–45</td>
<td>18 (19.1%)</td>
<td>216 (23.8%)</td>
<td>N = 234</td>
</tr>
<tr>
<td>46–55</td>
<td>21 (22.3%)</td>
<td>123 (13.5%)</td>
<td>N = 144</td>
</tr>
<tr>
<td>56–65</td>
<td>15 (15.9%)</td>
<td>48 (5.2%)</td>
<td>N = 63</td>
</tr>
<tr>
<td>66+</td>
<td>4 (4.2%)</td>
<td>13 (1.4%)</td>
<td>N = 17</td>
</tr>
</tbody>
</table>

Table A3: Treatment group is more male than our control group $X^2(1, N = 1,000) = 29.35 p < 0.001$.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Treatment</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>72 (76.5%)</td>
<td>428 (47.2%)</td>
<td>N = 500</td>
</tr>
<tr>
<td>Female</td>
<td>22 (23.4%)</td>
<td>478 (52.7%)</td>
<td>N = 500</td>
</tr>
</tbody>
</table>

Table A4: Our treatment group is similar to our control group based upon urban/rural identity $X^2(1, N = 994) = 0.727 p = 0.393$.

<table>
<thead>
<tr>
<th>Urban/Rural?</th>
<th>Treatment</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>46 (49.4%)</td>
<td>404 (44.8%)</td>
<td>N = 450</td>
</tr>
<tr>
<td>Rural</td>
<td>47 (50.5%)</td>
<td>497 (55.1%)</td>
<td>N = 544</td>
</tr>
</tbody>
</table>

Table A5: Our treatment group is similar to our control group based upon religious identity $X^2(3, N = 990) = 2.09 p = 0.553$.

<table>
<thead>
<tr>
<th>Religion</th>
<th>Treatment</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muslim</td>
<td>87 (95.6%)</td>
<td>827 (91.1%)</td>
<td>N = 914</td>
</tr>
<tr>
<td>Christian</td>
<td>4 (4.3%)</td>
<td>58 (6.4%)</td>
<td>N = 62</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0.0%)</td>
<td>13 (1.4%)</td>
<td>N = 13</td>
</tr>
<tr>
<td>None</td>
<td>0 (0.0%)</td>
<td>1 (0.1%)</td>
<td>N = 1</td>
</tr>
</tbody>
</table>
Pilot study suggests online media literacy programming reduces belief in false news in Indonesia

by Pamela Bilo Thomas, Clark Hogan-Taylor, Michael Yankoski, and Tim Weninger.

First Monday, volume 27, number 1 (January 2022).

doi: https://dx.doi.org/10.5210/fm.v27i1.11683