

# "Reenviado Muchas Veces": How Platform Warnings Affect WhatsApp Users in Mexico and Colombia

Kevin Munger, Angel Villegas-Cruz, Jorge Gallego & Mateo Vásquez-Cortés

To cite this article: Kevin Munger, Angel Villegas-Cruz, Jorge Gallego & Mateo Vásquez-Cortés (14 Mar 2024): "Reenviado Muchas Veces": How Platform Warnings Affect WhatsApp Users in Mexico and Colombia, Political Communication, DOI: [10.1080/10584609.2024.2326130](https://doi.org/10.1080/10584609.2024.2326130)

To link to this article: <https://doi.org/10.1080/10584609.2024.2326130>



Copyright © 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 14 Mar 2024.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)



This article has been awarded the Centre for Open Science 'Open Data' badge.



OPEN ACCESS



# "Reenviado Muchas Veces": How Platform Warnings Affect WhatsApp Users in Mexico and Colombia

Kevin Munger<sup>a</sup>, Angel Villegas-Cruz<sup>a</sup>, Jorge Gallego<sup>b</sup>, and Mateo Vásquez-Cortés<sup>c</sup>

<sup>a</sup>Department of Political Science, Pond Laboratory, State College, Penn State University, PA, USA; <sup>b</sup>Department of Economics, Universidad del Rosario, Bogota, Colombia; <sup>c</sup>Department of Political Science, University of California, San Diego, USA

## ABSTRACT

Digital literacy affects how people use the internet. However, we argue that the concept of "digital literacy" cannot usefully be applied to all internet users; there is simply too much heterogeneity across devices, platforms, and social contexts. We conduct surveys in Mexico and Colombia to understand how these people use the internet. We find that WhatsApp is by far the dominant platform across all sectors of society. In contrast to evidence from the US, we find that education is a better predictor of two measures of digital literacy than is age. We then conduct a survey experiment to test how people understand news shared on WhatsApp. We find that the inclusion of a link to source material increases respondents' trust in a message shared on WhatsApp, but that the platform-supplied note that a message has been shared many times has no such effect.



## Keywords


WhatsApp; digital Literacy; Colombia; Mexico

## Digital Literacies and Platform Affordances

Like previous innovations in media technology, the internet and social media have enabled an explosion in the production of media and thus the amount of content that average citizens can consume. In both cases, however, there were and are significant inequalities in the *capacities* of citizens to understand and take advantage of this media. Addressing the current "secondlevel digital divide" (Hargittai, 2001) requires the cultivation of digital literacy. Although scholars of digital literacy are generally nuanced in their treatment of the topic (Gilster, 1997; Koltay, 2011), the extent of the heterogeneity and complexity of the digital media environment presents a novel challenge for both citizens and how researchers understand their experiences.

Research on digital literacy has seen renewed interest in the wake of concern about the spread of misinformation on social media in the past decade. In particular, a series of large-scale analyses in the US demonstrated overwhelming relationship between age and the propensity to view and share misinformation (Barbera, 2018; Grinberg et al., 2019; A. Guess et al., 2019a). Age itself is unsatisfying as a theoretical mechanism; however, age is correlated with a host of other individual characteristics, including digital literacy (A. M. Guess & Munger, 2023), which *does* have implications for how people evaluate news online.

**CONTACT** Kevin Munger  [kmm7999@psu.edu](mailto:kmm7999@psu.edu)  Department of Political Science, Pond Laboratory, State College, Penn State University, PA 16801, USA

 Supplemental data for this article can be accessed on the publisher's website at <https://doi.org/10.1080/10584609.2024.2326130>.

Copyright © 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

A. M. Guess and Munger (2023) define digital literacy as consisting of a *skills* component and an *information literacy* component. At this level of abstraction, the definition is plausible, but the specific way that scholars have rationalized these two components cannot be trivially transported out of the context in which they were developed. The paradigmatic citizen in this context is located in Europe or the US; uses a personal computer (desktop or laptop); has no limits on data availability; and encounters information in a platform-driven social media context.

Contrast this with the modal internet user in Latin America who uses a mobile device; with a pre-paid phone plan with data caps; and who encounters information in a WhatsAppstyle peer-to-peer messaging context. Although cross-national differences dominate, this mode of internet use is far more common worldwide than is the European/US context that serves as an implicit baseline internet experience in both the minds of many researchers and the research itself.

WhatsApp in particular has received insufficient attention compared to platforms like Facebook and Twitter. The app ranks third in worldwide Monthly Active Users, behind Facebook and YouTube, yet there are fewer than 10% as many Google Scholar results for WhatsApp than for Facebook, YouTube or Twitter. In addition to regional and linguistic bias – as well as the non-trivial barrier to collecting data from WhatsApp relative to platforms with more open APIs – the fact that WhatsApp is “merely” a messaging app has accelerated its naturalization: the process by which technology ceases to be recognized as technology and instead becomes an invisible part of daily life (Cruz & Harindranath, 2020).

This ubiquity can sometimes overshadow the specific affordances of WhatsApp: the limits the platform has imposed on the sharing of messages across groups, the limits on the number of pictures that can be shared, and most centrally, the warning the platform provides when a message has been re-shared many times.

We therefore investigate two dimensions on which the “banal” reality (to use Treré (2020)’s evocative phrase) of WhatsApp use for political communication in Mexico and Colombia differs from the use of platform-based social media in the US and Europe.

First, at the individual level, we investigate the relationships between demographic characteristics like age and education, the use of the internet and social media, and a variety of measures of digital literacy and platform-specific knowledge. Our aim here is to verify the extent to which well-known relationships from the US context (like the negative relationship between age and digital literacy) apply in Mexico and Colombia. This is essential for how scholars can adapt findings from one context to another, how they should change (or not) their theoretical expectations. Based on existing literature for contexts other than the US, we propose, as our first testable hypothesis, that the correlation between digital literacy and education is stronger than that between digital literacy and age in Colombia and Mexico.

Second, with respect to the specifics of the affordances of WhatsApp, we conduct an experiment to analyze the impact of two kinds of cues on how citizens evaluate whether a news story shared on WhatsApp is credible, and whether they are willing to read it and/or share it. The first experimental manipulation is an example of a user-specific cue: the inclusion of a link to the quoted news story. Again, based on the existing literature, our second testable hypothesis states that the inclusion of a source has a positive effect on the probability that a WhatsApp user reads, shares and believes in news received through this platform.

The second manipulation is an example of a platform-specific cue: the addition of the phrase “Reenviado Muchas Veces” and its characteristic “double arrow,” which signals that a message has been forwarded five or more times, instead of the message that it has simply been forwarded (“Reenviado”). Our study thereby applies and extends the theory developed in Tandoc et al. (2022)—the most related study of which we are aware – from Singapore to Mexico and Colombia, countries with similar levels of WhatsApp use but very different media environments and levels of digital literacy. The existing evidence here is less conclusive and we are therefore a bit more agnostic regarding this cue, and hypothesize that the tag only has an effect on the probability that users find a news story credible.

To corroborate our hypotheses, we conduct an original survey with a sample of over 3,000 internet users in both Colombia and Mexico. We collaborated with Offerwise, an established consulting company with over 16 years of experience in online market research in Latin America. Offerwise has one of the largest and most representative panels of respondents in Latin America and among U.S. Hispanics, and while the sampling frame does not allow for making population-inferences about base rates of internet use or digital literacy, the correlations between these characteristics and the results of our experiments are still useful for answering our central research question.

One major finding in the US context is that older people are on average much more likely to be exposed to misinformation than are young people (Barbera, 2018; Grinberg et al., 2019; A. Guess et al., 2019b) and that this tendency can be in part explained by digital literacy (A. M. Guess & Munger, 2020). This could be primarily because of age per se or primarily because of cohort- or network-based effects. Data from a single country are insufficient to disentangle these mechanisms, motivating one dimension of the present analysis. As stated above, we hypothesize that the latter effects dominate, and that the digital literacy gradient in Latin America will be driven more by education than by age. This is what we find: across two scales developed to measure digital literacy in the US, more of the variance is explained by educational attainment than by age, in both Mexico and Colombia and in contrast to the US.

Moreover, we find that the relationship between age and internet skills is non-monotonic, with middle-aged respondents scoring higher than either the young or the old. In contrast, according to our new scale of WhatsApp knowledge, young respondents score highest. And in both Colombia and Mexico, the relationship between education and WhatsApp knowledge is non-monotonic: respondents with either the lowest or highest level of educational attainment evince less WhatsApp knowledge than respondents in the middle of the range.

Turning to our experimental results, we find consistent evidence across both countries and multiple outcome measures that the addition of the link to the original story increased the credibility of the message being shared, as well as their reported propensity to read and share the message. This result is consistent with our second testable hypothesis. However, we find little evidence that switching the WhatsApp cue from “Reenviado” to “Reenviado Muchas Veces” had any effect, in contrast to what we propose in our third hypothesis. The one exception is that in Mexico (but not Colombia), the addition of “Muchas Veces” caused respondents to be more likely to *share* the message, without any effect on the *credibility* of the news or their propensity to read it.

We also explore treatment effect heterogeneity along the various dimensions that have been found to moderate digital treatment effects in other contexts. In contrast to our expectations and our fourth testable hypothesis, we find little evidence of treatment effect heterogeneity; the one exception is along the dimension of age, although this effect appears only for some outcome measures.

In summary, our results provide important context for scholars studying digital media, platform affordances and media credibility across multiple contexts. Although digital literacy has been shown to be correlated with age than with education in the US, we find the opposite result in Mexico and Colombia; furthermore, this important moderating variable was not shown to have any effect of how citizens interpret the platform-specific cue of interest. But we also find that “Mexico and Colombia” cannot be considered a single context, with different experimental results in each country.

The question of how best to translate knowledge across contexts is a difficult one, but the solution is not to shy away from the problem, to retreat to the best-studied countries and platforms. Our study combines the examination of individual- and platform-level variables in the aim of balancing the process of theoretical generalization with careful empirical checks on the intuitions underlying that theory.

## WhatsApp and Political Communication

### *WhatsApp in Colombia and Mexico*

WhatsApp comprises much of the internet for many people in Latin America. On this dimension and in terms of human capital and the prevalence of mobile internet access, this region is far more representative of the rest of the world than are the developed countries that currently dominate research on online communication. According to the *Latinobarómetro* (2018), 64% of the population in Latin America used this app, with similar numbers for Colombia (67%) and Mexico (60%). Treré (2020) explains WhatsApp’s appeal in these contexts: “[I]ts simplicity, reliability and accessibility are some of its core selling points, especially for people in the Global South where online services are accessed mainly through cellphones and several connectivity and bandwidth issues are often thwarting a rapid and fluid digital experience.”

Even though Colombia and Mexico have similar characteristics in terms of education, internet use and WhatsApp popularity, these countries provide a contrast in terms of the political orientation of the government in power. Mexico is governed by a left-wing government led by Andres Manuel Lopez Obrador, who won the election in 2018 with an agenda contrary to the traditional parties of Mexico (PRI and PAN). His party, MORENA, describes itself as a democratic left-wing party with an agenda opposing the neoliberal economic policies that Mexico began adopting in the 1980s. In contrast, Ivan Duque won the Colombian elections in the same year of 2018 with the support of the Centro Democrático party, founded by former right-wing president Alvaro Uribe Velez. The party’s political agenda is based on security, private investment, and social cohesion.

We consider these two countries given the political orientation of their governments and the increasing relevance of WhatsApp in their elections. Political candidates and organizers rely on WhatsApp to reach voters, mobilize supporters, raise funds, and increase poll turnout. However, the spread of disinformation and polarization in elections is a significant concern in both countries.<sup>1</sup> For instance, prior to the 2018 parliamentary elections in Colombia, a fake image claiming that the Governor had suspended the elections in the port city of Tumaco was widely circulated on WhatsApp.<sup>2</sup>

In Colombia and Mexico, as in other countries in the Global South like Brazil, India, and Nigeria (Ceci, 2021), WhatsApp is one of the most important channels for disseminating

fake news (Cheeseman et al., 2020; Davies, 2020; Santini et al., 2021). Given the rapidly changing nature of political discourse on digital platforms, it is crucial to have a comprehensive understanding of WhatsApp's role in political matters beyond a single country.

### **Digital Literacy and (Mis)information on WhatsApp**

Political misinformation is troublingly common on WhatsApp, especially during elections (Garimella & Eckles, 2020; Machado et al., 2019). Chain messages or *cadena*s—communications in the form of text, audio or video forwarded massively across users – are frequently used for political, commercial, and sometimes fraudulent purposes (Cortes & Peñarredonda, 2018). The technological affordances of WhatsApp include the fact that messages are encrypted, which means that the company (and its parent corporation, Meta, which also operates Facebook) cannot deploy the same suite of content moderation tools that have been developed in the context of platform-centric social media. There is no algorithm governing the display of content on WhatsApp, evidence that misinformation can flourish even without the attention-based algorithmic feedback loops that have inspired concern in the platform context.

Still, the social dynamics that determine the flow of content on WhatsApp – what gets posted, shared, engaged with, and whether this happens in one-on-one chats, small friend group chats, large family group chats, or extremely large (up to the limit of 256 people) group chats organized around shared interests – are complex. The individual user must evaluate the text, audio recordings, videos, images and links that are shared with them. They must also decide what they themselves want to share, and with whom, and in what contexts.

When a person receives a message on WhatsApp, they have information about the person who sent it and how many times it has been forwarded; the recipient can also easily forward it to other groups of people within WhatsApp. A user can do this by selecting the link and then choosing the person within their WhatsApp contact list who will receive the message. However, and more relevant to our study, a user can also share information directly without opening the content of the message (in the case of an Internet link, for example). Baulch et al. (2020) provides a useful overview of research on WhatsApp and political communication.

There is some evidence that the restrictions on re-sharing decreased the speed of spread of misinformation in Brazil and India (Melo et al., 2019). However, the individual mechanisms that produce this macro-level tendency are not yet known; our goal is to better understand how users decide how to evaluate information they encounter on the app. Furthermore, Santos et al. (2020) shows how strong restrictions on the use of WhatsApp in Brazil – specifically, a brief shutdown of the app – spurred future skill acquisition. This points to the challenge of studying the effect of platform changes in the presence of response by users.

In research on digital literacy in the United States, one variable has been found to have the most significant explanatory power: age. This variable is of course far from deterministic, and scholars have been careful to point out that narratives about “digital natives” and “digital immigrants” can obscure more than they explain (Hargittai, 2010). Given the distinct pathways by which individuals in the Global South have come to use the internet, we expect the relationship to be even weaker there. Some of the difference in our



expectations across national contexts stems from the empirical distribution of age and education. The age pyramid in many advanced democracies is unusually top-heavy, in contrast to developing countries like Colombia, and especially Mexico, which have younger populations.

There are many mechanisms that can explain the correlation between age and digital literacy in the United States. At least some of the explanation, though, seems to be related to the evolution of cognitive plasticity throughout the life cycle, which is the kind of effect that we expect to be relatively constant across national contexts. In the United States, 17% of the population is at least 65 years of age, compared to just 9% in Colombia and 8% in Mexico. Pew reports that 75% of US citizens in this age range use the internet (Pew, 2021), compared to 53% of citizens living in Latin American.<sup>3</sup> Age also influences the rate of social media adoption, but compared to the Internet, the gap between the number of social media users in the US vs. Latin America seems to be wider. Pew reports that 45% of US citizens in this age range (65 years or more) use at least one social media app (Pew, 2021), compared to less than 3.4% in Mexico<sup>4</sup> and 4.1% in Colombia.<sup>5</sup>

So while we expect age to be important in explaining digital literacy in Colombia and Mexico, we expect it to be less relevant than in the United States. In contrast, we expect that human capital will be more relevant to how citizens of these nations interpret information on WhatsApp. There is significantly more variation in education level in these nations, a proxy for human capital. Furthermore, given that the content encountered on WhatsApp is even more driven by social networks than content on platform-based social media, we expect that the level of human capital of the social contexts in which Mexicans and Colombians are embedded to explain the kind of content they encounter and decide to share.

We operationalize digital literacy in three ways. First, we adapt two batteries developed in other contexts: the “internet skills” measure developed in Hargittai (2005); Hargittai and Patrick Hsieh (2012) and the “power user” scale developed by Sundar and Marathe (2010). The former consists of a battery of questions about the familiarity of the respondent with a variety of terms related to computers and the internet. We follow the selection of terms and translation used by Correa et al. (2022) in the Chilean context. Respondents were asked how familiar they were with the following: Advanced search, PDF, Spyware, Wiki, Bookmark, and JPG. The outcome is a five-point scale ranging “not at all” to “a great deal.”

The latter scale consists of a series of questions about the habits and confidence of technology users. This scale consists of twelve questions asking respondents how they interact with technology, on a 9-point agree-disagree scale. These questions are intentionally more vague, and include “I make good use of most of the features available in any technological device” and “Using information technology makes it easier to do my work.”

Finally, we developed a short measure of advanced WhatsApp-specific knowledge. There are important restrictions on how messages can be shared that we expect expert WhatsApp users to be aware of. We asked the following three questions: “How many photos can you send via WhatsApp at the same time?,” “To how many people can you forward content via WhatsApp?” and “How many times can content be forwarded on WhatsApp?”

It is possible that some respondents searched on the internet to find the correct answers. While this is itself an important measure of digital literacy, in the future, we would choose to measure these two aspects directly: to either instruct respondents not to look up the answer, to include a time limit, or to include a separate information-retrieval task.

## ***The Effects of Self- and System-Generated Cues on Credibility***

We explore two aspects of political communication on WhatsApp that have been developed and applied to other populations to understand trust in the information shared on the platform. First, we study the impact of content label cues such as the aforementioned forwarded tag (Santini et al., 2021; Tandoc et al., 2022). Second, we revise the impact of sources cues following growing efforts to understand and curb the spread of misinformation in private communication applications (Masip et al., 2021; Rossini, Anita Baptista, et al., 2021; Sumitra et al., 2022).

The two experimental treatments we deploy are based on two distinct types of cues which have been shown to affect credibility across a variety of online contexts: system-generated cues and self-generated cues (Shan, 2016).

The central system-generated cue we study is the message informing the user if an message has been “forwarded many times” (“*Reenviado muchas veces*” in Spanish) or only “forwarded.” The “*Reenviado muchas veces*” dimension is relevant in light of one of the main features of WhatsApp. As the messages between users are encrypted, the company cannot moderate its content, as it is done in other platforms such as Facebook or Twitter. Therefore, the application uses informational strategies of this type to alert users about possibly problematic content. There is anecdotal evidence that the label was interpreted by users in Mexico<sup>6</sup> and Colombia<sup>7</sup> as an attempt to stop the spread of fake news (especially images).

Tandoc et al. (2022) provide a deeper qualitative understanding of the way that WhatsApp users understand this system-generated cue: “Although the original intention is to signal receivers that they are receiving a message that had been shared by others, the forwarded tag, just like other system-generated cues online, may have intended and unintended uses and understandings (p1856).” On average, the Singaporean subjects studied by Tandoc et al. (2022) rated messages as lower in credibility when the “forwarded” message was present; however, there was substantial heterogeneity that emerged in the qualitative interviews. Of particular relevance to the current study is the interviewed that “referred to the difference between a single arrow and a double arrow. . .The participant said: ‘The two arrows means that it’s been forwarded more times, so it’s more reliable’ (p1860).”

This finding demonstrates the possible tension between the platform designers’ intended purposes when implementing new affordances and their takeup by large and diverse userbases. The discussion of digital literacy above is particularly relevant to this problem, and it motivates out extension of Tandoc et al. (2022)’s study from the context of Singapore to the two Latin American countries in the current study. In particular, we investigate further the distinction between the demonstrated effect of the “two arrows” in reducing the credibility of messages while having no effect on the change in attitudes intended in those messages.

The primary self-generated cue we study is the presence or absence of a link to a news outlet with the full story in addition to the screenshot image in our simulated messages. This second dimension of interest, rather than evaluating a company’s strategy to combat misinformation, aims to assess the effect that the format of the news circulating on the platform can have on users’ perceptions and behavior. In WhatsApp, news circulating without any link redirecting to a site are common. In theory, these types of messages are more problematic as it is more difficult for the user to verify the source and judge whether it is credible or not. However, the very absence of a link can signal the trustworthiness of the content.



Understanding the impact of these self-generated cues through this arm of the experiment is relevant if we start from the idea that there is a high level of concern about the quality and accuracy of the information that circulates on WhatsApp (Rossini, Anita Baptista, et al., 2021; Zhu et al., 2022), and that the news could be consumed out of context of the media brands (Masip et al., 2021). However, the evidence here is limited. Some studies have found that the addition of a source cue on WhatsApp does not mediate the impact of other efforts, such as corrections (Sumitra et al., 2022) or interpersonal discussion (Vermeer et al., 2021).

An external link may have intended and unintended consequences that are worth exploring experimentally. In particular, our designs separately analyze the effect of possible cue sources on engagement, shareability, and credibility. Importantly, doing it experimentally allows us to control for the type of information that is transmitted. Exploring these possible differentiated effects is essential if we consider that the information communicated on WhatsApp can have an impact on offline behavior (de Zúñiga et al., 2021; Rossini, Stromer-Galley, et al., 2021) and that its nature of private communication allows the proliferation of corrections by part of the users (Badrinathan, 2021; Rossini, Stromer-Galley, et al., 2021).

Our experiments complement these efforts not only by exploring this type of strategy in other contexts and with two much more representative experimental designs but because it allows us to see variation within application-specific system-generated cues (forwarded vs. forwarded many times), directly comparing these strategies against environmental self-generated cues like the inclusion of a link to the news source, and by exploring essential heterogeneities by age, digital literacy, education, and political affiliation.

### ***Research Question and Hypotheses***

The above discussion of previous research ties together various strands of literature from a variety of contexts, so we now provide several formal hypotheses to translate our expectations into testable predictions. The overarching Research Question, then, is as follows:

**RQ1:** How well does previous research on digital literacy and platform cues translate to the context of WhatsApp use in Mexico and Colombia?

This Research Question is broken down into the following hypotheses:

**H1:** The correlation between education and digital literacy will be stronger than the correlation between age and digital literacy.

Our read of the literature on the drivers of digital literacy, as well as our substantive knowledge about the role of educational attainment in the context of Latin American society, motivates Hypothesis 1. Note that this Hypothesis is the opposite of results on this question in the context of the United States discussed above. This

fundamentally descriptive (that is, non-causal) hypothesis is important because of how it relates to the premises underlying theories of the spread of misinformation in different contexts. Our goal here is to test whether these contexts do in fact differ on this dimension.

**H2:** The inclusion of user-specific cues (here, a source cue) will cause an increase in the perceived credibility of the news, as well as the propensity to read and share it.

There is a substantial literature on the connection between user-specific cues and especially source cues and the credibility of news stories on social media more broadly. We therefore have strong expectations that this cue will have an effect here, across all three of our outcome measures.

**H3:** The inclusion of platform-specific cues (here, the “Muchas Veces” tag) will cause a decrease in the perceived credibility of the news, but an increase in the propensity to read and share it.

The literature on platform-specific cues provides more mixed findings, and of course different cues have different effects. Research on this specific “Forwarded Many Times” tag is inconclusive, and points to the possibility of different effects on the different outcome measures of interest.

**H4:** The main effects in H2 and H3 will be significantly moderated by age and by digital literacy.

These effect heterogeneities have been observed in a variety of experimental interventions on digital media in the United States discussed above. We investigate whether the same variables moderate the experimental cues in this case, and whether they apply here in Mexico and Colombia.

In the sections that follow, we test these four hypotheses using two sources of evidence: first, descriptive evidence from surveys conducted in these two countries on the socio-demographic characteristics of WhatsApp users in Colombia and Mexico. Second, experimental evidence from a randomized controlled trial, in which we varied the exposure to user- and platform-specific cues for the interviewed subjects.

## Sampling Internet Users

One challenge for research on low digital-literacy populations is that online convenience samples tend to dramatically undersample them (Munger et al., 2021). We use Offerwise, an established consulting company with over 16 years of experience in online market research in Latin America. Offerwise has one of the largest and most representative panels of respondents in Latin America and among U.S. Hispanics.<sup>8</sup> Recent published work using Offerwise to recruit Colombians include (Taillie et al.,

2020), and recruiting U.S. Hispanics include (Street et al., 2015). The survey and experimental design was approved by the Institutional Review Boards of participant universities.

The survey of 1,745 respondents was conducted online in all 32 states of Mexico in December 2021. The distribution of responses by state is similar to the national one (national data for Mexico comes from INEGI). Our sample has an average age of 36 compared to the national average of 27. Education levels are higher in our sample as we have 52% with higher education while the national average is 18%. Also, 60% of our sample were women compared to 51% of the country.

In the case of Colombia, the survey of 1,519 respondents was conducted in 28 of the 32 departments of the country in February 2022. The average age in our sample is 34 compared to the national average of 31 (data on Colombian nationals come from DANE and the World Bank). In terms of education, 51% have higher education in the sample compared to 13% at the national level. Additionally, 50% of the sample are women compared to 51% at the national level. The number of participants from indigenous groups was 2.3% and from AfroColombians 5.5% (compared to the national average of 4.3% and 6.8%, respectively).

Our sample therefore falls somewhere between an online convenience sample from familiar US-based survey respondent providers like MTurk or Lucid and a gold-standard, nationally representative survey like the AmericasBarometer or Latinobarómetro. According to Latinobarómetro (2020), for example, the average age for Mexican and Colombian participants is 45 and 43 years old, respectively. In terms of age, our Mexican sample's age (35) is between the national average (27) and the nationally representative survey. Our Colombian sample (33) is also closer to the national average (31) than the Latinobarómetro.

The younger age and higher education levels observed in our sample may be attributed to our focus on the WhatsApp users of both countries as our target population. According to Latinobarómetro (2020), WhatsApp users tend to be younger with an average age of 39 years old, which is consistent with our sample. Moreover, the percentage of individuals with higher education is also higher among WhatsApp users, with 30% in Colombia and 29% in Mexico, compared to the national average of 13% and 18%, respectively. These characteristics suggest that our sample closely resembles the online community in these countries.

While we therefore avoid making descriptive inferences about a general population of either Mexico or Colombia, our sample is appropriate to the analyses we conduct. We are interested in studying the use of a digital tool among a group of people who are already somewhat familiar with digital tools and, in any case, could use a platform like WhatsApp. That is, there are many citizens in these countries who cannot be reached by Offerwise's online survey – but these people are not part of our population of interest, which are the *internet users* of these two countries. Furthermore, based on accumulating evidence about the generalizability of survey experiments from online convenience samples to nationally representative target populations (Coppock, 2019; Coppock et al., 2018; Peyton et al., 2021), the estimated treatment effects from our central experiment should apply to a larger population of interest.

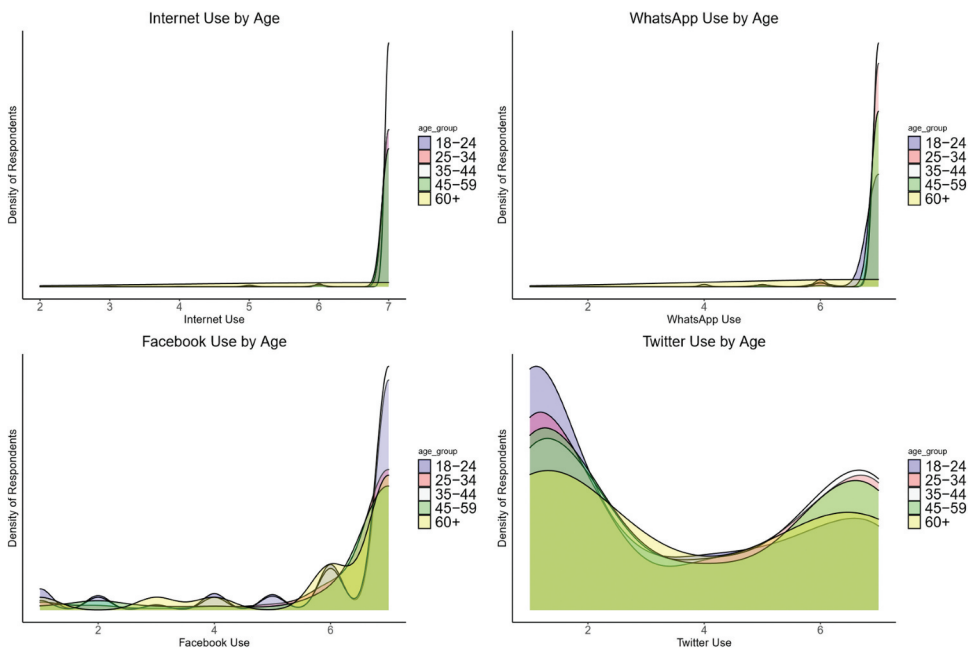


Figure 1. Age and internet, app use in Colombia.

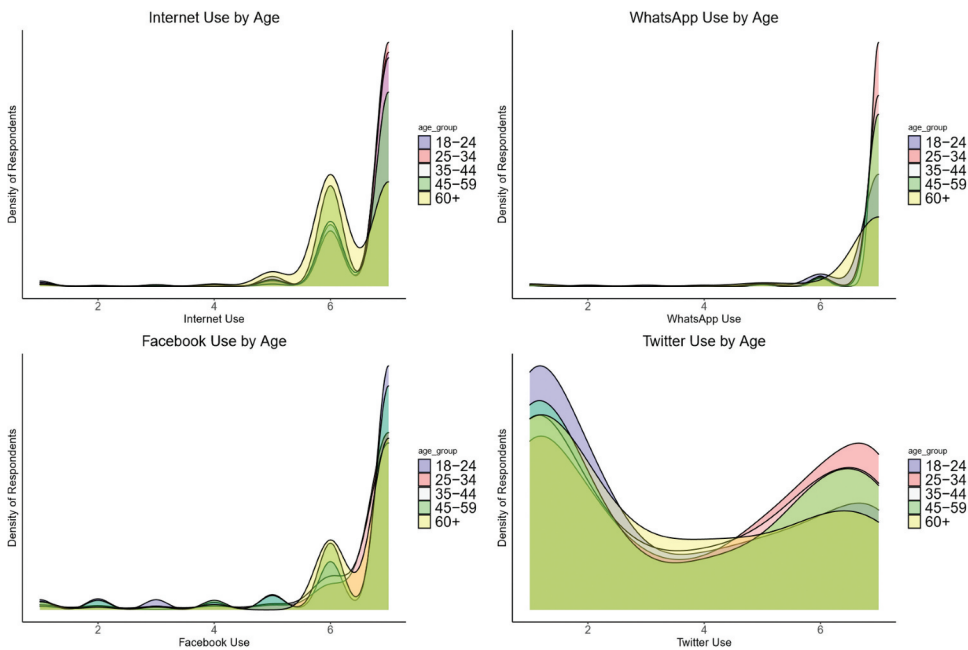


Figure 2. Age and Internet, app use in Mexico.

## Describing the Platform and User Environment

Figure 1 displays the density plots of internet and app use by respondent age in Colombia; Figure 2 does the same for Mexico. We ask participants to report the frequency with which they use the internet, WhatsApp, Facebook and Twitter. The x-axis represents the frequency

Of internet and app use from never at all (1) to multiple times a day to (7), while the y axis shows the percentage of respondents in each age category. The top row of each figure replicates previous findings that nearly every respondent in our survey uses WhatsApp many times a day, almost as much as use the internet. Most participants, across all age groups, use WhatsApp multiple times a day: 96% in Colombia and 91% in Mexico. In both countries, Facebook use is nearly as common as WhatsApp use, with 72% and 73% in Colombia and Mexico, respectively.

Next, we describe the distribution of our samples (Figure 3 for Colombia and Figure 4 for Mexico) along the three dimensions of digital literacy outlined above. The top row of each figure displays the distribution of the “Internet Skills” measure along both age and

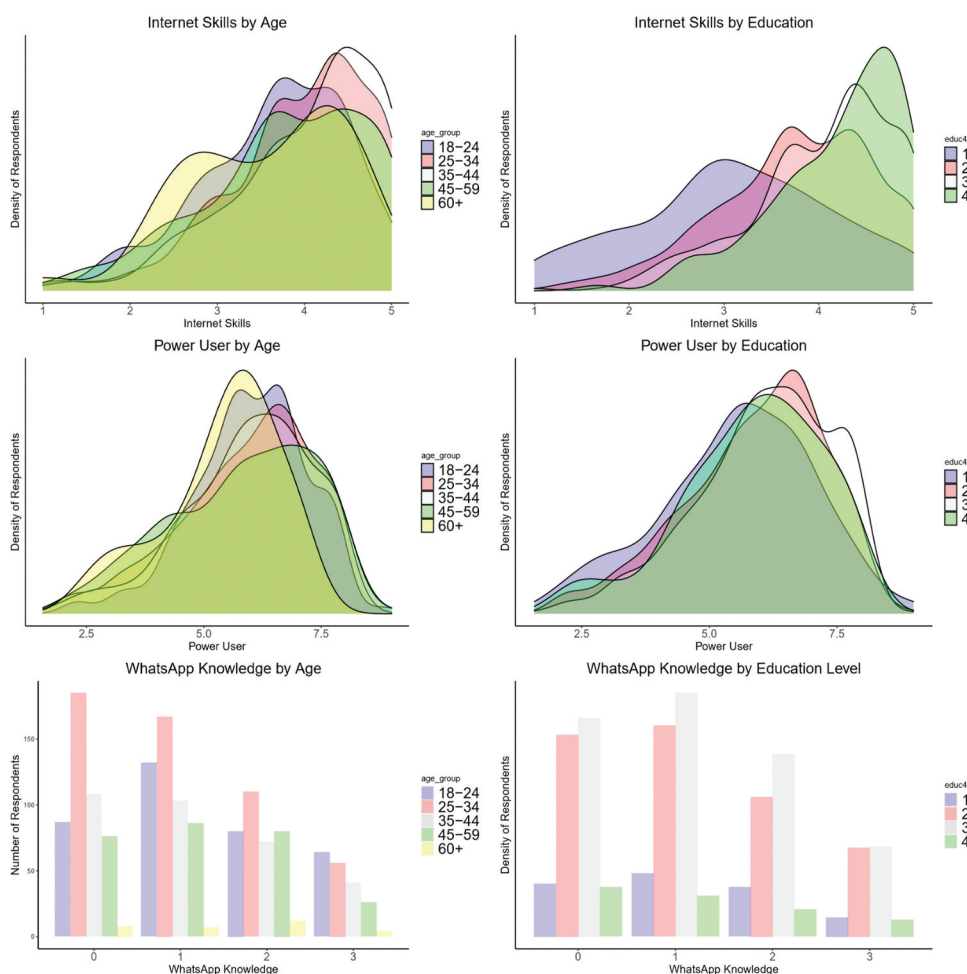
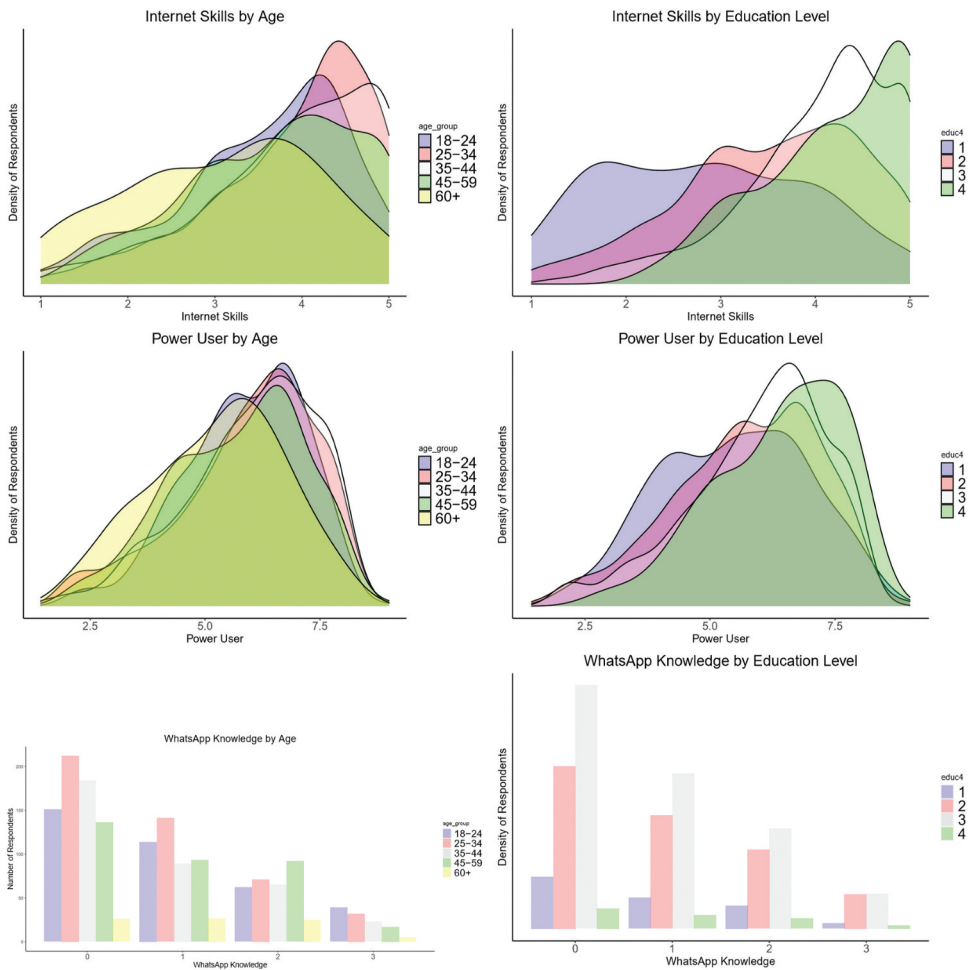


Figure 3. Age, education and internet skills, power use, advanced WhatsApp knowledge in Colombia.



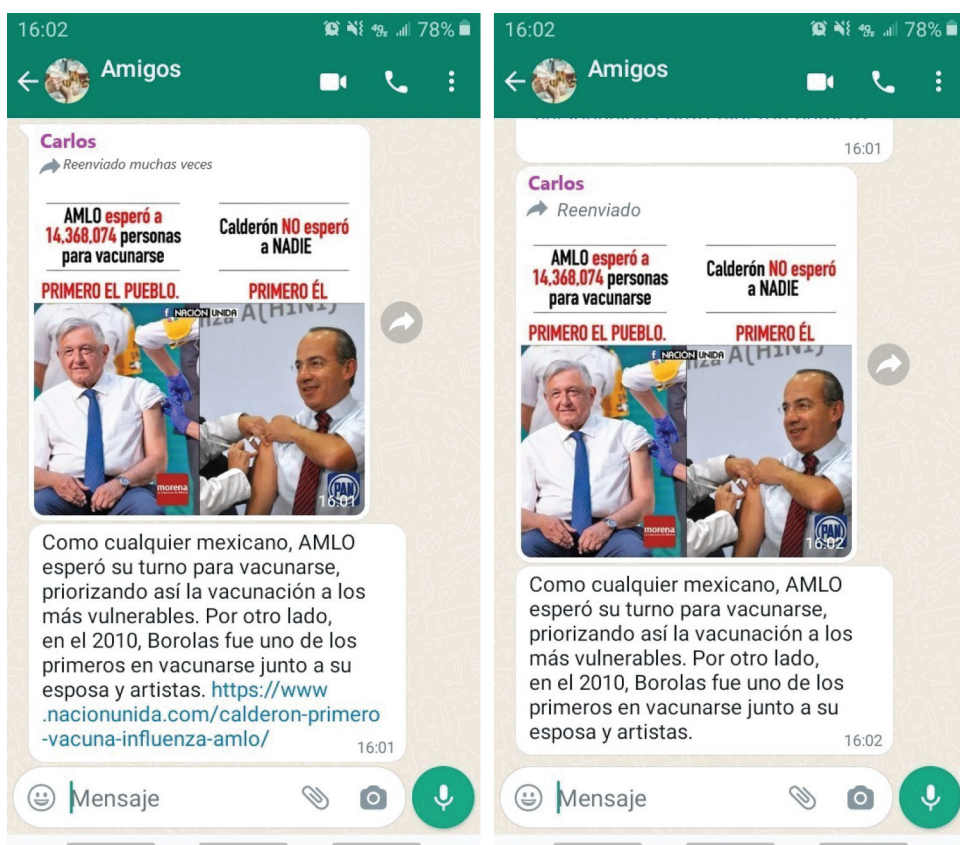
**Figure 4.** Age, education and internet skills, power use, advanced WhatsApp knowledge in Mexico.

education. We see significant heterogeneity on both dimensions and for both countries; in both cases, and in contrast to findings from the US, there is more variance along education than age. Furthermore, we find noticeable non-monotonicity in Internet Skills by age: in both countries, middle-aged respondents score higher than either younger or older respondents.

The second row of each figure displays respondents' scores on the "Power User" scale. There is less variance along this dimension in either age or education, although the oldest category of respondent scores lower than other age categories. The exception is in Mexico, where we observe significant and monotonic increase in "Power User" score by education. The difference between the two countries in this respect points to some significant difference in the role that formal education plays in either how people verify information they encounter online or how society is stratified more broadly; we leave this distinction to future research.

The bottom row of each figure demonstrates the scores on our novel measure of "Advanced WhatsApp Knowledge." This is a four-point measure, so the visualization is





**Figure 5.** Reenviado Muchas Veces and Link (left); Reenviado and no Link (Right).

bunched along those four points. Within the two countries, there are noticeable but not consistent trends in advanced WhatsApp Knowledge by both age and education. Particularly noteworthy, however, is that the effect is not monotonic: in both countries, the highest-education category is less likely to answer all three questions correctly than are people in the middle range of educational attainment.

In sum, the descriptive analysis presented in this section, particularly the analysis associated with the “Internet Skills,” of the respondents in Mexico and Colombia, provides evidence in favor of hypothesis H1 posed in [Section 2.4](#): the correlation between education and digital literacy is greater than between age and digital literacy in these two countries.

## WhatsApp-Specific and User-Specific Cues Experiment

### Design

We deployed the experimental section at the end of each survey in both countries. The two experiments follow the same design: in each study, we showed each respondent a series of ostensible screenshots of WhatsApp messages that included recent news

stories from the respective countries. We rely on easily recognizable political events for selecting the news. The news stories referred to visible figures of the left, right, and center in Colombia and left and right in Mexico. The political parties in Mexico have recently been divided between right and left, with no strong centrist option. We classified traditional parties based on their support or opposition to the leftist party coalition, MORENA, and thus focused solely on the right and left classification in Mexico. In contrast, we adopted a three-part distribution in Colombia that included a center category. Although some parties have shifted between right and center during certain historical periods, we analyzed separately the parties that were clearly classified as center during our study period.

To select the news stories, we sampled from real recent news stories rated by fact-checking websites as false in each country. Specifically, we used *ColombiaCheck* and *La Silla Vacía* in Colombia and *Animal Político* in Mexico.<sup>9</sup> Like PolitiFact or FactCheck.org in the U.S., these websites are dedicated to rate the accuracy of media content, including viral pictures on social media, claims by elected officials, articles in newspapers, etc. These news generally presented information about salient leaders in both countries, such as Alvaro Uribe and Gustavo Petro in Colombia, or Andres Manuel Lopez Obrador and Felipe Calderon in Mexico. In an environment of high polarization around the elections, this is the type of information circulating massively on WhatsApp in both countries.<sup>10</sup>

We used a  $2 \times 2$  factorial design, where we manipulated two aspects of the image: whether the linked news item included the WhatsApp-supplied label “*Reenviado muchas veces*” or simply “*Reenviado*”<sup>11</sup>; and whether the linked items contained a link to a news outlet or was simply an image and a text taken from the news outlet. We call these the *Shares Treatment* and the *Source Treatment*, respectively.<sup>12</sup>

After the respondents answered the demographic and social media usage questions, they were equally likely to be assigned to one of the experimental conditions described above. Figure 5 shows example stimuli that demonstrate the relevant dimensions of manipulation. The left panel demonstrates the “*Reenviado muchas veces*” message and includes a link to the original story; the right panel demonstrates the “*Reenviado*” message with no link. The Appendix includes a description of the instrument with all the images and text used.

After each image or text, participants answered a series of outcome questions. In both studies we analyzed three main outcomes related to the respondent’s trust in the information presented. We asked respondents three questions about the content on the image or text: a five-point scale of “How likely are you to read the article associated with the screenshot?,” a binary question about “How credible do you think the article associated with the screenshot is?,” and a five-point scale of “How likely is it for you to share the article associated with the screenshot?” We analyze each of the outcomes separately.

We estimate the effect of the intervention on each of the outcomes via OLS. We compute robust standard errors clustered at the respondent level. We estimate the following equation:

$$Y_i = \alpha + \beta_1(\text{Source}) + \beta_2(\text{Source} + \text{Muchas}) + \beta_3(\text{No} - \text{Source} + \text{Muchas}) + \gamma X_i + \varepsilon_i,$$

where  $Y_i$  is the outcome variable of interest, “Source” denotes assignment to the Source Treatment, “Source + Muchas” denotes assignment to both the Source and Shared Treatments, and “No-Source + Muchas” denotes assignment to the Shared Treatment only. Finally,  $X_i$  is the vector of covariate values for individual  $i$ ,  $\varepsilon_i$  is the error term and  $\alpha$ ,

$\beta_1, \beta_2, \beta_3$ , and  $\gamma$  are the parameters to be estimated. In this way, the coefficients  $\beta_1, \beta_2$  and  $\beta_3$ , in addition to the analysis of heterogeneous treatment effects that we describe in the following section, allow us to corroborate hypotheses H2-H4 of [section 2.4](#), related to the causal effects of user- and platform-specific cues on the behavior of WhatsApp users in Colombia and Mexico.

## Experimental Results

We turn now to the results of our survey experiment designed to test the effects of different WhatsApp-specific source cues. [Table 1](#) displays the results from the experiment implemented in Colombia. There are a number of strong but non-experimental relationships between individual characteristics and each of the outcome measures, and while they are not our focus here, they may be of interest to inspire future research.

The experimental results can be found in the first three rows of [Table 1](#). Relative to the condition with simply “Reenviado” and no link, the addition of a link (rows 1 and 2) increases the likelihood that the respondent will read, share and find the message credible. However, the addition of “Muchas Veces” without a source cue (row 3) fails to produce any significant effect.

**Table 1.** Mexico: Treatment Effects (Standard Errors Clustered at Respondent).

	Read (1)	Read (2)	Share (3)	Share (4)	Credible (5)	Credible -6
Source	0.08* (0.05)	0.20 (0.14)	0.08** (0.04)	0.12 (0.12)	0.07*** (0.02)	0.14*** -0.05
Source + Muchas	0.07 (0.05)	0.16 (0.15)	0.07* (0.04)	0.12 (0.12)	0.05*** (0.02)	0.12** -0.05
No Source + Muchas	0.05 (0.05)	0.18 (0.14)	0.04 (0.04)	0.13 (0.12)	0.01 (0.01)	0.06 -0.05
Female	-0.18*** (0.06)	-0.18*** (0.06)	-0.21*** (0.05)	-0.21*** (0.05)	0.0005 (0.02)	0.0003 -0.02
Education	-0.19*** (0.04)	-0.19*** (0.04)	-0.13*** (0.04)	-0.13*** (0.04)	-0.02* (0.01)	-0.02* -0.01
Skills	0.03 (0.04)	0.03 (0.04)	0.03 (0.04)	0.03 (0.04)	0.004 (0.01)	0.004 -0.01
Power	-0.12*** (0.02)	-0.12*** (0.02)	-0.16*** (0.02)	-0.16*** (0.02)	-0.01 (0.01)	-0.01 -0.01
Knowledge	0.04 (0.03)	0.04 (0.03)	0.01 (0.03)	0.01 (0.03)	0.004 (0.01)	0.004 -0.01
Age	-0.001 (0.003)	0.002 (0.004)	-0.004 (0.002)	-0.002 (0.003)	-0.003*** (0.001)	-0.002* -0.001
Center Party	0.32*** (0.08)	0.32*** (0.08)	0.38*** (0.07)	0.38*** (0.07)	0.02 (0.02)	0.02 -0.02
Right Party	0.20** (0.09)	0.20** (0.09)	0.25*** (0.08)	0.25*** (0.08)	0.02 (0.02)	0.02 -0.02
Left Party	0.39*** (0.08)	0.39*** (0.08)	0.23*** (0.08)	0.23*** (0.08)	0.05** (0.02)	0.05** -0.02
Share Political News	0.05** (0.02)	0.05** (0.02)	-0.02 (0.02)	-0.02 (0.02)	-0.001 (0.01)	-0.001 -0.01
Source X Age		-0.004 (0.004)		-0.001 (0.003)		-0.002* -0.001
Source + Muchas X Age		-0.003 (0.004)		-0.002 (0.003)		-0.002 -0.001
No Source + Muchas X Age		-0.004 (0.004)		-0.003 (0.003)		-0.001 -0.001
Constant	3.60*** (0.21)	3.51*** (0.23)	3.28*** (0.20)	3.23*** (0.21)	0.50*** (0.06)	0.45*** -0.06

Turning to heterogeneous treatment effects, we find in column 6 some evidence that older people are less responsive to the addition of a link than are younger people. However, we caution that this result is significant only at  $p < .1$ .

Table 2 displays the results from the experiment we implemented in Mexico. In this case, the addition of the source cue in treatments 1 and 2 also has an effect on each outcome in at least one specification. Here, however, the addition of “Muchas Veces” in row 3 causes a significant increase in the likelihood that the response would share the story, while leaving the other two outcomes unchanged. In terms of heterogeneous effects, we only find evidence of an interaction between age and treatment 1 on the likelihood of reading the story.

All in all, we obtain support for some but not all of the specified experimental hypotheses. First, news that includes the user-specific cue of an identifiable source are more likely to be read and shared. They are also reported to be more credible, although we point out that this effect size is roughly twice as large in Colombia as in Mexico. This generally provides support for Hypothesis H2 of section 2.4. However, the platform-specific cue that the article has been shared many times did not have any effect in the case of Colombia. In Mexico, in according with our Hypothesis H3, we find the expected increase in the likelihood of sharing the story, but no effect on the other two outcomes of interest.

**Table 2.** Colombia: Treatment Effects (Standard Errors Clustered at Respondent).

	Read (1)	Read (2)	Share (3)	Share (4)	Credible (5)	Credible –6
Source	0.01 (0.04)	0.25** (0.13)	0.07* (0.04)	0.15 (0.12)	0.03* (0.01)	0.08* –0.04
Source + Muchas	0.01 (0.04)	0.10 (0.13)	0.09** (0.04)	0.02 (0.12)	0.01 (0.01)	0.05 –0.04
No Source + Muchas	0.04 (0.04)	0.005 (0.13)	0.09** (0.04)	0.08 (0.12)	0.02 (0.01)	0.07 –0.05
Female	0.07 (0.06)	0.07 (0.06)	0.02 (0.05)	0.02 (0.05)	0.01 (0.01)	0.01 –0.01
Education	–0.16*** (0.04)	–0.16*** (0.04)	–0.09** (0.04)	–0.09** (0.04)	–0.01 (0.01)	–0.01 –0.01
Skills	–0.05 (0.03)	–0.05 (0.03)	–0.04 (0.03)	–0.04 (0.03)	0.002 (0.01)	0.002 –0.01
Power	–0.05** (0.02)	–0.05** (0.02)	–0.12*** (0.02)	–0.12*** (0.02)	–0.01 (0.01)	–0.01 –0.01
Knowledge	0.06** (0.03)	0.06** (0.03)	0.01 (0.03)	0.01 (0.03)	0.01** (0.01)	0.01** –0.01
Age	0.002 (0.002)	0.004 (0.003)	–0.01*** (0.002)	–0.01*** (0.003)	–0.004*** (0.001)	–0.003*** –0.001
Right Party	0.23*** (0.07)	0.23*** (0.07)	0.36*** (0.06)	0.36*** (0.06)	0.07*** (0.02)	0.07*** –0.02
Share Political News	0.07*** (0.02)	0.07*** (0.02)	0.03 (0.02)	0.03 (0.02)	0.01 (0.01)	0.01 –0.01
Left Party	0.52*** (0.06)	0.52*** (0.06)	0.56*** (0.06)	0.56*** (0.06)	0.10*** (0.01)	0.10*** –0.01
Source X Age		–0.01** (0.003)		–0.002 (0.003)		–0.002 –0.001
Source + Muchas X Age		–0.003 (0.003)		0.002 (0.003)		–0.001 –0.001
No Source + Muchas X Age		0.001 (0.004)		0.0001 (0.003)		–0.001 –0.001
Constant	3.26*** (0.20)	3.18*** (0.21)	3.16*** (0.18)	3.16*** (0.20)	0.42*** (0.05)	0.39*** –0.05

The evidence for our Hypothesis H4 is generally very weak. We find two instances of effect heterogeneity in age, but they are spread across different outcome measures and countries without any informative pattern. And for effect heterogeneity in digital literacy, operationalized in three different ways, we find consistently null results that we do not report here for space reasons. For H4, then, we find null results.

In the Appendix, we conduct exploratory heterogeneity analysis in order to take full advantage of our dataset and explore whether there are differences in treatment effects for participants who previously encountered fake news on WhatsApp. As Appendix Table A1 and Table A2 show, the results remain the same even after controlling for previous exposure to fake news.

We also examine whether the treatments have different effects based on the partisanship of the party or candidate being attacked (Appendix Table A3 and Table A4). Although partisanship influences people's perception of information on social media, we find no consistent partisan differences in treatment effects. For example, the effect of the "Reenviado" label in Colombia is always positive, regardless of whether the news targets the left, center, or right. However, there is a general

partisan difference in the reception of misinformation, particularly in Mexico. Participants on the left (right) are more likely to find misinformation attacking right- (left-) wing parties or figures credible. Nonetheless, all participants are equally likely to read and share misinformation attacking the left or right. In Colombia, those who identified with left- (right-) wing parties are less likely to read, share, or find credible misinformation attacking the left (right). Also, participants who identified with a centrist party are equally likely to read, share, or find credible misinformation attacking rightist, leftist, or centrist parties.

## Conclusion

The technological shocks of the internet, smartphones and social media are re-organizing the social and political landscape. In particular, the information flows that underwrite the functioning of democracy have been transformed. Acknowledging the magnitude of these shocks does not entail a naive technological determinism. Indeed, it is essential that political communication researchers recognize the dizzying variety of forms that digital communication have taken as different communities have incorporated the internet differently.

Our paper is a minor step toward addressing the widely-bemoaned inequality in the scholarly attention paid to social media use in the Global North. Rather than starting from scratch in each new sociotechnical context, we adapted relevant theories and methods to reflect the dominant ways in which social media is used for political communication in Mexico and Colombia.

"Digital Literacy" is an increasingly prominent conceptual frame that emphasizes the heterogeneity in how citizens interpret and interact with information encountered on the internet. But we argue that a universal theory of "Digital Literacy" that measures the same skills and capacities regardless of context is worse than useless, advocating instead for theoretically-informed "digital literacies" that can be empirically validated based on the contexts they aim to explain.

Whereas the paradigmatic internet user in existing conceptions of digital literacy is English speaking, using a desktop, navigating a feed-driven social media platform, above

the global median in formal education, and unconstrained by data access, we develop and apply the concept in the context of a paradigmatic user located in the Global South, using a mobile device, navigating a group chat-driven social media platform, potentially quite low in formal education, and potentially constrained by mobile data credits. From this premise, we begin to establish descriptive facts about how Colombians and Mexicans use the internet and various social media platforms.

We then deployed existing survey measures of digital literacy in order to establish a baseline for future research. As expected, these measures are not perfectly suited to the WhatsApp context, but neither are they irrelevant. The most striking contrast to trends observed in the United States is that the Internet Skills measures tend to be more tightly correlated with education levels than with respondent age. This result corroborates the first testable hypothesis that we proposed in this study. Our theoretical understanding of the social and political phenomena that tend to develop digital literacy causes us to predict that the United States is the outlier here. Similar to Mexico and Colombia, most nations in the Global South have a lower number of college degree graduates and a higher variation of education levels than the United States, making education (a proxy for human capital) more relevant than age to how citizens of these nations understand the information they see on the internet. Future analysis of digital literacy in the Global South should pay particular attention to respondent education levels.

Finally, we deployed a survey experiment to understand how our respondents evaluate information they encounter on WhatsApp. Mimicking the format of a news story shared within a group chat, we manipulated two features of the story: whether it included a user-specific cue in the form of a link to the source, and whether it included a platform-specific cue in the form of the WhatsApp-supplied label “Reenviado Muchas Veces.”

We describe a variety of demographic correlations that bear future investigation – for example, that older respondents in both countries found all of the stories less credible than did younger respondents, and that more educated respondents were consistently less likely to want to read or share the stories. Among our experimental manipulations, we find that the addition of the userspecific cue increases the likelihood that respondents will want to read, share and find credible the story in a majority of specifications. This allows us to corroborate the second testable hypothesis raised in the study. In contrast, there is only evidence that the platform-specific “reenviado muchas veces” label affected one of these outcomes, the likelihood of sharing the news story (and here, only in Mexico and not in Colombia). In contrast to our third Hypothesis, we find no effect of this cue on the respondent’s estimation of the credibility of the news or likelihood of reading it.

These findings offer support for the growing consensus in the misinformation literature that the propensity to share news is not driven (solely) by credibility; indeed, following the results in Tandoc et al. (2022), we find that the platform-specific cue had an effect on the desire to share the news without affecting perceived credibility.

Finally, and contrary to our fourth hypothesis, we find that null results for treatment effect heterogeneity in both age and digital literacy, with the latter operationalized in three different ways. This result connects to one of the central motivations for the paper: the difficulty in transporting both theoretical and empirical knowledge from one context to substantially different ones. We do not conclude from these findings that the effects of the platform- and user-specific cues are constant across all types of people in Mexico and



Colombia, but rather that there is more work to be done in both conceptualizing and operationalizing the relevant moderating variables.

It is possible, for example, that there is a distinct operationalization of “digital literacy” which would reveal that this is in fact a relevant treatment effect moderator. However, it is also possible that the underlying cultural or psychological mechanism is distinct from what we have theorized. This kind of question is best answered by qualitative scholars, and we encourage experts in these methods to continue to explore the topic.

Another crucial limitation of these findings stems from the challenges presented by quantitative research based on sampling from an unknown population. We are particularly concerned with understanding the experience of low digital literacy citizens in Mexico and Colombia, but this group is definitionally unlikely to appear in any online survey sample. Although we worked with an established market researcher specializing in this region, we cannot be certain that we reached a representative sample because of this structural bias. We encourage future research – likely qualitative, based on a small but intentional sample – to understand the magnitude of this bias.

We encourage further research on how people in different contexts understand the information they encounter on the internet. There is a growing body of research on digital communication in the Global South focused on “misinformation,” but we argue that this topic can only be addressed superficially without a richer understanding of the underlying information ecosystems.

## Notes

1. <https://jp.reuters.com/article/mexico-facebook/in-mexico-fake-news-creators-up-their-game-ahead-of-election-idUSL1N1TT29Z>. Last accessed: 03/29/2023.
2. <https://www.lasillavacia.com/historias/silla-nacional/la-falsa-noticia-de-la-suspension-de-elecciones-entumaco>. Last accessed: 03/29/2023.
3. See <https://www.integracionsocial.gov.co/index.php/noticias/116-otras-noticias/4517-como-usan-latecnologia-las-personas-mayores>. Last accessed: 10/12/2022.
4. See <https://es.statista.com/estadisticas/1139347/distribucion-redes-sociales-usuarios-edad-generomexico/>. Last accessed: 10/12/2022.
5. See <https://branch.com.co/marketing-digital/estadisticas-de-la-situacion-digital-de-colombia-en-el-20202021/>. Last accessed: 10/12/2022.
6. See <https://dplnews.com/whatsapp-desde-ahora-avisara-al-usuario-si-un-mensaje-ha-sido-reenviadomuchas-veces/>. Last accessed: 10/5/2022.
7. See <https://www.enter.co/chips-bits/apps-software/whatsapp-mensajes-reenviados-2/>. Last accessed: 10/5/2022.
8. See <https://www.offerwise.com> for more information.
9. See <https://colombiacheck.com/>, <https://www.lasillavacia.com/la-silla-vacia/detector-de-mentiras/>, and <https://www.animalpolitico.com/sabueso/>. Last accessed: 4/2/2022.
10. The focus on fake news is due to the interest in curbing the communication of this type of content. Therefore, our results apply to this type of news, not necessarily news and other information shared on the platform. The focus on fake news may reduce variation in the probability of respondents sharing, reading or trusting the information.
11. “Forwarded” messages can be sent to up to five chats at a time, while “forwarded many times” messages can only be forwarded to one chat at a time. The “forwarded many times” feature was introduced in 2019, in an attempt to combat misinformation and fake news. See <https://tech.hindustantimes.com/tech/news/whatsapp-now-tells-users-when-a-message-has-beenforwarded-many-times-story-IRFr4I0jLGFM0VNQZLrFPP.html>. Last accessed: 4/2/2022.

12. To assess whether or not WhatsApp's "Reenviado muchas veces" strategy was effective, it seems reasonable to use as a comparison the placebo strategy in which the message is simply labeled "Reenviado." It would have been useful to study how either of these two labels compares with a pure control in which no label is used at all. Unfortunately, since we had another dimension (inclusion of a link) in our factorial design, for reasons of statistical power we could not include this scenario.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The work was supported by a Facebook Foundational Integrity Research given to the corresponding author in 2020.

## Notes on contributors

*Kevin Munger* is an Assistant Professor of Political Science and Social Data Analytics at Penn State University. He studies the internet and politics.

*Angel Villegas-Cruz* is a Ph.D. Candidate in Political Science and Asian Studies at the Pennsylvania State University. His research interests focus on the relationship between digital technologies and international politics, with an area specialization in China and Latin America.

*Jorge Gallego* is a senior economist at the Office of Evaluation and Oversight (OVE) of the Inter-American Development Bank and an associate professor of Economics at Universidad del Rosario.

*Mateo Vásquez-Cortés* (Ph.D., New York University) is an Assistant Professor in the School of Global Policy and Strategy at the University of California, San Diego. His research explores questions related to conflict processes, identity, and migration in the Global South.

## Data availability

The data will be made available upon request by contacting the corresponding author

## data availability

The data will be made available upon request by contacting the corresponding author

## Open scholarship



This article has earned the Center for Open Science badge for Open Data. The data are openly accessible at <https://osf.io/kf8ex>

## References

- Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in India. *American Political Science Review*, 115(4), 1325–1341. <https://doi.org/10.1017/S0003055421000459>
- Barbera, P. (2018). Explaining the spread of misinformation on social media: Evidence from the 2016 U.S. Presidential election. In *Symposium: Fake news and the Politics of Misinformation*. APSA.
- Baulch, E., Matamoros-Fernández, A., & Johns, A. (2020). Introduction: Ten years of WhatsApp: The role of chat apps in the formation and mobilization of online publics. *First Monday*, 25(1). <https://doi.org/10.5210/fm.v25i12.10412>
- Ceci, L. (2021). Global WhatsApp users in selected countries 2021. <https://www.statista.com/statistics/289778/countries-with-the-most-facebook-users/>
- Cheeseman, N., Fisher, J., Hassan, I., & Hitchen, J. (2020). Social media disruption: Nigeria's WhatsApp politics. *Journal of Democracy*, 31(3), 145–159. <https://doi.org/10.1353/jod.2020.0037>
- Coppock, A. (2019). Generalizing from survey experiments conducted on mechanical Turk: A replication approach. *Political Science Research and Methods*, 7(3), 613–628. <https://doi.org/10.1017/psrm.2018.10>
- Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, 115(49), 12441–12446. <https://doi.org/10.1073/pnas.1808083115>
- Correa, T., Valenzuela, S., & Pavez, I. (2022). For better and for worse: A panel survey of how mobile-only and hybrid internet use affects digital skills over time. *New Media & Society*, 26(2), 995–1017. <https://doi.org/10.1177/14614448211059114>
- Cortes, C., & Peñarredonda, D. (2018). *Politics (on WhatsApp) is dynamic. Disinformation and the spread of political chains in Colombia*. Technical report Linterna Verde.
- Cruz, E. G., & Harindranath, R. (2020). WhatsApp as 'technology of life': Reframing research agendas. *First Monday*. <https://doi.org/10.5210/fm.v25i12.10405>
- Davies, W. (2020). *What is wrong with WhatsApp*. <https://www.theguardian.com/technology/2020/jul/02/whatsapp-groups-conspiracytheories-disinformation-democracy>
- de Zúñiga, G., Homero, A.-A.-A., & Casero-Ripollés, A. (2021). WhatsApp political discussion, conventional participation and activism: Exploring direct, indirect and generational effects. *Information, Communication & Society*, 24(2), 201–218. <https://doi.org/10.1080/1369118X.2019.1642933>
- Garimella, K., & Eckles, D. (2020). Images and misinformation in political groups: Evidence from WhatsApp in India. *Harvard Kennedy School Misinformation Review*, arXiv preprint arXiv:2005.09784. <https://doi.org/10.37016/mr-2020-030>
- Gilster, P. (1997). *Digital literacy*. John Wiley & Sons, Inc.
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. *Science*, 363(6425), 374–378. <https://doi.org/10.1126/science.aau2706>
- Guess, A. M., & Munger, K. (2020). Digital literacy and online political behavior. *Political Science Research and Methods*, 11.1(2023), 110–128.
- Guess, A. M., & Munger, K. (2023). Digital literacy and online political behavior. *Political Science Research and Methods*, 11(1), 110–128. <https://doi.org/10.1017/psrm.2022.17>
- Guess, A., Nagler, J., & Tucker, J. (2019a). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1). <https://doi.org/10.1126/sciadv.aau4586>
- Guess, A., Nagler, J., & Tucker, J. (2019b). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- Hargittai, E. (2001). Second-level digital divide: mapping differences in people's online skills. *First Monday*, 7(4), arXiv preprint cs/0109068. <https://doi.org/10.5210/fm.v7i4.942>
- Hargittai, E. (2005). Survey measures of web-oriented digital literacy. *Social Science Computer Review*, 23(3), 371–379. <https://doi.org/10.1177/0894439305275911>
- Hargittai, E. (2010). Digital natives? Variation in internet skills and uses among members of the "net generation". *Sociological Inquiry*, 80(1), 92–113. <https://doi.org/10.1111/j.1475-682X.2009.00317.x>

- Hargittai, E., & Patrick Hsieh, Y. (2012). Succinct survey measures of web-use skills. *Social Science Computer Review*, 30(1), 95–107. <https://doi.org/10.1177/0894439310397146>
- Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media, Culture & Society*, 33(2), 211–221. <https://doi.org/10.1177/0163443710393382>
- Latinobarómetro. (2018). Informe 2018. Technical report Corporación Latinobarómetro.
- Latinobarómetro. (2020). Informe 2020. Technical report Corporación Latinobarómetro.
- Machado, C., Kira, B., Narayanan, V., Kollanyi, B., & Howard, P. (2019). A study of misinformation in WhatsApp groups with a focus on the Brazilian presidential elections. In *Companion proceedings of the 2019 World Wide Web conference*. pp. 1013–1019.
- Masip, P., Suau, J., Ruiz-Caballero, C., Capilla, P., & Zilles, K. (2021). News engagement on closed platforms. Human factors and technological affordances influencing exposure to news on WhatsApp. *Digital Journalism*, 9(8), 1062–1084. <https://doi.org/10.1080/21670811.2021.1927778>
- Melo, F., de, P., Coimbra Vieira, C., Garimella, K., Melo, P. O., & Benevenuto, F. (2019). Can WhatsApp counter misinformation by limiting message forwarding? In *International conference on complex networks and their applications*, Pittsburgh, PA (pp. 372–384). Springer.
- Munger, K., Gopal, L., Nagler, J., & Tucker, J. A. (2021). Accessibility and generalizability: Are social media effects moderated by age or digital literacy? *Research & Politics*, 8(2), 20531680211016968. <https://doi.org/10.1177/20531680211016968>
- Pew. (2021). Internet/Broadband fact sheet. *Fact Sheet*.
- Peyton, K., Huber, G. A., & Coppock, A. (2021). The generalizability of online experiments conducted during the COVID-19 pandemic. *Journal of Experimental Political Science*, 9.3(2022), 379–394.
- Rossini, P., Anita Baptista, É., Veiga de Oliveira, V., & Stromer-Galley, J. (2021). Digital media landscape in Brazil: Political (mis) information and participation on Facebook and WhatsApp. *Journal of Quantitative Description: Digital Media*, 1. <https://doi.org/10.51685/jqd.2021.015>
- Rossini, P., Stromer-Galley, J., Anita Baptista, E., & Veiga de Oliveira, V. (2021). Dysfunctional information sharing on WhatsApp and Facebook: The role of political talk, cross-cutting exposure and social corrections. *New Media & Society*, 23(8), 2430–2451. <https://doi.org/10.1177/1461444820928059>
- Santini, R. M., Tucci, G., Salles, D., & de Almeida, A. R. D. (2021). Do you believe in fake after all? WhatsApp disinformation campaign during the Brazilian 2018 presidential election. In Guillermo López-García (Ed.), *Politics of Disinformation: The Influence of Fake News on the Public Sphere* (pp. 49–66).
- Santos, M., Saldaña, M., & Rosenberg, A. (2020). From access deprivation to skill acquisition: Cluster analysis of user behavior in face of a 12-hour legal blockage of WhatsApp in Brazil. In *First Monday*.
- Shan, Y. (2016). How credible are online product reviews? The effects of self-generated and system-generated cues on source credibility evaluation. *Computers in Human Behavior*, 55, 633–641. <https://doi.org/10.1016/j.chb.2015.10.013>
- Street, A., Zepeda-Millán, C., & Jones-Correa, M. (2015). Mass deportations and the future of latino partisanship. *Social Science Quarterly*, 96(2), 540–552. <https://doi.org/10.1111/ssqu.12158>
- Sumitra, B., & Chauchard, S. (2022). ”I don’t think that’s true, bro!” Social Corrections of Misinformation in India”. In *The International Journal of Press/Politics*. 19401612231158770.
- Sundar, S. S., & Marathe, S. S. (2010). Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research*, 36(3), 298–322. <https://doi.org/10.1111/j.1468-2958.2010.01377.x>
- Taillie, L. S., Hall, M. G., Fernando Gómez, L., Higgins, I., Bercholz, M., Murukutla, N., & Mora-Plazas, M. (2020). Designing an effective front-of-package warning label for food and drinks high in added sugar, sodium, or saturated fat in Colombia: An online experiment. *Nutrients*, 12(10), 3124. <https://doi.org/10.3390/nu12103124>
- Tandoc, E. C., Jr., Rosenthal, S., Yeo, J., Ong, Z., Yang, T., Malik, S., Ou, M., Zhou, Y., Zheng, J., Mohamed, H. A. B., & Tan, J. (2022). Moving forward against misinformation or stepping back? WhatsApp’s forwarded tag as an electronically relayed information cue. *International Journal of Communication*, 16, 18.

- Treré, E. (2020). The banality of WhatsApp: On the everyday politics of backstage activism in Mexico and Spain. *First Monday*, 25. <https://doi.org/10.5210/fm.v25i12.10404>
- Vermeer, S. A., Kruikemeier, S., Trilling, D., & de Vreese, C. H. (2021). WhatsApp with politics?! Examining the effects of interpersonal political discussion in instant messaging apps. *The International Journal of Press/politics*, 26(2), 410–437. <https://doi.org/10.1177/1940161220925020>
- Zhu, Q., Esteve-Del-Valle, M., & Meyer, J. K. (2022). Safe spaces? Grounding political talk in WhatsApp groups. *New Media & Society*, 14614448221136080. <https://doi.org/10.1177/14614448221136080>