A Mixed Methods Study of Public Perception of Social Distancing: Integrating Qualitative and Computational Analyses for Text Data

Pauline Ho¹, Kaiping Chen¹, Anqi Shao¹, Luye Bao¹, Angela Ai¹, Adati Tarfa¹, Dominique Brossard¹, Lori Brown¹, and Markus Brauer¹

Abstract
In a rapidly changing public health crisis such as COVID-19, researchers need innovative approaches that can effectively link qualitative approaches and computational methods. In this article, computational and qualitative methods are used to analyze survey data collected in March 2020 (n = 2,270) to explore the content of persuasive messages and their relationship with self-reported health behavior—that is, social distancing. Results suggest that persuasive messages, based on participants’ perspectives, vary by gender and race and are associated with self-reported health behavior. This article illustrates how qualitative analysis and structural topic modeling can be used in synergy in a public health study to understand the public’s perception and behavior related to science issues. Implications for health communication and future research are discussed.

Keywords
computational social science, mixed methods research, public health research, qualitative method, science communication

In the absence of pharmaceutical measures to cure and prevent the spread of coronavirus disease 2019 (COVID-19), nonpharmacological interventions such as social distancing, hand washing, and mask wearing were cautionary measures implemented to delay the spread of the virus in the spring of 2020 (World Health Organization [WHO], 2020a). The practice of social distancing is associated with a lower likelihood of an individual contracting COVID-

¹University of Wisconsin, Madison, USA

Pauline Ho and Kaiping Chen contributed equally to this work.

Corresponding Authors:
Pauline Ho, Department of Educational Psychology, University of Wisconsin–Madison, WI 53705, USA
Email: pho25@wisc.edu

Kaiping Chen, Department of Life Sciences Communication, University of Wisconsin-Madison, WI 53705, USA.
Email: kchen67@wisc.edu
19 (Clipman et al., 2020). As such, it was imperative early in the pandemic to explore the factors that informed a person’s decision to practice social distancing. The substantive aim of this study is to integrate quantitative and qualitative data to gain a synergistic understanding of individuals’ perceptions of social distancing as they considered persuasive messages that could influence their social distancing behavior in the earlier stages of the pandemic. Thus, our methodological aim is to integrate computational and qualitative findings to understand individuals’ perceptions of persuasive messages and their practice of health-related behaviors. Our study illustrates how both computational and qualitative methods can be used synergistically in a public health research study to understand the public’s perceptions and behavior during a pandemic. We start by reviewing the use of mixed methods in public health research, followed by an overview of COVID-19 and social distancing. We then discuss our study in greater detail.

The Use of Mixed Methods in Public Health Research

The concept of incorporating both qualitative and quantitative methods in research design, data collection, analysis, and interpretation in a single study or series of studies is known as mixed methods research (MMR; Tashakkori & Creswell, 2007; Tashakkori & Teddlie, 2010). This methodology has been used in public health research to examine community compliance with control measures for other viral outbreaks such as the Ebola virus disease (Caleo et al., 2018). Risk perceptions related to the human immunodeficiency virus global pandemic have been examined using mixed methods to inform the design of appropriate preventative measures (Baidoobonso et al., 2013). Additionally, MMR is pertinent to research that examines individuals’ perceptions of practicing healthy behaviors (Plow et al., 2017). An advantage of using MMR is that it can lead to gaining a holistic understanding of health issues while keeping individuals’ needs and concerns at the center of the study (Rafei, 2020; Tashakkori & Teddlie, 2010). There are seven reasons for conducting an MMR study: compensation, complementarity, completeness, development, diversity, expansion, and corroboration/triangulation (Greene et al., 1989; Venkatesh et al., 2013). With the complexity of the pandemic and its threat to public health, studying the public’s response to COVID-19 demands the use of a holistic approach, thus the use of MMR.

COVID-19 and Social Distancing

COVID-19 was declared a pandemic by WHO on March 11, 2020 (WHO, 2020b), and it has since been the cause of more than 400,000 deaths in the United States alone as of February 5, 2021 (Centers for Disease Control and Prevention [CDC], 2021). It has been recognized that asymptomatic carriers can transmit COVID-19, leading to community spread (Bai et al., 2020; Ghinai, 2020). Until the end of 2020, there was no cure for COVID-19, and vaccine deployment had been slow (Mills & Salisbury, 2020). Therefore, the main strategy has been community mitigation measures to prevent the spread of the disease. One of these measures is social distancing, where individuals keep a certain amount of physical space between themselves and others who live outside their residence (CDC, 2020). The importance of these community mitigation measures to contain COVID-19 has been emphasized, and many governments enacted policies to enforce these tactics (Lewnard & Lo, 2020; Patrick et al., 2020; Wee et al., 2020; Wilder-Smith & Freedman, 2020), with the help of insights from the social sciences (e.g., National Academies of Sciences, Engineering, and Medicine, 2020). Despite the consensus among scientists and health professionals on the necessity for social distancing, evidence showed that there
were individuals not performing social distancing, thus increasing the likelihood of community transmission (Andrew, 2020; Cummins, 2020; Glanz et al., 2020).

Determining strategies to encourage preventive behaviors such as social distancing can be complicated, and multiple factors (e.g., ideologies, value systems, cultural backgrounds, socioeconomic status, and lifestyles) must be considered when determining solutions (National Academies of Sciences, Engineering, and Medicine, 2020). Like many public health issues, the COVID-19 pandemic may be deemed a “wicked problem,” meaning there is no definite, clear solution, and the consequences of both the problem and its solutions often involve difficult compromises (Rittel & Webber, 1973). In an emerging crisis, it is extremely difficult to identify the best available science for a given policy choice (Scheufele et al., 2020); thus, scholars usually argue that collecting community input and feedback is important for informing policy design and implementation (CDC, 2018; Chen, 2021; Dietz, 2013; Fishkin, 2018; Raisio, 2010). Therefore, understanding public attitudes is essential for recognizing how different populations perceive the risk dimensions of the wicked problem and how these populations weigh the costs and benefits of precautionary actions to cope with these risks. This understanding can lead to improved scientific communication and public policy responses.

Substantive Aims

The substantive aim of this study is to understand how individuals’ perceptions of persuasive messages relate to their self-reported social distancing behavior. Depending on their perception of COVID-19 and social distancing, the messages that individuals consider persuasive enough to change their behavior may vary. The Health Belief Model, one of the most widely applied theories of health behavior, suggests that individuals’ beliefs about the health issue and their perceived barriers and benefits of actions help predict health-related behavior (Rosenstock, 1974). The Health Belief Model posits that an individual’s likelihood of engaging in a health-related behavior is determined by six constructs: (1) perceived susceptibility (chances of contracting the disease), (2) perceived severity (how serious the disease is and its potential consequences), (3) perceived benefits (the efficacy of the health behavior), (4) perceived barriers (the cost of the health behavior), (5) cues to action (something that prompts a desire to make the change) and (6) self-efficacy (the ability to make a change). This suggests that persuasive messages are those that include information related to these constructs (e.g., discussing susceptibility, severity, benefits, etc.). The Health belief model also suggests that modifying factors such as diverse demographic variables may influence perceptions and thus influence health-related behavior (Sheeran & Abraham, 1996; O’Connor et al., 2014; Wright et al., 2012). The Health Belief Model has been used successfully to design many health behavior interventions and health education campaigns (Champion & Skinner, 2008). Using the Health Belief Model as the theoretical framework, this study examined the content of what individuals considered as persuasive messages, how this interpretation varied among social groups, and how it is associated with their practice of social distancing.

Methodological Aims

The methodological aim of this article is to illustrate how a mixed methods approach, combining computational with qualitative analysis methods, can be used to develop a more comprehensive understanding of public perception regarding an emergent and complex public health issue. Additionally, we examined the association between public perception and the practice of health-related behavior. In this article, “methods” refers to a set of systematic procedures that
fall under an approach (Onwuegbuzie & Frels, 2016). We used a mixed methods approach for two purposes (see Figure 1):

1. **Corroboration**—to assess the credibility and trustworthiness of the results gained from one method (Bryman, 2006). In this article, we used a computational method, structural topic modeling, to discover topics from the textual data systematically and objectively, while minimizing the typical biases found with human coders (Evans, 2014; Lewis et al., 2013). Structural topic modeling is useful for the descriptive analysis of the data by analyzing simple expressions of words and systematically inferring the prevalence and content of topics in each open-ended response. However, structural topic modeling is an unsupervised approach that captures the structure and characteristics of data (e.g., word usage patterns) without contextualizing them. As a result, a qualitative analysis of the data is conducted to validate the topics derived by the structural topic model. Unlike structural topic modeling, qualitative analysis requires close reading of the text, allowing researchers to consider the social-contextual features (e.g., the context) of the response to derive a deeper analysis of the data.

2. **Expansion**—to extend the breadth and range of inquiry (Greene et al., 1989). Some state-of-the-art topic modeling methods allow researchers to study the effect of different demographic characteristics or how public perception may vary by external variables (e.g., race and gender). Structural topic modeling is used to estimate the topic differences among social groups and analyze the relationship between the participants’ responses and their self-reported social distancing behavior.

**Method**

**Data Sources**

The study data were obtained from the “COVID-19 and Social Distancing” online survey conducted by a group of interdisciplinary researchers at a large research public university in the Midwest. The survey assessed individuals’ beliefs, attitudes, feelings, and self-reported behaviors related to social distancing and COVID-19. It was approved by the university’s institutional review board (IRB). On March 19, 2020, the online survey was launched, and from March 19 to March 24, a sharable link to the survey was posted on social media platforms to collect data ($N = 49,029$). Among the 49,029 individuals who participated in the survey, 38,329 completed the entire survey (~80% completion rate). Participants also had the option of not
answering questions, therefore gender or ethnicity demographic information was not always completed. The participants thus constitute a convenience sample that is not necessarily representative of the entire population (for details, see the online Supplementary Section I).

**Target Group.** The target group of this study was individuals who reported that they were currently not practicing social distancing, determined by their answer to the survey question “Do you currently practice social distancing?” The target group rated themselves from 1 to 3 on a 5-point Likert scale where 1 = *No, not at all* and 5 = *Yes, very much*. These participants were selected on the assumption that they are the ones who need to change their behaviors to limit the spread of COVID-19. Understanding their perceptions and attitudes will help design effective communication strategies (Campbell & Brauer, 2020; Lee & Kotler, 2019). Among the 38,329 valid participants in our convenience sample, 3,207 (8.37%) were target participants, and only 2,270 participants provided a response to the open-ended question. To demonstrate how MMR can be used for data analysis, our final analytic sample size is 2,270 (Figure 2). Table 1 presents a detailed description of our sample’s demographics. In this article, our analyses all focus on this target group.

**Measures**

The survey consisted of six parts: (1) self-reported social distancing behavior, (2) media habits, (3) effective messaging, (4) barriers to social distancing, (5) sociodemographic characteristics, and (6) participants’ open-ended comments. Our study focuses on the answers to the open-ended questions and the sociodemographic questions to examine how the content of persuasive messages varies among different social groups. We also examined how these open-ended responses were related to the primary outcome variable (self-reported social distancing behavior).

**Self-Reported Social Distancing Behavior.** The participants were asked, “Do you currently practice social distancing (in other words, do you deliberately increase the physical space between you
and other individuals to avoid spreading illness)?” The participants gave their responses on a 5-point Likert scale with endpoints labeled from “No, not at all” (1) to “Yes, very much” (5).

Sociodemographic Variables. The participants were asked to provide information about their age, gender identity, and racial/ethnic identity; the state they live in; the type of community they belong to; and their education level.

Open-Ended Responses. The participants were asked an open-ended question about the messages that could persuade them to perform social distancing and the factors that facilitate or prevent them from practicing social distancing. In this article, we focus on participants’ responses to the persuasion question “What could someone say to you that would make it more likely that you practice more social distancing?”

Data Analysis
To address our methodological aims, we first merged the open-ended survey responses with the quantitative data (i.e., demographic characteristics, self-reported social distancing behavior). Then, we analyzed the open-ended survey responses using a combination of computational methods and qualitative analysis methods (Figure 3, Boxes C-E). Figure 3 is an illustration of the analysis procedures used in this study.

Qualitative Analysis. We performed an inductive, exploratory thematic content analysis and analyzed the patterns of meaning that emerged from the data to identify central themes among the participants’ responses (Braun & Clarke, 2006; Thomas, 2006). To develop the codebook, we randomly selected a subsample of participants (~30% of the responses). We first used an open-coding procedure to generate as many in vivo codes as possible, which refers to words or short phrases that are used by participants (Saldaña, 2016). Then, a preliminary codebook was created inductively with categories. For example, we grouped the following codes under the category “Information About COVID-19”: infection risk, virus transmission, severity. Examples of categories are “Information About COVID-19,” “Information About Social Distancing,” and “Practical Concerns.” The codebook—which included detailed definitions, and inclusion and exclusion criteria—was subsequently reviewed, refined, and then tested on a new set of responses to ensure that saturation had been achieved. The codebook was finalized through analytical discussions about the data and the codes.

The coding team consisted of one master coder who coded all the responses and two trained research assistants who served as reliability coders. The reliability coders coded separate portions of the data set; as such, all responses had two separate coders—the master coder and a reliability coder to ensure consistency. Interrater reliability was high (calculated by the percentage of agreement between the coders, .95; Cohen’s Kappa = .78). Only the coding of the master coder was used in this study.

Table 1. Characteristics of the Target Group (N = 2,270).

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Descriptive statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female (84%) Male (16%)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>Caucasian (93%) Non-Caucasian (7%)</td>
</tr>
<tr>
<td>Education</td>
<td>Bachelor’s degree or higher (89%) (Some) high school education (11%)</td>
</tr>
<tr>
<td>Community</td>
<td>Large city (31%) Small city, rural area, or suburb (69%)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>$M = 37.43$, $SD = 13.76$, among those people who disclosed their age</td>
</tr>
</tbody>
</table>
A coder was used in the final analysis. The codes were not mutually exclusive as such a single response could receive multiple codes. For example, a participant could both mention the risk of infection and ask for information about how the virus is transmitted in their response.

**Computational Methods.** The computational method used in this study to analyze the open-ended text data is a type of unsupervised computational approach for text analysis—more specifically, the topic model. Topic models refer to “distributions over a vocabulary of words that represent semantically interpretable themes” (Roberts et al., 2019, p. 3). Thus, topic models estimate “the co-occurrence of words across documents” (Grimmer & Stewart, 2013, p. 283). There are various topic modeling methods, from the Latent Dirichlet Allocation (Blei et al., 2003) to the structural topic model used in this study. Unsupervised computational text analysis methods such as topic modeling are particularly useful when the research topic is understudied or when researchers do not have predetermined categories—for example, in an unprecedented global pandemic (see Grimmer & Stewart, 2013, for a detailed discussion).

Topic modeling methods have been used in MMR previously. For instance, when comparing textual analyses using topic modeling and grounded theory, prior mixed methods studies found surprisingly similar results despite the different epistemology and relationships to theories (Baumer et al., 2017; for a review, see Isoaho et al., 2021). Among the computational methods that identify topics, we chose the structural topic model to analyze the participants’ open-ended responses (e.g., Chen et al., 2020; Roberts et al., 2014). This approach is found to be the most effective among various unsupervised methods in identifying topics and frames against manual coding, especially for narrow-scope data sets (Nicholls & Culpepper, 2021), and it provides more accurate estimation than other automated approaches, such as the Latent Dirichlet Allocation (Roberts et al., 2014).

Structural topic modeling has several methodological advantages over qualitative analysis methods. Similar to qualitative analysis methods, the structural topic model also thematically organizes the corpus of data to uncover prevalent topics that might be difficult to discover from reading the raw data (Günther & Quandt, 2016). However, structural topic modeling “makes analyzing open-ended responses easier, more revealing, and capable of being used to estimate

![Figure 3. An illustration of the analysis procedures used in the study.](image-url)
treatment effect’’ (Roberts et al., 2014, p. 1064). For instance, rather than counting the frequency of phrases in a corpus, the structural topic model generates a variety of keywords and example documents (e.g., text content) that are highly associated with each topic for researchers to deliberate about the meaning of each topic. In addition to identifying prevalent topics that frequently appear in open-ended responses, the structural topic model allows researchers to examine the relationship between documental-level metadata and the explored prevalent topics (Roberts et al., 2019). For example, this study incorporated participants’ gender and race as metadata in the models. Given that our target sample was predominantly Caucasian (93%) and women (84%), controlling for these two variables in the model can reduce their confounding effects on the topic prevalence. In addition, we can also estimate how race and gender are associated with the participants’ preferred persuasive message about social distancing.

The stm R Package for structural topic models (Version 1.3.5) was applied to identify the prevalent topics (Roberts et al., 2019). We first used the “textProcessor” function to clean the data by removing punctuations and stopwords and simplifying words with stemming. Then, the “stm” function generated the 10 most prevalent topics. As shown in the diagnostic values plot with the number of topics ($K$) from 5 to 20 (for details, see the online Supplementary Section II), the optimal choice for the topic number should be $K = 10$. When $K = 10$, the held-out likelihood is the largest, indicating the topics’ best performance to generalize the documents (Wallach et al., 2009), and the semantic coherence remains high, indicating a good performance of interpretability. Due to the short length of an open-ended response, generated topics became hardly interpretable for $K$ larger than 15, and merely 5 topics did not present as a full epitome of all the varied responses.

Each of the 10 topics includes five keywords that had high frequency and exclusiveness. Figure 4 presents the distribution of the top 10 prevalent topics summarizing our target participants’ responses to which persuasive messages are effective enough to persuade them to adopt social distancing behavior. To interpret the computational results, several researchers, including those who conducted the computational topic modeling and a qualitative researcher, discussed how they interpreted the keywords of each topic and which topic generated by the computer model is meaningful.

Next, we integrated the structural topic model results with regression analysis to investigate whether mentions of certain topics in the open-ended responses relate to the degree of self-
reported social distancing. First, we transformed the open-ended responses into categorical variables so that we could analyze them using standard statistical techniques. The “theta” output from the “stm” function allowed us to categorize each open-ended response into a topic for which the response has the largest likelihood. Therefore, we had 10 dichotomous variables for each response, indicating whether it identifies with a certain topic (coded as “1”) or not (coded as “0”). As no answers were categorized into Topic 8 with this method and because Topics 2 and 6 are not meaningful based on our interpretation from the structural topic model results (discussed in detail in the section “Prevalent Computational Topics” below), we excluded Topics 2, 6, and 8 in the following analyses. To regress the model on self-reported social distancing among the target participants, we controlled for demographic variables—age, gender, race, education level, and community type (see Table 1)—as well as political and economic ideologies using a 5-point liberal–conservative Likert-scale.

Results

Persuasive Messages

Qualitative Themes and Subthemes. We identified three major themes and eight subthemes that represented the types of messages participants wanted to hear in order to be persuaded to perform social distancing: (1) communicating the threat of COVID-19 (e.g., describing the risk, its severity, and its transmission), (2) addressing barriers to actions (e.g., issues related to feasibility, social norms, and strategies), and (3) promoting the benefits of social distancing behavior (e.g., explaining social distancing and its effectiveness). Figure 5 shows a hierarchy of themes and subthemes.

Communicating the Threat of COVID-19. Many participants mentioned that emphasizing the threat of COVID-19 would increase their social distancing behavior. Within the broad category of threat, we found that many individuals focused on messages that describe the risk of infection or the severity of COVID-19. For example, one participant wrote, “That my area has more cases and we’re at higher risk of community spread, or that I am at higher risk of complications if I got it.” Other participants wanted messaging around the transmission of COVID-19. One such participant commented that sharing information such as “COVID can be spread by
asymptomatic individuals and can survive on inorganic surfaces” would make a person more likely to increase social distancing.

To better understand risk perceptions, we further explored who the risk is supposed to affect (risk target). There are three risk targets specified: (1) risk affecting the participants themselves (personal), (2) risk to family members (familial), or (3) a general risk of infection (general). Among those who discussed the risk of infection, most participants highlighted personal risk. This shows that participants believed that messages that reflect an increased personal risk of contracting COVID-19, in addition to messages that emphasize the severity of the infected state, would increase social distancing behaviors.

**Addressing Barriers to Actions.** Participants tended to describe various barriers to social distancing. For example, many individuals discussed the feasibility of social distancing, such as their occupation and/or financial concerns making it harder for them to practice social distancing. One participant said,

> You will be paid the full amount on all the contracts you had that have been annulled due to the virus. (I need money, and my work is the one place where I don’t really practice social distancing effectively currently.)

Other participants cited social factors and existing social norms as a barrier to social distancing. Some discussed that they recognize that social distancing is important but that due to the actions of others, they do not feel that their efforts would make a difference. One respondent candidly said,

> It’s more in actions. I know it’s a good thing to do and I ought do it; it’s just very hard to practice when social distancing requires a combined effort of those around you and those you live with. If one person in my house doesn’t believe in it and continues meeting with others, I feel as though my contributions do nothing.

Along with describing barriers to social distancing, participants wanted strategies to circumvent these barriers. One response read,

> Give me more ideas for what to do indoors. I feel like there isn’t anything to do if I stay home all day, and it’s harder for me to be productive, so I would really appreciate tips for working/studying from home.

Thus, addressing the barriers preventing social distancing, including financial and social barriers, in addition to providing strategies for managing barriers, may help improve social distancing behaviors.

**Promoting the Benefits of Social Distancing.** Some individuals discussed how increased messaging around promoting the benefits of social distancing would change their behaviors. Participants were interested in messages about the actual effectiveness of social distancing behaviors. A crucial point from this theme is that some individuals do not understand why social distancing works and would need more concrete evidence that it is effective as a preventative tool. For example, one participant wrote, “Explain the exponential increase of cases when social distancing is in place versus when is not, maybe through graphics.” Participants also expressed the need for more clarity about social distancing behavior due to a lack of understanding of what qualifies and does not qualify as social distancing. One participant asked, “What activities you can and can’t do? Can you have a friend over? Can I pick up a pizza?” Messaging that
addresses the positive impact of social distancing and clarifies the specifics of social distancing may help increase social distancing behaviors.

Subthemes by Race and Gender. We further explored the prevalence of these subthemes by key demographic groups (see Table 2). In terms of gender, females mentioned the feasibility of social distancing more often than males (38% vs. 24%). Males were more likely to mention the severity of COVID-19 (20% vs. 14%) and social factors that they considered as barriers to social distancing (7% vs. 5%). Both mentioned risk in an almost equal proportion (36% of female participants and 36% of male participants). When examining responses based on self-reported race, those who identified as Caucasian were more likely to mention the feasibility of social distancing (33%). Participants who identified as non-Caucasian mentioned themes regarding the threat of COVID-19 (38%).

Prevalent computational topics. Table 3 provides the keywords and researchers’ interpretations of each topic (for example responses, see the online Supplementary Section III). The interpretation of each topic includes the computational researchers’ interpretation of the keywords and the interpretation from a qualitative researcher. Topics 2 and 6 were considered not meaningful, because the highly associated examples of these topics did not provide a consistent topic.

We observed four major themes from the eight meaningful topics that target participants related to the persuasive social distancing messages they wanted to hear. The first theme referred to the risks of getting and spreading COVID-19. Specifically, these risks include how fast the virus could spread (Topic 4), how much harm the virus would cause to their family members (Topic 5), and how likely one would get COVID-19 by being exposed to public spaces (Topic 9). The second theme was related to the severity of the current COVID-19 situation, such as the number of people who are asymptomatic carriers (Topic 1) and how severe it will be for pregnant women (Topic 3). The third persuasive theme centered around the feasibility of practicing social distancing. Particularly, these feasibility issues include the difficulty for one to practice social distancing while working (Topic 7) and to maintain their current jobs and essentials (Topic 10). The fourth theme related to the effectiveness of performing social

<table>
<thead>
<tr>
<th>Themes</th>
<th>Female (%; n = 1,561)</th>
<th>Male (%; n = 293)</th>
<th>Caucasian (%; n = 1,901)</th>
<th>Non-Caucasian (%; n = 221)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicating the threat of COVID-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk</td>
<td>35.55</td>
<td>35.84</td>
<td>30.51</td>
<td>37.56</td>
</tr>
<tr>
<td>Severity</td>
<td>13.71</td>
<td>19.45</td>
<td>12.52</td>
<td>16.74</td>
</tr>
<tr>
<td>Transmission</td>
<td>6.53</td>
<td>7.85</td>
<td>5.58</td>
<td>9.50</td>
</tr>
<tr>
<td>Promoting the benefits of social</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distancing behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity and specificity</td>
<td>2.63</td>
<td>2.73</td>
<td>2.21</td>
<td>3.17</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>12.24</td>
<td>12.63</td>
<td>10.84</td>
<td>11.31</td>
</tr>
<tr>
<td>Addressing barriers to action</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasibility</td>
<td>38.24</td>
<td>23.89</td>
<td>32.56</td>
<td>23.53</td>
</tr>
<tr>
<td>Strategies</td>
<td>4.29</td>
<td>3.75</td>
<td>3.79</td>
<td>2.71</td>
</tr>
<tr>
<td>Social factors</td>
<td>4.80</td>
<td>7.17</td>
<td>4.63</td>
<td>4.07</td>
</tr>
</tbody>
</table>

Note. N is the number of respondents who indicated that they currently did not engage in the recommended social distancing and who provided a response for this question. Each response could have multiple codes, so the total percentage does not add up to 100.
Table 3. Comparison of Researchers' Interpretations.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Keywords</th>
<th>Interpretation (computational)</th>
<th>Fit with the qualitative interpretation?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>think, covid, serious, number, see</td>
<td>Information about COVID-19 infected symptoms</td>
<td>Yes, the severity of COVID-19</td>
</tr>
<tr>
<td>2</td>
<td>like, make, week, love, live</td>
<td>Foresights</td>
<td>No, not meaningful</td>
</tr>
<tr>
<td>3</td>
<td>get, die, sick, feel, abl</td>
<td>Information about the detailed severity of COVID-19 infection</td>
<td>Yes, the severity of COVID-19</td>
</tr>
<tr>
<td>4</td>
<td>people, virus, someone, can, person</td>
<td>Information about the susceptibility of COVID-19 infection</td>
<td>Yes, the risk of getting/spreading COVID-19</td>
</tr>
<tr>
<td>5</td>
<td>will, noth, famili, risk, much</td>
<td>Information about family risks of COVID-19 infection</td>
<td>Yes, the risk and severity of COVID-19 among family members</td>
</tr>
<tr>
<td>6</td>
<td>know, alreadi, need, one, spread</td>
<td>Suggestions about effective communication strategies</td>
<td>No, not meaningful</td>
</tr>
<tr>
<td>7</td>
<td>distanc, social, practice, job, time</td>
<td>Difficulties of social distancing while working</td>
<td>Yes, feasibility at work</td>
</tr>
<tr>
<td>8</td>
<td>infect, covid-, contact, current, import</td>
<td>The importance of social distancing</td>
<td>Yes, the effectiveness and importance of social distancing</td>
</tr>
<tr>
<td>9</td>
<td>case, test, possil, sure, other</td>
<td>Concerns about possible COVID-19 infections from others</td>
<td>Yes, the risk of getting COVID-19</td>
</tr>
<tr>
<td>10</td>
<td>work, home, dont, stay, close</td>
<td>Willingness to practice social distancing if work modality changes</td>
<td>Yes, the feasibility of practicing social distancing</td>
</tr>
</tbody>
</table>
distancing to help people return to normal life activities earlier, or to go to bars and restaurants (Topic 8).

**Prevalent topics by gender and race.** In addition to examining the prevalent persuasive topics for our target participants, we used computational methods to investigate how various social groups (i.e., gender and race) differed in their answers to the persuasion question. For instance, Figure 6 presents how different racial groups differed in their answers to what messages can convince them to perform social distancing. The numbers on the $x$-axis reflect the difference in the average topic proportion between the chosen subgroups (e.g., non-Caucasian vs. Caucasian). For example, the black dot on the left side of Figure 6 indicates that Topic 5 was more prevalent in the non-Caucasian group and was less likely to be discussed by Caucasian participants. The racial difference in mentioning Topic 5 is 2%, significant at the .05 level. Topic 10 was 4% more likely to be mentioned by the Caucasian group than by the non-Caucasian group, also significant at the .05 level.

Examining the keywords associated with these two topics and the example responses participants wrote, we found that being identified as Caucasian was associated with higher demand for messages telling them what they could do while staying at home. For instance, one participant pointed out,

> Give me more ideas for what to do indoors. I feel like there isn’t anything to do if I stay home all day, and it’s harder for me to be productive, so I would really appreciate tips for working/studying from home.

Alternatively, being in the non-Caucasian group was associated with greater willingness to hear risk communication about the pandemic. For instance, a typical response was to ask for

> a better statement explaining the risk categories for this virus: for ages (young and old) and other conditions that suppress the immune system. The messages I have seen are too generic and have heard individuals rationalize the warnings and it is not specific enough to apply to them.

---

**Figure 6.** The difference in average topic proportion between different race groups ($n = 2,270$).

Note. We set the “Caucasian participants” group as the reference, and a negative coefficient for topic prevalence would mean the topic’s appearance less in the Caucasian and more in the non-Caucasian group. The black dots in the figure indicate the coefficient of the relationship between race and the likelihood of mentioning a certain topic, and the horizontal lines represent the confidence intervals.
Figure 7 presents how different gender groups differ in their answers to the persuasive questions. We found that Topic 7 is 3% more likely and Topic 10 was 7% more likely to be mentioned by participants who identified as women, both significant at the .001 level. Male participants were 3% more likely to mention Topics 1 and 5 and 1% more likely to discuss Topic 8, also all with a significance level at .001. From inspecting the topic keywords, we could conclude that women participants were more concerned about their jobs and working at home. Men participants put more emphasis on the risk of the virus and requested information about their family and public health.

Comparing the computational and qualitative results. To demonstrate how computational and qualitative approaches can corroborate each other, we compared the results from each approach that analyzed the most prevalent topics and themes that appeared in the responses to the open-ended survey question asking what persuasive messages could motivate individuals to change their behaviors (see Figure 8). Overall, the two approaches yielded similar results, showing that preferred persuasive messages are those that highlight the threat of COVID-19 (the risk, transmission, and severity of COVID-19: computational = 52%, qualitative = 48%) and address barriers to social distancing (work issues: computational = 48%, qualitative = 38%). The percentage of responses assigned to each theme varies because the computational method assigned responses to the topic that has the highest maximum likelihood whereas qualitative analysis allows a response to be coded under multiple themes. Hence, the qualitative results not only helped us identify meaningful computational topics but also provided a more nuanced understanding of the data by revealing three new insights: (1) strategies to deal with barriers, (2) social concerns related to practicing social distancing, and (3) the need for more information about social distancing (effectiveness and specificity). Furthermore, both approaches also showed that public perceptions vary by demographic characteristics. Both types of results found that members of non-Caucasian groups and males tended to mention the threat of COVID-19 and members of Caucasian groups and females tended to mention the feasibility of social distancing in their responses. Besides, our computational findings revealed that members of the Caucasian group were more likely to demand persuasive messages to tell them what they could do when staying at home.
In this section, we investigated the relationship between social distancing behavior and responses to the persuasion question among the target group, by integrating structural topic modeling and regression analysis.

Table 4 presents a parallel bivariate coefficient table that shows each topic identifying variables regressed with social distancing behavior while controlling for other variables. Our results showed significant associations between answers to the persuasive message question and self-reported social distancing behavior among the target participants. We found that mentions of Topic 7 were more likely to be associated with better social distancing practices among target participants ($b = .09, p < .01$), while references to Topic 10 were associated with a lower level of practice of social distancing ($b = -2.09, p < .001$). Though both Topic 7 and Topic 10 were about working issues (see examples in Table 3), they are expressed differently. Topic 7 described participants’ difficulties with social distancing while they have work to do. For example, some participants claimed that they have practiced social distancing well when not at work and ascribed their improper practice of social distancing to their job requirements. Participants answering with Topic 10, however, were demanding conditional information that would convince them to conduct social distancing. The answer “If my employer let me work from home” is a prevalent wording in target participants’ responses featuring Topic 10. That is, they would not be likely to practice more social distancing unless their work modality changes. Comparing Topics 7 and 10, we see that while Topic 7 discussed the infeasibility of social distancing only at work, Topic 10 put working as a major issue that impedes participants from practicing social distancing, without referring to whether they followed social distancing the rest of the time.

Despite the statistically significant association between participants’ responses to the persuasive messages question and their self-reported social distancing behavior, the magnitude of the association is not large. This small correlation could be because our participants had not yet heard these persuasive messages in real life. The open-ended survey questions asked what messages they wanted to hear. After getting enough exposure to these persuasive messages in real life, they might adjust their behavior accordingly.

**Discussion**

In this article, we demonstrate how computational methods can be integrated with qualitative analysis methods to mine public opinion data sets in order to understand public perceptions of...
persuasive messages that could encourage individuals to practice more social distancing during a pandemic such as COVID-19. By drawing on the strengths of each method, a mixed methods approach can generate a more comprehensive understanding of how people weigh the perceived benefits of and barriers to social distancing, what their priorities are, and what other concerns they may have. This knowledge is especially important for public health communication because effective communication needs to address individuals’ concerns and to incorporate public values and concerns in the message design.

**Contributions to the Health Communication Literature**

Both computational and qualitative results suggest that the perception of threat to health (perceived risk and severity of infection) is one of the factors that influence health-related behaviors (N. T. Brewer et al., 2007; Ferrer & Klein, 2015). To practice social distancing, participants recommended the use of stronger messaging detailing the risk, severity, and transmission of the virus to aid with the individual level barriers to social distancing. There are several barriers to social distancing from an individual and organizational level. Barriers to social distancing can be at the individual level such as the individual’s perception of COVID-19 related messages, or at the organizational level such as being required to work in person (Coroiu et al., 2020). Additionally, individuals are interested in learning more about the benefits of social distancing. Many individuals also reported the organizational-level barrier of work constraints, which impacts an individual’s financial context.

Echoing the health belief model, our findings further revealed that what individuals consider as persuasive messages that could potentially change their behavior varied by gender and race. For instance, Caucasian and female participants tended to mention the feasibility of social distancing and working from home as a key challenge to performing social distancing. This difference could be due to the types of jobs they have and other family obligations (e.g., child care). These findings, together, suggest that even when facing a public health crisis, individuals’ perceptions that can influence their behavior are multidimensional and vary from person to person (Kreuter et al., 2000; Schmid et al., 2008). For health communicators, it is critical to understand the nuances of public perception and to be attentive to the unique needs, characteristics, and concerns of different individuals in persuading people to perform precautionary actions to fight COVID-19.

Examining the association between what messages individuals wanted to hear to be persuaded to practice social distancing and their self-reported social distancing, we found that there is a statistically significant association between hearing certain types of persuasive messages and self-reported social distancing behavior. Individuals who mentioned that their current jobs did not allow them to do a lot of social distancing (e.g., urgent-care provider) were more likely
to practice social distancing; individuals who wanted to hear persuasive messages about how working modality could change were less likely to practice social distancing. These findings suggest that individuals’ social distancing behavior depends on the feasibility of doing their jobs while still practicing social distancing. These findings echo literature in risk communication that stresses that individuals’ health choice behavior is highly contextual (Alaszewski, 2005; Duff, 2003).

Aligning with previous health behavior studies, these findings also suggest that perceived barriers could be the most powerful predictor of individuals’ social distancing behavior among all other constructs in the health belief model (Champion & Skinner, 2008; Janz & Becker, 1984). Therefore, an implication for health communicators who are designing message campaigns is that they need to assess the feasibility of performing social distancing among different population segments (e.g., health providers, grocery workers) and tailor their messages to populations differing in race, gender, occupation, and value system (Campbell & Brauer, 2020; Scheufele et al., 2020; CDC, 2018). These findings also suggest that adjustments in working policies, such as changes of working modality, need to come hand in hand with COVID-19 policies to support individuals in practicing social distancing.

**Contributions to the Mixed Methods Literature**

This present article demonstrates how a mixed methods approach can be used in public health research using both open-ended and closed-ended survey data. Integrating qualitative and quantitative data together allows researchers to achieve a more holistic understanding of a phenomenon. As Fetters and Freshwater (2015) stressed, addressing the “integration challenge” is a difficult task facing the mixed methods community. The goal of integration is such that when qualitative and quantitative strands are integrated, the value of the combination should be greater than the sum of the individual components (Fetters & Freshwater, 2015). More specifically, Fetters and Molina-Azorin (2017) suggest that integration can be achieved by intentionally integrating the qualitative and quantitative strands of the study through 15 dimensions of the MMR integration trilogy (philosophical, methodological, and methods). This article primarily focuses on integration in the “data analysis” and “interpretation” dimensions (Fetters & Molina-Azorin, 2017, p. 294).

**Integration in the Analysis Dimension.** While integrating qualitative and quantitative data during data collection has been addressed in the literature (Bazeley, 2012; J. Brewer & Hunter, 2006), practical techniques for merging qualitative and quantitative data during analysis remains limited (Fetters et al., 2013; Guetterman et al., 2015; Vogl, 2019). To illustrate how to implement these practical techniques of integration, we demonstrated how to combine qualitative and computational methods (1) to conduct independent analyses, followed by comparing both findings to identify what themes/topics are similar and different between the two methods to reach expansion and corroboration in MMR, and (2) to transform qualitative data into quantitative outcomes, followed by merging with other quantitative data (Fetters et al., 2013). Integrating both computational and qualitative methods in analysis allowed us to capitalize on the unique strengths of each method to gain a more nuanced understanding of the complexity of different individuals’ perceptions of COVID-19 risk and social distancing.

There are several advantages to using computational methods in combination with qualitative analysis. First, computational analysis can help researchers gain a general overview of the content of text data without being overwhelmed by the volume of data. The state-of-the-art topic modeling method used in this article, structural topic modeling, demonstrates its usefulness to social scientists in transforming text data into interpretable patterns by identifying key
topics and related keywords. Moreover, when the structural topic model generates the prevalent topics for a text data set, it also takes into consideration other document-level metadata that could influence how individuals express their opinions (e.g., gender, race). However, while topic modeling is good for descriptive analyses, it cannot replace the interpretative capacity of qualitative analysis. In fact, qualitative analysis is needed to make sense of the keywords identified in each topic in order to understand what theme a topic reflects and which topics belong to the same theme. Furthermore, unsupervised topic modeling could produce nonsensical topics that are not necessarily accurate or meaningful to the written text. To assess the credibility and trustworthiness of the prevalent topics, a qualitative analysis is conducted within each prevalent topic to validate that they are truly substantive topics and to corroborate to what extent the prevalent topics line up with researchers’ hand coding.

Moreover, structural topic modeling also transforms text data into quantifiable data. For instance, it allows researchers to know for each answer a respondent wrote, the maximal likelihood that this answer belongs to a topic. Because of this quantification, structural topic modeling allows researchers to conduct additional quantitative analyses with other survey data. As the present article demonstrates, we were able to extract the topics using the structural topic model and then performed regression analysis between public perception and self-reported social distancing behavior to find out what public perception is correlated with a higher likelihood of performing social distancing.

Integration in the Interpretation Dimension. This article presents both qualitative and computational findings through narrative in the results section using the “weaving approach” (Fetters et al., 2013). This approach involves connecting results to each other and weaving back and forth based on the thematic relevance. The assessment of “fit” of data integration, which refers to the coherence of both qualitative and quantitative findings, resulted in confirmation, complementarity, and expansion (Fetters et al., 2013). The qualitative and computational results on persuasive messages confirmed each other, allowing us to identify topics that are truly substantive and meaningful. Both results were compared using a comparison joint display (see Table 3). Since both methods provided similar results, the findings have greater credibility. The integration also expanded our understanding by revealing the complementary relationship between the qualitative and computational findings, where qualitative findings provide a more nuanced understanding of the content of persuasive messages and computational findings speak to the strengths of the associations between persuasive messages and demographic characteristics and self-reported health behavior.

Therefore, our research contributes to the field of MMR by (1) showing the credibility of computational methods in analyzing text data (corroboration) and (2) demonstrating how they can be used together with qualitative analysis to extend the breadth of inquiry (expansion). Our findings suggest that structural topic modeling is a comparable approach to an exploratory, descriptive qualitative analysis. There are a variety of computational text analysis methods besides the structural topic model that we used in this study, and readers can refer to Grimmer and Stewart (2013)’s article for a more detailed discussion of the different methods and their application for discovery, measurement, and inference in social science (Grimmer et al., 2021). We also want to call attention to a growing body of work that discusses how to integrate computational methods with qualitative methods across disciplines, such as informational science (Baumer et al., 2017), policy studies (Isoaho et al., 2021), and communication research (Ophir et al., 2020).
Limitations

Our convenience sampling did not provide a representative sample of the U.S. population, and females and Caucasians were oversampled. We made a necessary trade-off in the sampling method here to better serve the rapid response that was needed from the communication task force at our institution to circulate messages about the outbreak in the early stages of the pandemic in the spring of 2020. Acknowledging this limitation in the sample representativeness, we controlled for gender and race when we used structural topic modeling to generate the most prevalent topics to control for the influence of gender and race. It is important to keep in mind that the primary goal of the present article is not to explore what the average American thought about effective persuasions to perform social distancing. Instead, we utilized this data set to demonstrate how both qualitative and computational methods can be used to analyze data.

A limitation of the qualitative method in our study is the open-ended nature of the survey responses. Participants tended to provide short responses, which made it challenging for the coders to interpret the data. In terms of computational analysis, we used one specific method of automated text analysis, structural topic modeling, to examine the different themes in our open-ended question. There are other computational text analysis methods, including supervised, unsupervised, or a mix of the two, for analyzing text data quantitatively (for a detailed discussion, see figure 1 in Grimmer & Stewart, 2013). Future research could further examine how results from various computational methods corroborate and expand the qualitative method.

In this article, we focused on integrating mixed methods in the analysis and interpretation dimensions (Fetters & Molina-Azorin, 2017). It will be fruitful for future research to also consider integration in other dimensions, such as theoretical, research design, and sampling. For example, researchers can use convergent, sequential exploratory, sequential explanatory, or multistage designs to collect both qualitative and quantitative data (e.g., Leech & Onwuegbuzie, 2009; Teddlie & Tashakkori, 2009). In the context of COVID-19, we found that individuals will be more likely to be persuaded when they hear messages that highlight the consequences of inaction and address barriers to make social distancing feasible (e.g., work issues). This evidence can inform future experimental designs regarding what messages are more likely to increase individuals’ social distancing behavior. In a way, evidence from the qualitative data can help health communicators design and study the effects of message campaign strategies. The quantitative data (e.g., self-reported behavior) obtained from the message campaign experiment can further inform research on why some individuals choose to perform social distancing while others do not.

Conclusion

By integrating both computational and qualitative approaches, this article contributes to the emerging public health literature on public perceptions of effective, persuasive messages to social distance. It provides an in-depth investigation of persuasive messages that could potentially change individuals’ health behavior. Methodologically, this article supplies an example of mixed methods that can be used to gain a more comprehensive understanding of a complex public health issue. Because quantitative data assessed individuals’ health behaviors and qualitative data explored individuals’ concerns, using multiple methods allows researchers to extend the range of inquiry to examine the relationship between the two. Mixed methods research has an immense potential to enrich the analysis and findings of public health research addition to increasing its rigor.
Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs
Pauline Ho https://orcid.org/0000-0002-6010-4173
Luye Bao https://orcid.org/0000-0003-3355-6630
Angela Ai https://orcid.org/0000-0002-6064-4700

Supplemental Material
Supplemental material for this article is available online.

References


