Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Usability of state public health department websites for communication during a pandemic: A heuristic evaluation

Amirmasoud Momenipour\textsuperscript{a}, Salvador Rojas-Murillo\textsuperscript{b}, Brandon Murphy\textsuperscript{c}, Priyadarshini Pennathur\textsuperscript{c}, Arunkumar Pennathur\textsuperscript{c,\ast}

\textsuperscript{a} Engineering Management, Rose-Hulman Institute of Technology, USA
\textsuperscript{b} Industrial Engineering, Hofstra University, USA
\textsuperscript{c} Industrial and Systems Engineering, University of Iowa, USA

ARTICLE INFO

Keywords:
Usability
Pandemic
Heuristic evaluation
Public health websites
COVID-19
Emergency communication

ABSTRACT

Study aims were to investigate how usable COVID-19 dedicated state public health websites in the US were, and whether case counts in different geographical regions in the US were related to website usability. 16 state websites representing the 2 highest and the 2 lowest case count states in each region were selected. Five experts used a heuristic evaluation procedure to independently rate all 16 websites on a severity scale of 0–4. Usability criteria published by the US Department of Health and Human Services and criteria on risk communication and data dashboards were used. Analyses involved cross tabulation of usability criteria with case counts, comparison of usability scores using Mood’s median tests, test of differences in average usability scores using ANOVA and post-hoc tests, and identification of correlations between case counts and usability scores. Results from the Mood’s median test showed that the median usability scores for the states were significantly different from each other at the 5% level of significance ($df = 15$, chi-square $= 38.40$, $p = 0.001$). ANOVA showed statistically significant differences between the mean usability scores for the states at the 5% level of significance ($F = 6.33$, $p < 0.05$). Although not statistically significant, results from a correlation analysis between case count and usability scores showed a negative correlation ($r = -0.209$, $p = 0.437$) indicating that the higher the case count, the better the usability score. Overall, the websites fared well on usability, but many websites were used as an information and data repository. These websites must communicate infection risk better.

Relevance to Industry: The study applies to public health agency websites that communicate essential information during a pandemic.

1. Introduction

COVID-19 continues to spread worldwide and has been causing widespread deaths and economic hardships. Amidst this pandemic, to communicate accurate, credible, life-saving information to the public in a timely manner, government entities at all levels, including county, state and federal health departments and agencies have created COVID-dedicated websites. These websites have included general information on COVID symptoms, guidelines for preventing its spread, dashboards on case counts, information on testing, and recommendations for safely reopening businesses and educational institutions. The public, healthcare professionals including frontline workers and contact tracers, and more recently, decision makers and workers in businesses and educational institutions, have come to rely on these websites as a primary daily source of up-to-date information for making life-saving decisions. It is critical, therefore, that these websites be usable.

2. Background and related work

2.1. Usability and its evaluation

Usability in general, and website usability in specific, are not new considerations in design. The most prominent usability metrics have been derived from the ten usability heuristics for user interface design developed by Nielsen (Nielsen, 1994, 2005). Nielsen’s heuristics include ensuring that the status of a system is visible to users, matching the system and the real world, enabling user control and freedom, maintaining consistency, and adhering to standards, preventing errors,

\footnotesize\textsuperscript{\ast} Corresponding author.
E-mail address: arunkumar-pennathur@uiowa.edu (A. Pennathur).

https://doi.org/10.1016/j.ergon.2021.103216

Received 8 February 2021; Received in revised form 2 September 2021; Accepted 9 September 2021

Available online 20 September 2021

0169-8141/© 2021 Elsevier B.V. All rights reserved.
designing for recognition rather than recall, making the system flexible and efficient in use, making the design aesthetic and minimalist, helping users recognize, diagnose and recover from error, and providing help and documentation to users. The ISO 9241–11:2018 standard defines usability as the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use (International Standards Organization, 2018).

Guidelines such as Nielsen’s heuristics and the ISO standards provide a general framework with which to evaluate usability. But, researchers suggest tailoring and adapting these general heuristics and guidelines to specific contexts. Using context specific guidelines when evaluating interactive systems have been shown to result in an increase in the number of usability problems uncovered when compared to evaluations that only use generic heuristics (Adacan and Erbug, 2009; Jimenez et al., 2012). Further, usability evaluations need to be domain specific – usability criteria in consumer products design can significantly differ from usability criteria for the transportation industry, for example (Mator et al., 2020). Additionally, usability evaluations that incorporate a combination of methods such as heuristic evaluations and user testing when feasible, allow for discovery of more usability problems within a shorter duration (Solano et al., 2016).

2.2. Usability of websites

Scholars have developed many guidelines for web developers to use to make websites usable – notable among these are guidelines Nielsen originally developed, all the other guidelines based on Nielsen’s original work (Nielsen, 1994, 2000, 2005; Nielsen and Tahir, 2001), and work from other researchers (Bevan, 2001, 2005; Cosgrove, 2018; Krug, 2009). Despite the wide availability of usability guidelines and principles, studies show that even website designers and developers trained in usability find it difficult to readily apply the guidelines and principles either because they perceive that a significant effect may be required on their part when following the principles, or because they do not have the technology to readily code website elements that will make the sites usable. These barriers affect the degree of compliance of website developers to usability guidelines and standards (Alonso-Virgos et al., 2019). Additionally, studies (Hertzum, 2010) show that people in general, and web developers, by extension, may perceive the goals of usability engineering and its meaning in different ways. Some may think usability is about making websites easy for all to use (the universality principle); others may view usability as situational and may focus their efforts on the quality of use of the website for different tasks and users; some others may design a website based on a user’s direct perception when they interact with the site; yet others may focus only on the hedonic aspects of the website by prioritizing joy of use rather than ease of use. Usability metrics may also be based on organizational goals (e.g., to allow people to collaboratively use and share information via a website) (Ludwig et al., 2015, 2017) and may have cultural undertones based on user background. Hence, when designing websites, the interface aspects, the color, and the organization of the content, to name a few, may not be the only considerations for a developer. There is consensus that web developers and designers can stay informed of the most recent guidelines and standards in usability by building successful websites and gaining real world experience, and by solving website development problems by participating in a community of practitioners (Sayyad Abdi et al., 2016).

2.3. Usability of health communication websites

Our focus in this paper relates to one specific aspect of website usability - namely what factors impact the usability of websites created for health communication during emergencies such as a pandemic. The growth of the internet, and the concomitant growth in people’s online presence and their use of social media have enabled federal, state, and local government agencies to leverage internet-based communication tools to manage emergency planning, and to communicate about public health emergencies. While the main challenges in designing these emergency communication websites initially focused on managing the amount of trustworthy information to share with the public (Green III, 2001), more recent studies evaluating government websites for emergency management and communication have focused on usability and accessibility (Youngblood and Youngblood, 2018). Venkatesh et al. (2017) examined the usability of the healthcare.gov website using the US Department of Health and Human Services usability evaluation guidelines (Shneiderman et al., 2006). The healthcare.gov website was designed to provide health insurance access to the public. From their survey of 374 citizen users, they concluded that the website rated poorly on usability. They recommended that designers of government health websites consider the educational backgrounds of users, the access of citizen users to technology, literacy of the users, and the cultural diversity of the citizen users. Youngblood and Youngblood (2018), in interpreting the work of many other scholars in this area (Grove, 2013; Fogg et al., 2003; Huang et al., 2009; Scott, 2005) have echoed the idea that website usability is a critical design factor that significantly impacts the degree to which a government website can be trusted as having credible, useful, and transparent content, thereby enabling people to interact with the website to transact their business.

Digital inclusion in relation to emergency communication is another important research area scholars have investigated, particularly with the growth in the smartphone availability and use among all age groups and people in all socioeconomic and demographic strata (Pew Research Center, 2017). Furthermore, studies have shown that smartphone users rely on their phones for a wide variety of tasks including for accessing information on websites, further highlighting the need for website designers to carefully consider the responsiveness and usability of their websites on mobile devices (Gebauer et al., 2008; King and Youngblood, 2016; Marcotte, 2017). The problem of website usability on mobile devices is exacerbated further given the small screen real estate available on these devices and given the limitations with broadband access to the internet via smartphones. Reliance on smartphones for accessing websites is particularly prevalent among younger users, and among Hispanic, and non-Hispanic black users (Brown et al., 2016) further highlighting the need to consider user characteristics in website design.

Compounding the problem of accessing emergency information via smartphones is the availability and use of social media in getting information quickly out to users. Undoubtedly, there are advantages to using social media for disseminating information (for example, targeted location-based information pushes) (Merchant et al., 2011), but there is also a need to balance the convenience and speed of information dissemination with the accuracy and authenticity of information pushed out via social media. Given the ease with which social media can be embedded into websites, the impact of social media on website usability becomes important and interesting to examine.

2.4. Summary of literature

To summarize, our brief review of the usability literature indicates that we know much about how to incorporate usability into product designs. We also know how websites, in general, can be made usable. Further, given the vital role that websites disseminating and communicating emergency information to the public play in emergency management, researchers have clarified specific design criteria these websites must be based upon. What we do not know, however, is how usable the COVID-19 dedicated websites that public health agencies in the US have created are.

Many characteristics of the pandemic make it difficult for the public health agencies to make the COVID-dedicated websites usable. Given the rapidly spreading pandemic of this scale, scope, and complexity, and given the ever evolving and changing nature of the underlying
symptoms of the pandemic, public health agencies have had to manage not only the voluminous critical health information with backend digital data storage systems, but they have also had to manage the roll out, the updating and regular dissemination of this vital health information to the public and other consumers via their websites. They have also had to incorporate current trends such as data dashboards, and social media embeds into the public health webpages. Furthermore, the capability to embed live and recorded media broadcasts into websites via platforms such as YouTube, Twitter and Facebook, have imposed an additional burden on the health agencies - the line between people consuming information primarily via traditional broadcast media such as television, and people relying on new media platforms as a primary source of information has blurred significantly, forcing public health agencies to keep up with the new media for dissemination. Given these competing demands, it is especially critical at this time to evaluate how usable these public health emergency dedicated websites are, and if users can readily access, understand and use the information on these websites.

3. Study objectives

The main objectives of our study were to (1): investigate how usable COVID-19 dedicated public health websites in the US were; (2) investigate whether case counts in states in different geographical regions had a relationship with the usability of the websites – our expectation was that states that had higher case counts and spread of infection would have more usable content than states with lower case counts and infection spread.

4. Material and methods

We illustrate our study’s overall methods architecture in Fig. 1. Our study activities and methods contained 3 broad phases – an initial planning and study requirements gathering phase, an evaluation and rating phase, and an analysis and reconciliation phase. We describe our study activities and methods in these three phases in detail in the following sections.

4.1. Study design and selection of websites for study

We selected COVID-dedicated websites of 16 state public health agencies in the US for our study based on the following two criteria: (1) that the state represented a particular geographical region of the country (Northeast, Midwest, South, West) based on the census classification of the geographical regions in the United States by the US Department of Agriculture (USDA Geographical Regions, 2020); and (2) that within a geographical region, the selected states represented the 2 highest and the 2 lowest COVID-19 case counts. We finalized the selection of states to include in the study in May 2020 to ensure that at the time we picked these states, their high and low case counts were approximately similar in magnitude across the four geographical regions.

4.2. Usability guidelines and criteria used in the website evaluations

In looking to guidance from researchers on how to systematically select and/or develop context and application specific usability criteria and heuristic evaluation guidelines, we used a systematic approach like what Rusu et al. (2011) suggest. They suggest a 6-step methodology for establishing new application specific usability heuristics. In step 1, they suggest exploring the research literature related to the core research topic, and the characteristics of the specific application domain. For our study, the application domain and the context of evaluation are related to government healthcare websites and portals targeted to communicating information about the pandemic. Therefore, we first gathered research literature focused on usability evaluations of government health websites. As part of this literature search, we also looked for validated and published usability evaluation criteria applicable specifically to healthcare websites – our search yielded the guidelines published by Shneiderman et al. (2006), specific to our application context – that is, to evaluate the usability of websites related to healthcare communication by government entities. Therefore, for this study, we used selected guidelines and criteria from the research-based web usability design and evaluation guidebook developed by the U.S. Department of Health and Human Services and the U.S. General Services Administration (Shneiderman et al., 2006) to conduct our heuristic
evaluation of the COVID-19 state public health websites.

Similar to what Rusu et al. (2011) suggest as step 2 (descriptively formalizing the main concepts to focus on for the evaluation), and step 3 (correlating application specific usability needs with traditional usability heuristics and criteria), from among the usability evaluation criteria available in Shneiderman et al. (2006) guidebook (for the complete guidebook, please see US Department of Health and Human Services Usability Criteria), we selected 130 criteria to use in our evaluation through a careful process of individual review and selection, and a team discussion and finalization. The guidebook provides not only the relative importance of the criteria vis-à-vis usability based on expert consensus, but it also documents the strength of evidence for every criterion based on research findings.

To first decide which criteria to use for the study, we (the authors of this paper) each independently reviewed every criterion from the guidebook for its usefulness, its relative importance, the strength of evidence for that criterion, and its relevance to present day websites. Then, in a spreadsheet listing all the criteria, we each separately voted an “yes” if we wanted the criterion included in the study, or a “no” to indicate our judgement to exclude a criterion from the study. We then compiled our independent yes/no ratings for every criterion into one spreadsheet. Then, all five of us met and discussed our ratings. If at least 3 of us voted “yes” to include a criterion, that criterion was automatically selected for the study without discussion. If only 2 of us voted “yes”, we discussed that criterion further to reach consensus on whether to include or exclude it from our study.

After our discussion process, we had excluded 54 criteria from our study (see appendix 1 for a list of all excluded criteria) because of the following reasons:

1. if the criteria directly involved users and user testing, we excluded those; for example, establishing user requirements and involving users in establishing user requirements; ours was a heuristic evaluation and did not directly involve end users.
2. if the criteria were meant to capture the design process of the website designer when they designed the site; for example, setting and stating design goals or using personas for design. We evaluated only the already designed and publicly available website, and not the designer.
3. if the criteria were irrelevant or outdated to present day webpages; for example, we felt that the criterion “avoid screen flicker” was irrelevant to most present-day websites.
4. if the criteria were such that we were unable to objectively measure or subjectively assess; for example, the criterion attend to home page panel width or designing for commonly used screen resolutions or using glosses to assist navigation was beyond our assessment and scope of measurement.

The criteria we selected for the study represented a variety of usability categories including user experience, hardware and software requirements, homepage usability, page layout, website navigation, scrolling and paging functions in the websites, website headings, links in the websites, text appearance, lists in the sites, graphics design in the websites, content organization, and search functionality.

In addition to the criteria from the US Department of Health and Human Services guidebook (Shneiderman et al., 2006), we selected usability guidelines from Youngblood and Youngblood (2018) for evaluating readability in the websites, a website’s findability, and the extent of responsive design in the site.

Additionally, we evaluated dashboard usability based on Lechner and Fruhling’s (2014) work on dashboard design. These criteria include whether dashboards conveyed the state of urgency, whether they provided actionable information for users, whether they supported correct data interpretation, whether they aggregated information appropriately, whether they adhered to dashboard design conventions, whether their design aesthetics were minimalist, whether they included GIS interfaces, whether they included relevant content that minimized cognitive processing, and whether they provided temporal trends.

Finally, we added our own criterion that we labeled “conveys risk information” to evaluate the degree to which the COVID-19 dedicated websites conveyed risk information in readily usable form to the users about the pandemic. We did not find this criterion in any of the other guidelines we referenced for the study – the closest was the state of urgency criterion specific only to the dashboard, and not to the website.

After finalizing the 148 evaluation criteria to use in our study, much like what Rusu et al. (2011) suggest as the explicative stage and step 4 in their methodology, we created a master spreadsheet to use as our template for the evaluation with the criteria in rows and an individual evaluator’s ratings in columns, with one set of columns for each of the 16 states in our study. We created drop-down lists in each cell in Excel for evaluators to pick and enter their ratings. The plan was for each of us to fill the blank cells in the spreadsheet with our rating for every criterion for each state when we independently reviewed each website. We describe our rating scale further in section 4.4.

4.3. Evaluators’ background

Researchers conducting heuristic evaluations have found that with five evaluators, they were able to increase the number of usability problems uncovered (Afacan and Erbug, 2009; Nielsen, 1994, 2000). All five of us (the authors of this paper) served as evaluators in the study, and participated in the selection of websites, the selection of the usability criteria to use in the study, and the heuristic evaluation of the websites. We are all trained human factors engineers and researchers and are well-versed in conducting usability evaluation and data analyses. Our criteria selection process, and the use of five experts for the heuristic evaluation reflects experiences of other researchers who have performed usability studies based on instruments with more than 100 items (Cosgrove, 2018; Park et al., 2018), and where the number of expert evaluators was not large to identify the main usability problems (Virzi, 1992).

4.4. Heuristic evaluation procedure

We used a heuristic evaluation procedure to evaluate the public health websites. In following a structured process for the heuristic evaluation, our focus was to minimize subjectivity. Therefore, we took many precautions when we conducted the evaluation. Similar to recommendations from Solano et al. (2016), because we did not use a combination of usability evaluation methods, before beginning the heuristic evaluation, we agreed on a set of activities each evaluator would perform, the specific deliverables and due dates for those activities, how we would communicate and discuss the evaluation, and how we would reconcile our evaluations. The five of us first independently evaluated all 16 state public health agency COVID-specific webpages using the evaluation criteria (please see an example snapshot from our Excel spreadsheet in Fig. 2).

We randomized the order in which each of us individually evaluated a state’s website. We linked the home pages of the public health agency websites in a cell in the master spreadsheet so that all of us could begin the heuristic evaluation from a consistent, single location. After navigating to a state’s public health agency site, we specifically reviewed COVID-19 dedicated webpages, paying particular attention to how the COVID-dedicated pages were displayed. If the health agency provided a separate link or an external website for COVID-19 information from within the public health website, we followed the link to the COVID site for the heuristic evaluation. We used the usability criteria described in section 4.2. As mentioned previously, when finalizing the criteria, we discussed them to ensure that all of us understood what each criterion meant and how to apply it when evaluating a website.

We independently evaluated all the 16 websites, identified whether the webpage violated criteria and provided severity ratings for each
usability criterion by responding to a binary “yes” or “no” to the question “Was the criterion violated?” If the answer was a no, the usability rating was coded as a 0 (no usability problem). If the answer to the question was yes, we each rated the usability problem indicated by the criterion using severity ratings ranging from 1 (cosmetic problem) to a 4 (a usability catastrophe) (see Table 1 for a complete listing of our rating scale). Each severity rating also indicated the urgency of action required to resolve the usability problem. For example, the rating 4 (usability catastrophe) indicated that the website could not be used without addressing these problems first, whereas a rating of 2 (a minor usability problem) indicated that fixing the usability problem would have a low priority for the designers. We also added detailed comments in our individual spreadsheets whenever needed to provide context into the usability problem we saw, and to explain our rating so that we could discuss the ratings during the reconciliation process.

Once we completed all the individual evaluations, we transferred the individual ratings into a single master spreadsheet (please see Fig. 3 for an example snapshot of our master spreadsheet with all our ratings in one sheet) to begin the ratings comparison, discussion, and reconciliation process.

In a series of meetings held from June 2020 to October 2020, the five of us met via Zoom every week for about 2 hours comparing our ratings, discussing our ratings, and adjusting and reconciling our ratings, for every usability criterion, for all 16 states. Our discussions proceeded state by state. We first developed two rules to decide whether we would discuss our ratings further: (1) we would not discuss the criteria if all 5 of us had a consensus “no” response to whether a website violated a criterion; and (2) we would discuss a criterion if at least one of us had responded “yes” to whether a website violated a criterion and had rated the violation with a 2 or above score.

During the discussion, we provided each other a verbal justification for our numerical ratings. During the meeting, we also referred to the corresponding webpage to facilitate discussion and pinpoint specific features on the website that we felt violated a criterion as a basis to justify our rating. This process of discussion and reconciliation helped us reduce variability from three sources: (1) one of us missing a usability problem; (2) avoiding as much as possible any personal opinions about the website; and (3) minimizing order effects pertaining to the order in which each of us evaluated each state, particularly given the large number of criteria we used and given that each of us evaluated 16 states.

After a thorough discussion of each criterion, for every state, we could either revise our ratings based on the discussion or retain our original ratings.

To estimate the degree of agreement in our ratings, we computed three different measures: (1) first, we computed a simple frequency distribution on how many times we agreed (the interrater agreement) by extracting the raw scores from the master spreadsheet, across all 16 states and all 148 usability criteria, much like what Georgsson et al. (2014) suggest; (2) using an assessment agreement consistency rating for each of the 148 criterion across all states and evaluators as suggested by Landis and Koch (1977) with the consistency rating scale ranging from poor to almost perfect (see Table 2); and (3) the Fleiss Kappa consistency rates (Falotico and Quatto, 2015) for each criterion across all states and evaluators with the consistency rating scale again ranging from poor to almost perfect. We had 148 usability criteria in the study. Minitab’s Fleiss Kappa analyses computed the incidences of agreements and assigned it a strength of assessment agreement for each usability criterion – we then summed all the incidences for each category of agreement strength; for example, the five of us had an almost perfect agreement on 61 of the 148 criteria across all states. 61 almost perfect agreements out of 148 assessments represents 41.2%. The Fleiss Kappa Rates in Table 2 were calculated similarly.

From the simple frequency analysis, it is evident that after the reconciliation discussions, 68% of the time, all five of us agreed and assigned the exact same rating for a criterion, whether it be a rating of 4, 3, 2, 1, or 0.16% of the time, 4 of us agreed on an identical rating for a criterion, and 12% of the time, 3 of us agreed on an identical rating. Therefore, at least 3 of us (60% of us) agreed on a severity rating 96% of the time after the reconciliation discussion.

Table 2, it is evident that 83% of the time our rating agreement consistency was moderate to almost perfect over all the 148 criteria. The Fleiss Kappa rates further indicate that the strength of those agreements was reliable 69% of the time. Nearly 96% of the time, the consistency of our agreement over all criteria was fair to almost perfect, with an 88% reliability as indicated by the Fleiss Kappa rates.

The strength of the agreements between our ratings indicate that the rating reconciliation processes we used and the discussions we held helped us significantly clarify our ratings and reach a good degree of consensus in our usability evaluations for all the criteria.

Once we completed all discussions, we ranked our final usability ratings based on the severity of the usability problems to generate further descriptive results described in section 5, and to generate recommendations for any improvements for websites that may be developed in the future for pandemics or other public emergencies.

### 5. Results

When the COVID-19 pandemic created a high-risk public health emergency worldwide, the state public health agencies in the US created COVID-19 dedicated websites to disseminate information to the public. Given the rapid creation of many of these state websites, we set out to

| Usability rating | Meaning of the rating |
|------------------|-----------------------|
| 0                | Not a usability problem at all |
| 1                | Cosmetic usability problem only; need not be fixed unless extra time is available. |
| 2                | Minor usability problem; fixing this should be given low priority |
| 3                | Major usability problem; fixing this should be given high priority |
| 4                | Usability catastrophe; imperative to fix before release |

**Table 1**

Usability rating scale used in our study.

**Fig. 2.** A snapshot from a consolidated Excel spreadsheet showing our individual ratings and comments for various usability criteria before our reconciliation analyses.
To determine how usable the websites were, we undertook two sets of analyses. First, we computed a criterion wise score for each usability criterion to determine how usable the websites were with respect to every criterion. Second, we grouped the individual usability criteria into broader usability categories and computed a similar category wise score. Each category comprised many individual criteria, and hence represented the usability aspects under consideration broadly. For example, navigation was a category comprised of specific criteria such as navigation options, tabs, menus etc.

To explore the relationships between case counts of COVID-19 in the states and the usability scores for the websites, we undertook four distinct sets of quantitative analyses including cross-tabulations and non-parametric and parametric analyses as follows: (1) from the master spreadsheet of usability scores for all criteria, we extracted all the criteria that had an average rating of 3 and 4, as these represented the most severe usability problems, and cross-tabulated these with high and low case counts; (2) we conducted non-parametric comparisons of usability scores among high and low case count states using Mood median tests; (3) we conducted an ANOVA to test for differences in the ability scores among high and low case count states using Mood median tests; and (4) we conducted a correlation analysis between actual case counts and average usability scores for that state. We present results of each of these sets of quantitative analyses in the following sections.

5.1. Criterion and category wise usability scores

5.1.1. Criterion wise usability scores

From the final reconciled master spreadsheet of usability ratings, we computed a proportion usability score for each criterion across all states and all evaluators. The proportion score for a criterion represents the ratio of the score we assigned for a certain criterion to the maximum score possible (worst score) for that criterion. The total maximum usability score for a criterion is 320, obtained as the product of 16 states x 5 evaluators x the worst score of 4 on the rating scale. The actual score is

| Table 2 | Assessment agreement consistency ratings and Fleiss Kappa Rates. |
|---------|---------------------------------------------------------------|
| Strength of agreement | Agreement Between Evaluators | Fleiss Kappa Rate |
|                        | Incidence | Percentage | Incidence | Percentage |
| Poor                  | 0           | 0.0%       | 8          | 5.4%       |
| Slight                | 5           | 3.4%       | 9          | 6.1%       |
| Fair                  | 20          | 13.5%      | 29         | 19.6%      |
| Moderate              | 25          | 16.9%      | 43         | 29.1%      |
| Substantial           | 37          | 25.0%      | 24         | 16.2%      |
| Almost Perfect        | 61          | 41.2%      | 35         | 23.6%      |
| Total Assessments     | 148         | 148        |            |            |

investigate how usable these COVID-19 dedicated public health websites were, and whether case counts in these states in different geographical regions correlated with the usability of the websites – our expectation was that states that had higher case counts and spread of infection would have more usable content than states with lower case counts and infection spread.

To determine how usable the websites were, we undertook two sets of analyses. First, we computed a criterion wise score for each usability criterion to determine how usable the websites were with respect to every criterion. Second, we grouped the individual usability criteria into broader usability categories and computed a similar category wise score. Each category comprised many individual criteria, and hence represented the usability aspects under consideration broadly. For example, navigation was a category comprised of specific criteria such as navigation options, tabs, menus etc.

To explore the relationships between case counts of COVID-19 in the states and the usability scores for the websites, we undertook four distinct sets of quantitative analyses including cross-tabulations and non-parametric and parametric analyses as follows: (1) from the master spreadsheet of usability scores for all criteria, we extracted all the criteria that had an average rating of 3 and 4, as these represented the most severe usability problems, and cross-tabulated these with high and low case counts; (2) we conducted non-parametric comparisons of usability scores among high and low case count states using Mood median tests; (3) we conducted an ANOVA to test for differences in the ability scores among high and low case count states using Mood median tests; and (4) we conducted a correlation analysis between actual case counts and average usability scores for that state. We present results of each of these sets of quantitative analyses in the following sections.

5.1. Criterion and category wise usability scores

5.1.1. Criterion wise usability scores

From the final reconciled master spreadsheet of usability ratings, we computed a proportion usability score for each criterion across all states and all evaluators. The proportion score for a criterion represents the ratio of the score we assigned for a certain criterion to the maximum score possible (worst score) for that criterion. The total maximum usability score for a criterion is 320, obtained as the product of 16 states x 5 evaluators x the worst score of 4 on the rating scale. The actual score is

The product of 16 states x 5 evaluators x the actual score on the rating scale. The smaller the actual score on the usability rating scale, the better. Hence, the lower this proportion, the better the website with respect to that criterion. Therefore, the closer the proportion is to 1, the worse the website fared on criteria.

Table 3 presents the usability criteria that had a proportion score of 0.2 and above. Results indicate that the “designating used link” criterion had a proportion usability score of 0.5 – many websites did not change link color once a user had visited a link. Providing printing options on the webpage and the ability to convey overall risk information each had scores of 0.45, indicating poor user options for printing in general, and poor conveyance of overall pandemic risk in a particular geographical region. While the dashboards provided updated case numbers, they did not necessarily convey risk information in a readily usable form. Websites also had trouble optimizing display densities of content and information, given the volume of information that was being posted to the websites daily. The rest of the usability criteria scored below 0.4.

| Table 3 | Usability criteria with a 0.2 proportion score and above across all states. |
|---------|--------------------------------------------------------------------------|
| Criteria | Proportion score |
| Designating used links | 0.50 |
| Providing printing options | 0.45 |
| The webpage conveying overall risk information | 0.45 |
| Optimizing display density | 0.43 |
| Availability of page for mobile device viewing | 0.39 |
| Placing primary navigation menus in the left panel | 0.38 |
| Formatting information for reading and printing | 0.36 |
| Organizing information in the content clearly | 0.36 |
| Developing pages that will print properly | 0.32 |
| Ensuring the homepage looks like a homepage | 0.31 |
| Minimizing the number of clicks on pages | 0.30 |
| Using images to facilitate learning | 0.29 |
| Ordering list elements to maximize user performance | 0.28 |
| The data dashboard conveying the state of urgency | 0.28 |
| Enabling access to the homepage | 0.28 |
| Differentiating and grouping navigation elements | 0.27 |
| Providing feedback on user’s location | 0.26 |
| Responsiveness to resizing the page on a desktop browser | 0.26 |
| Avoiding cluttered displays | 0.25 |
| Understanding and meeting user expectations | 0.25 |
| Organizing content to facilitate scanning | 0.25 |
| Placing important items at the top center | 0.24 |
| Establishing the levels of importance | 0.23 |
| Reducing the user’s workload | 0.23 |
| Displaying information in a directly usable format | 0.22 |
| Using fluid layouts | 0.22 |
| Options to view pages in a language other than English | 0.22 |
| Limiting prose text on the homepage | 0.21 |
| Avoiding misleading cues to click on links | 0.21 |
| The data dashboard having minimalist aesthetics by reducing non-data ink | 0.21 |
enabling usable search results, designing a search engine to search the entire site, providing a search option on each page on the website, designing search around users’ search terms, and notifying users when multiple search options exist in the page. From the final reconciled master spreadsheet of usability ratings for the usability categories, we computed a proportion usability score for each category across all states and all evaluators. The proportion score for a category represents the ratio of the score we assigned for a category to the maximum score possible (worst score) for that category. The worst score possible on each criterion is a 4, so, for example, the worst score possible on the category “Search” is \(4 \times 5 = 20\). With 5 evaluators and 16 states in the evaluation, the total worst possible score for this category hence is \(20 \times 5 \times 16 = 1600\). Hence, the smaller the actual score the better the usability. The closer to 1 the proportion, the worse the website fared in that category.

Table 4 presents the usability categories that had a proportion score of 0.2 and above. Navigation had a proportion usability score of 0.2 among all usability categories evaluated in the study, indicating some scope for improvement of the various navigational options in the pages, and how those navigational options were organized on the pages. The design of the homepage and the overall page layout also scored close to 0.2. We felt that several homepages did not completely reflect the look and feel of a typical homepage; further, some pages communicated the purpose of the website poorly and contained voluminous prose text. The major options available on the website were also poorly communicated on the homepage in a few instances. As for the page layout category, some websites contained cluttered pages, and were inconsistent in locating important items within their pages. A few websites also had alignment problems and exhibited poor use of white space. Overall, though, when grouped into major usability categories, on an average across all states and evaluators, the websites did not score poorly in the important usability categories.

5.2. Case count-based differences in usability

5.2.1. Cross-tabulation of major usability problems by high and low case counts

Tables 5 and 6 present results extracted from the master spreadsheet for usability criteria in any state that received an average rating of 4 (usability catastrophe) and a 3 (major usability problem) respectively, cross-tabulated by high and low case count states. On the usability scale, a 4 represented a usability catastrophe which was imperative to fix. A 3 represented a major usability problem, important enough to fix. Results indicate that problems in search functionality, and problems with providing alternate language options on the website received the worst usability ratings from us consistently in both high and low case count states.

### Table 4

| Usability category | Proportion score |
|--------------------|------------------|
| Navigation        | 0.20             |
| The homepage      | 0.17             |
| Page layout       | 0.17             |
| Optimizing the user experience | 0.16 |
| Content organization | 0.15 |
| Design process and evaluation | 0.14 |
| Readability, findability, and responsiveness | 0.14 |
| Search            | 0.13             |
| Links             | 0.13             |
| Dashboard design  | 0.12             |
| Lists             | 0.12             |
| Headings, titles, and labels | 0.09 |
| Scrolling and paging | 0.08 |
| Text appearance   | 0.07             |
| Graphics, images, and multimedia | 0.07 |
| Writing web content | 0.03 |
| Hardware and software | 0.03 |
| Screen-based controls (widgets) | 0.02 |

### Table 5

| Usability criteria with an average rating of 4 |
|-----------------------------------------------|
| Case count category (high and low)            |
| High                                          |
| • Communicating the website’s value and purpose |
| • Option to view the website in a language other than English |
| • Designing the search engine to search the entire site |
| • Providing a search option on each page in the website |
| Low                                           |
| • Option to view the website in a language other than English |
| • Designing the search engine to search the entire site |
| • Providing a search option on each page in the website |
| • Designing search around users’ search terms |

### Table 6

| Usability criteria with an average rating of 3 |
|-----------------------------------------------|
| Case count category (high and low)            |
| High                                          |
| • Presenting tabs effectively                 |
| • Using at least 12-point font                |
| • Mobile device viewing                       |
| • GIS interface (zooming in and out to view detailed information, and providing situational awareness of activities and trends) |
| • Using fluid layouts                         |
| • Designing used links                        |
| • Clarifying clickable regions of images      |
| • Minimizing user data entry                  |
| • Providing useful content                    |
| • Limiting prose text on the homepage         |
| • Placing primary navigation menus in the left panel |
| • Providing consistent clickability cues      |
| • Enabling access to the homepage             |
| • Differentiating and grouping navigation elements |
| Low                                           |
| • Displaying only necessary information       |
| • Using fluid layouts                         |
| • Desktop browser rezining                    |
| • Mobile device viewing                       |
| • Enabling access to the homepage             |
| • Including actual data with the data graphics |
| • State of urgency (organizing information by color, graphs/trends, emergency/urgent alert that allow user to determine if a state of urgency exists) |
| • Supporting correct data interpretation (that is, helping the user understand information and perform actions like meaningful comparisons correctly) |
| • Placing primary navigation menus in the left panel |
| • Differentiating and grouping navigation elements |
| • Using appropriate menu types                |
| • Designing used links                        |
| • Ensuring that website images convey intended messages |
| • Enabling usable search results              |
| • Conveying risk information                  |
| • Aligning items on a page                   |
| • Providing feedback on user’s location during navigation |

Fig. 4 below depicts an example of a good homepage on a COVID-dedicated site that provided immediate language options for the user in the top banner of the page. The page also translated the data tables into Spanish – given the diversity of the population in the US, these options were especially valuable and important in the pages. Usability criteria that received an average rating of 3 in both high and low case count states included search functionality, page layout and...
Fig. 4. An example of a well-designed homepage providing direct links (indicated by the red arrows) for language options. This page immediately translated all content into Spanish. The data dashboards also were translated into Spanish. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

resizing problems, and failure to designate used links. In addition, low case count states also received an average rating of 3 from us for not indicating the urgency of the COVID status in the state, and for a lack of features for the user to interpret the data provided on the website.

One usability problem we detected in almost all pages was the lack of interpretation of the data presented to the user (see Fig. 5a and b). Fig. 5a is an example of a website that conveyed risk information to the user poorly. Although the page presented numbers in different categories, and a map with different color codes, the interpretation of risk was left to the user. In comparison, in Fig. 5b, the risk information in the page was coded from the data with an interpretation of the case characteristics with respect to target values, providing a ready indication to the user about warnings if the targets were being exceeded (for example, the warning sign in the new cases compared to target). Although these pages presented numbers, there was no way to interpret these numbers into an overall risk indicator. We felt this was an
important missed opportunity for these websites, given the initial case counts in selected geographical regions of every state (for instance rural counties had fewer cases than urban regions in a state) and given how travel restrictions and factors such as social distancing impacted the spread of infection and given how the overall risk changed from state to state when the pandemic spread across state borders.

Poorly built search functionality was another usability problem in many websites. We were particularly surprised by this problem, given how prevalent search engines are, and given the case with which searches can be incorporated into websites. Fig. 6 is an example of a clearly labeled and prominent search function (indicated by the red arrow) on a landing homepage. While many websites provided a global search engine (such as the one in Fig. 6), the search functionality came without advanced search options to narrow searches or to focus searches to specific topics of interest – we felt this feature was important given the volume of information these webpages carried and given that these websites were being updated with COVID information almost daily.

5.2.2. Median differences in usability scores between high and low case count states

To determine if the median usability scores of the states were different from each other, we conducted a non-parametric Mood’s median test on the medians of the totals of the usability scores we each gave each state. From the table of total scores, the medians of the total scores for each state were then compared to the overall median. Table 7 presents the observed and the expected frequencies from the Mood’s median test. The region and case count categories are coded in the table – W represents West, H represent high case count, and 1 refers to one state from that region; MW represents Midwest, L represents a low case count, and 2 represents a second state sampled from that same region. E represents the eastern region of the US, and S represents the southern region of the US. As an example of how we computed the medians, for WH1, the totals of the usability ratings from each of the 5 evaluators was 65, 74, 93, 94, 93 and 93 respectively. The median of these 5 totals is 93. We computed medians of our rating totals for all other states in the same way and ran the Mood’s median test against an overall median of 67.5. Results from the Mood’s median test showed that the overall median usability scores for the states were significantly different from each other at the 5% level of significance ($\chi^2 = 38.40; p = 0.001$).

In addition to the overall median test, we tested separately if the median usability scores within the high case count states were different from each other, and if the scores within the low case count states were different from each other. Results (Tables 8 and 9) indicate that the median usability scores within high case count states are statistically significantly different from each other at the 5% level of significance ($\chi^2 = 20.8; p = 0.004$), and that the scores within low case counts exhibit a similar statistically significant difference ($\chi^2 = 20.8; p = 0.004$). However, the median usability scores between high case count states and low case count states are not significantly different from each other at the 5% significance level.

| Region and Case Count Code | Median | N ≤ Overall Median | N > Overall Median | Q3-Q1 | 95% CI on Median |
|---------------------------|--------|--------------------|--------------------|-------|-----------------|
| WH1                       | 93     | 1                  | 4                  | 24.0  | (65, 94)        |
| WH2                       | 73     | 2                  | 3                  | 52.0  | (60, 134)       |
| WL1                       | 54     | 4                  | 1                  | 30.0  | (32, 79)        |
| WL2                       | 42     | 5                  | 0                  | 7.5   | (36, 47)        |
| MWH1                      | 44     | 4                  | 1                  | 42.5  | (32, 102)       |
| MWH2                      | 50     | 5                  | 0                  | 15.0  | (34, 51)        |
| MWL1                      | 61     | 3                  | 2                  | 21.5  | (38, 94)        |
| MWL2                      | 74     | 1                  | 4                  | 16.5  | (63, 85)        |
| EH1                       | 38     | 4                  | 1                  | 40    | (32, 95)        |
| EH2                       | 59     | 3                  | 2                  | 21    | (55, 81)        |
| EL1                       | 62     | 3                  | 2                  | 19.0  | (52, 74)        |
| EL2                       | 54     | 4                  | 1                  | 13.5  | (52, 74)        |
| SH1                       | 93     | 0                  | 5                  | 31.5  | (90, 137)       |
| SH2                       | 77     | 0                  | 5                  | 16.0  | (75, 92)        |
| SL1                       | 106    | 0                  | 5                  | 20.5  | (93, 127)       |
| SL2                       | 90     | 0                  | 4                  | 56.0  | (62, 148)       |
| Overall                   | 67.5   | 4                  | 4                  | 24.0  | (32, 95)        |

Table 8 Mood’s median test results on median values of the total usability score within high case count states. The computation of frequencies in the test are similar to the description for Table 7.

5.2.3. Mean differences in usability scores among high and low case count states using ANOVA

To test if the mean differences in usability scores were different among high and low case count states, we conducted an analysis of variance (ANOVA). First, we tested the normality assumption for the overall usability score for each state. The normal probability plot (Fig. 7) and the Anderson-Darling test for normality (Fig. 8) show that the overall usability scores for the states averaged over all criteria as rated by the evaluators were normally distributed ($p < 0.05$). Because the normality assumption was satisfied, we conducted an ANOVA to test if
the mean usability ratings for the states were different from each other. Results (Table 10) show statistically significant differences between the mean usability scores for the states at the 5% level of significance ($F = 6.33, p < 0.05$). Further post-hoc tests for pairwise differences in usability scores between states using the Tukey’s Honestly Significant Difference test (Table 11) reveal that there are more instances of pairwise differences in the mean usability scores between high and low case count states, compared to the number of instances of significant differences between just low case count state pairs, or between just high case count state pairs.

5.2.4. Correlation between case counts and usability scores
We also performed correlations between case counts per 10,000 people in every one of the 16 states (classified by high and low case counts) and the corresponding average usability score of the state (Table 12). Although not statistically significant, results (Fig. 9) indicate a negative correlation ($r = -0.209, p = 0.437$) indicating that the higher the case count, the better the usability score.

6. Discussion
6.1. Overview
Results from our quantitative analyses of the usability scores indicate that although the state public health sites disseminating COVID-19 information fared more than adequately on usability metrics for the most part, they did exhibit a few areas of concerns. We outline some of these concerns in the following sections. Our interpretations of the quantitative results in the following sections are based upon the extensive

| Region and Low Case Count States | Median | N ≤ | N > | Q3 – Q1 | 95% Median CI |
|----------------------------------|--------|-----|-----|---------|---------------|
| E-L1                             | 62.0   | 3   | 2   | 19.0    | (52, 74)      |
| E-L2                             | 54.0   | 4   | 1   | 15.5    | (52, 74)      |
| MW-L1                            | 61.0   | 3   | 2   | 21.5    | (58, 94)      |
| MW-L2                            | 74.0   | 0   | 5   | 16.5    | (63, 85)      |
| S-L1                             | 106.0  | 0   | 5   | 20.5    | (93, 127)     |
| S-L2                             | 90.0   | 1   | 4   | 56.0    | (62, 148)     |
| W-L1                             | 54.0   | 4   | 1   | 30.0    | (52, 78)      |
| W-L2                             | 42.0   | 5   | 0   | 7.5     | (36, 47)      |
| Overall Median                   | 62.5   | 2   | 4   | 23.5    | (49, 77)      |

Fig. 7. Normal probability plot for the overall usability scores. Each of the 16 states received a rating for every usability criterion from each of the 5 evaluators for a total of $n = 80$ sample points for the test of normality.

Fig. 8. Anderson-Darling test for normality of the overall usability score.

Table 10
Analysis of variance.

| Source | df | Adj SS | Adj MS | F-Value | P-Value |
|--------|----|--------|--------|---------|---------|
| State  | 15 | 31669  | 2111.3 | 6.33    | 0.000   |
| Error  | 64 | 21336  | 333.4  |         |         |
| Total  | 79 | 53005  |        |         |         |

Table 9
Mood’s median test results on median values of the total usability score within high case count states. The computation of frequencies in the test are similar to the description for Table 7.
comments we each recorded in the master spreadsheet when we assigned individual quantitative ratings. They are also based upon the joint discussions of our quantitative ratings we held during our reconciliation process.

Overall, a main finding of concern was that many states used these COVID-dedicated websites as an information repository and as a daily data dump. The websites were not necessarily seen as a venue and a medium to communicate risk of infection or spread, or to extensively promote compliance behaviors. Additionally, we found problems in page navigation and page layout design, with both these problems heightened by either too many links to external references such as to CDC’s main pages, or by too much information (resulting again from perhaps viewing these websites as a large data repository to dump data into). These problems further had the potential to reduce the quality of user experience when navigating the sites.

Almost all websites used data dashboards, but many dashboards suffered from weaknesses such as too much data, and a lack of derived interpretation of the data into usable indicators of the immediate risk of infection and spread in local geographical areas.

Across all states, some usability problems received low quantitative usability scores indicative of high negative ratings from us. These included the organization of content and information on the pages, the search functionality in the pages, navigation among the pages in a site, the design and layout of the home page on the sites, the design of hyperlinks in the sites, limited provision for viewing the pages on mobile devices particularly dashboard information, and limited communication of risk information including on dashboards. We discuss these concerns further in the next sections.

6.2. Organization of content

Our findings indicate that the content on the pages in many websites was not always organized based on the importance of the information or based on what the user might expect to see on the page first, and what information, therefore, was absolutely necessary to have on the page. We rated the organization of content across all state websites to be moderately poor as indicated by the criterion wise proportion usability score of 0.36 for this criterion (Table 3) and a category wise usability proportion score of 0.15 (Table 4). We rated content organization of these sites poorly because critical information about COVID-19 symptoms was buried deep under educational materials and not presented to the user upfront. Similarly, another page presented “case investigations and contact tracing” upfront but led the user into deeper menus to find even basic information on the symptoms when the pandemic was in its initial phases and beginning to spread. The organization of the content was also impacted by the display density of information in the webpages. We rated websites poorly in relation to optimizing display density with a usability proportion score of 0.43 (Table 3). Page layout, as a category, was rated at 0.17 (Table 4). We felt that, initially at the beginning of the pandemic, there was poor balance in many cases between the need to push out vital information about the pandemic to the public, and the need to supply usable and actionable information to users visiting these websites. To balance the large volumes of information posted on the websites with organizing this information clearly, some sites utilized the “frequently asked questions” and the “search” functionality effectively to organize all content and to display information as needed by the user. For example, on one website, COVID-19 information was organized according to the common questions and concerns users might have, while another website had comprehensive search and filter functionality to support the organization and display of important COVID-19 content. But these cases were exceptions. We found that most of the websites were content- and information-dense.

6.3. Search functionality

From a usability standpoint, search functionality can make or break a website (Calisir et al., 2010). Users tend to look for search functionality when they first land on a website (Nielsen, 1999). We found that some websites used search functionality as the main method for users to seek information from the website by placing the search function in a central part of the website, and by building advanced and sophisticated Boolean filters. A few sites also used chatbots to integrate search and help functionalities. However, we quantitatively rated the search functionality moderately poorly with a proportion usability score of 0.13 (Table 4) because many websites had poor search functionality - problems included a display of search results not related to COVID-19, a lack of flexibility in what the user could input as search terms in text form, and inconsistent placement of search bars and search fields. We rated three criteria - designing the search engine to search the entire site, the ability to use filters to sort search results, and the ability to search around users’ search terms each with a quantitative usability score of 4 representing a usability catastrophe (Table 5). We were surprised that the search functionality was completely missing in a few sites, especially in this day and age, when searching for information is a natural, efficient and accepted way for finding information. Research
suggests that usable search functionality along with navigation makes a webpage easier to use (Palmer, 2002). In many sites though, we found that good search functionality masked navigation problems in the site — because the search was designed and built well, we could quickly search for specific information, in lieu of navigating through the poorly laid out pages to unearth the same information.

6.4. Site navigation

Effective navigation is especially important for sites that disseminate a large volume of vital information. Our quantitative evaluation found that some websites included unique navigation features and aids that improved usability such as expandable boxes to select and view more information on a topic, a clickable list of contents in heading form at the top of a page to enable the user to quickly jump to a topic, and a scrollable table of contents on the left panel of the page that moved with the page scroll. However, overall, we rated the usability of navigation features as problematic in many sites as indicated by the criterion wise proportion usability score of 0.38 for not placing primary navigation menus in the left panel, 0.27 for not differentiating and grouping navigation elements, and 0.26 for not providing feedback on users’ location when navigating (Table 3). We also rated navigation as a category with a proportion usability score of 0.2 – navigational problems scored the highest among all categories of usability (Table 4). Navigation problems in these websites point to scope for improvement of the various navigational options in the pages, and how those navigational options are organized on the pages. Many sites had a varied mix of menu types even within a page, making it difficult for us to discern which information was found where. In these pages, critical information such as basic COVID-19 symptoms were hidden behind menus that did not have clear labels to indicate what they represented.

6.5. Design of homepages

The homepage, considered the landing page for many websites, serves as the most visible part of a website (Bucy et al., 1999; Ha and James, 1998; Singh et al., 2005). Users expect to find the most important information upfront on the homepage. The usability of a website’s homepage can determine whether users eventually revisit the site (Nielsen, 2000).

Our quantitative results indicate that many COVID-dedicated homepages were navigation pages full of links. They did not look like typical website homepages as indicated by our criterion wise rating of 0.31 for homepage usability (Table 3). Given the volume of information about COVID-19, the continuous updates to symptoms, recommendations and guidelines, and the many external stakeholders involved in the process such as the World Health Organization (WHO), the Centers for Disease Control (CDC), local businesses, and personal protective equipment vendors, and the need for state health agencies to share information relevant to these various stakeholders, the homepages on many of these websites were lengthy, and contained very dense text — we found voluminous amounts of information and a variety of content on the homepages, particularly on the landing pages. Our quantitative rating for limiting prose text on the homepage was 0.21 (Table 3). On many websites, we found navigating back to the homepage after visiting links cumbersome, and rated access to the homepage with a usability score of 0.28 (Table 3). Taken collectively, as a usability category, we rated homepage design with a score of 0.17 (Table 4).

Also, some websites also did not display critical external links such as to the CDC and WHO on their homepages. We also found many homepages cluttered with a multitude of external links, voluminous text, mixed formatting of text, images, and videos, and vertically and horizontally organized content on the same homepages. A few webpages counteracted this clutter problem by highlighting important information via display banners, alerts, and key COVID-19 information highlights on the homepage.

6.6. Links in the websites

We rated links on websites, as a usability category, at 0.13 (Table 4) because we found many problems with links in these websites. All the state public health sites we sampled relied heavily on direct external links to the CDC, the WHO, and other official information sources to share guidelines and recommendations. Depending extensively on external links would create problems if the links broke. Also, the links were sometimes formatted inconsistently within a site - some links were
underlined when we hovered over them indicating they were a link, while others were not. The design of links varied widely across the different public health sites, in the formatting as well as the extent to which links were used to convey information. For example, some links used color to indicate they were links, while others used underlining. Sometimes, an underlined text, typically assumed by users to represent a link, merely represented emphasis of words or sentences. Because these cues were misleading and prompted us to click on them in the belief that they were links, we rated this criterion with a 0.21 proportion usability score (Table 3). Images were also used as links, but many times, the clicking cues were inconsistent – some images were linked, and a majority of others were not. Notably, we also found the links did not change color or format consistently to indicate a “visited link”. We rated the “designating used link” criterion with a proportion usability score of 0.5 (Table 3). While we evaluated this traditional usability criterion to be poor in the websites due to the concerns mentioned above, during our discussion of the ratings, we agreed that website designers did not appear to be consistently designing for this feature. In some websites, links were used to expand and collapse content within a page and to better organize information. However, we found that most of the links connected to an external web page. We found that some websites used links heavily and made that page their main home page – this forced us to always click and navigate to external sources for even basic COVID-related information. On these sites, even the basic COVID-19 symptoms and guidelines were links to the CDC. In our discussions, we felt that the website could have at least presented static content on the homepages with the basic COVID symptoms, and how it spread, without the user having to click on a link to get this information from CDC directly. We felt the tradeoff between obtaining authentic information from a primary source such as CDC, and the time and effort it would take to highlight basic information about COVID on a static page was not well thought out on these pages. Highlighting just the links as the primary source of information could reduce the user interaction experience and require users to hunt for relevant and state-specific basic information.

6.7. Data dashboards

A key feature in almost all websites was a data dashboard or a link to one. These dashboards often provided detailed case counts at county levels, interactive charts and maps with filters, and data tables for downloading. We found a variety of dashboard elements based on the software used to run and populate the dashboards at the backend – many websites, for instance, used Tableau.

We found some important usability problems with the dashboards. For one, the dashboards had a large volume of raw data, making navigating and understanding the data potentially difficult for common users. We rated whether and how well the dashboards conveyed risk information to the public with a proportion usability score of 0.28 (Table 3), and whether the dashboard had minimal data ink with a score of 0.21 (Table 3). The design of dashboards as a category scored 0.12 (Table 4). While the dashboards presented aggregate information and tied them to maps of counties and states very well, they did not interpret the raw data and translate it into a risk assessment for the public. Also, we found that the dashboards did not scale well to smaller devices, particularly mobile devices like phones. At a time when mobile devices are pervasive for accessing the information on websites (King and Youngblood, 2016; Marcotte, 2017; Pew Research Center, 2017), and given the usefulness of data dashboards in communication during public health emergencies, usability concerns in accessing dashboards via mobile devices needs further consideration and research. Also, many dashboards took a while to populate and update. Many also did not resize and respond to fluid page widths and sometimes required extensive horizontal scrolling from us.

6.8. Use of social media on websites

Although the scope of our work did not include a detailed analysis of the social media used on these websites, we noted that all 16 state public health sites used several types of social media platforms including Twitter, Facebook, YouTube and Instagram. Updates about COVID-19, particularly daily press conferences and tweets from health officials, were directly embedded in the websites. Additionally, some websites included videos to create awareness and to educate the public about COVID-19. Given the prevalence of social media, and the advantages and disadvantages in using social media (Merchant et al., 2011), further research comparing the effectiveness of the various types of social media platforms in disseminating accurate, timely information for public consumption during health emergencies will be beneficial.

6.9. Communication of risk about COVID-19

We evaluated whether the COVID-19 state public health websites communicated risk about the pandemic sufficiently. While we do not consider risk communication a usability criterion per se, we still evaluated this criterion with the reasonable expectation that a public health agency, as part of its primary function, would communicate detailed risk information about the pandemic. Except one state, none of the websites included any specific features to communicate risk information directly, therefore, we rated this criterion at 0.45 (Table 3). The one state with a risk indicator on the site designed the indicator to change dynamically with changes in case counts in that state (see Fig. 5b). All other websites used features such a color gradation in their data dashboards as the primary means of communicating risk. They all left the interpretation of the risk to the user.

6.10. Case counts and usability of websites

Results from our ANOVA (Table 10) on the mean usability scores across all states at the 5% level of significance indicate statistically significant differences in usability between the states. A further post hoc analysis reveals that there are more instances of significant differences in mean usability scores between high and low case count states than between just the high case count states or between just the low case count states (Table 11). The differences we observed in usability between the high and low case count states may be indicative of the attention, effort and resources high case count states may have brought to designing pandemic focused websites, compared to low case count states, given the scale and scope of pandemic cases in their respective states. The negative correlation (not statistically significant though) (Table 12 and Fig. 9) between case counts and usability scores also supports this reasoning. While we expected to find major differences in usability between sites representing states with high and low case counts perhaps based on how much resources high case count states would dedicate to dissemination efforts, results from our Mood’s median analyses (Table 7) indicated significant differences in usability across all states, and particularly within just high case count states (Table 8), and within just low case count states (Table 9). We assumed that high case count states would perhaps expend more resources to make their site usable, but our findings indicate that this was not always the case. The significant differences in median usability scores even within the high case count states suggest that even a pressing need to communicate and disseminate health information in high-risk locations in the midst of a pandemic does not necessarily result in a well-designed, usable public health website during a health crisis. Additionally, the usability scores we computed across high and low case count states show how variable the design features were on these websites – they point to localized, on-demand web development efforts with minimal resources.
7. Study limitations

Heuristic evaluation, by its very nature depends upon an expert’s assessment and may be construed as subjective. While use of a combination of methods such as heuristic evaluations, user testing and user experience studies can yield more data and can potentially unearth more usability problems (Rusu et al., 2011; Solano et al., 2016), the special circumstances in our study conducted during the height of the COVID-19 crisis in the United States did not allow for any other methodology to be employed, particularly any validation experiments using user testing (step 5 of what Rusu et al. (2011) suggest in their method) or any refinements based on user testing and validation (step 6 of the (Rusu et al., 2011) framework).

We conducted our study using remote meeting software and screen sharing due to the physical distancing and lockdown measures in place in our communities. However, to minimize subjectivity in the study, and to maximize our chances for uncovering the important usability problems in the websites, we followed a rigorous process of evaluation and reconciliation of our ratings, accounting for individual variations as described in our materials and methods.

We considered only 16 out of the 50 state public health websites in the United States, which is a study limitation. But the 16 state websites we selected for our study are representative of websites of states in their respective geographical areas - we used very specific and reasonable selection criteria, namely, 2 states representing the highest case counts, and 2 states representing the lowest case counts in all 4 geographical regions (Northeast, Midwest, South and West) in the United States as classified by the US Department of Agriculture (USDA Geographical Regions, 2020) for census classification purposes. In balance, based on the effort and time we spent for extensively reviewing, evaluating, and analyzing a website, and because we were specifically interested in how a sample of high and low case count states were coping with updating COVID related information and advisories in their websites, our study samples helped us uncover many major and minor usability problems in these sites.

8. Conclusions and design recommendations

The usability concerns we identified suggest the need for more effective mechanisms to communicate information during a pandemic or other high consequence, longitudinal events that require emergency responses. The COVID-19 pandemic necessitated continuous updates to symptoms, guidelines, and recommendations; furthermore, the information content in the updates varied from state to state. Additionally, given the long timeframe of this pandemic, there has been an explosion of information and disinformation about the pandemic and its health effects.

Although the traditional news media and the new social media outlets from these state public health agency sources disseminated information, websites can supplement as official and credible sources of information; furthermore, they can store voluminous, time-critical, static, and dynamic information for easy access and use. But their designs need careful consideration of usability and effectiveness, particularly because they have become easy and inexpensive to design, update and maintain.

Our evaluation also revealed a need to reconsider web usability criteria. Websites no longer use some traditional usability features. For instance, we found that in almost all sites, options for printing were limited to printing PDFs and to the printing function integrated into web browsers – only a few sites offered separate print options. However, we did not think this would limit users from using the website for critical health information – users may no longer have a need to print pages, and could instead use browser prints, even if it became clumsy.

We also found that we did not have detailed usability criteria to evaluate some new and emerging features such as data dashboards. Future web usability design and evaluation criteria must consider these emerging web technology integrations.

Although our study was primarily an evaluation study based on established usability criteria from the US Department of Health and Human Services (Shneiderman et al., 2006), and was not intended as a study to develop new guidelines, based on our evaluation and based on our prior work in designing interfaces for human interaction (please see Mital and Pennathur (2000)), similar to what Collazos et al. (2021) in their work on designing guidelines for online interactions during a pandemic provide, we suggest 11 specific design recommendations in the following paragraphs to emphasize the importance of usability criteria we found important for designers to address in their work when designing health emergency dedicated websites in the future. We premise our design recommendations on established general principles for designing user interactions and mediating these interactions via interfaces. The core factors when designing interfaces for users must include user characteristics, task characteristics, and the characteristics of the use context and environment, at a minimum. When considering users, designers must elicit background information about users including data about their demographic and cultural characteristics, and their personal preferences and constraints. The tasks users will perform with the websites must also be considered carefully – task characteristics such as task criticality and dependencies, action sequences in the interaction, task demands including the perceptual and cognitive demands, and the degree of user discretion and control when interacting with the website, can all impact a user’s effectiveness in using a website, particularly in emergency, high-risk situations such as a pandemic. Contextual and situational factors that could arise when users navigate the websites must also be given careful consideration. These factors could include access to and availability of technology, interruptions users may experience when using technology to access the web, and any policy changes including local laws that may affect user experiences. Our design recommendations based on our heuristic evaluation follow.

8.1. Design recommendation 1: health emergency websites must convey overall risk information to the end user

During an emergency, people may use statistics reported by the public agencies at the county, state, and national levels as metrics to assess the risk to themselves and their families and friends. Although during a public health emergency such as an infectious disease outbreak, the numbers of positive cases, deaths, and people who recover are essential metrics to report and publish to the public, they do not completely convey risk to the public from the infection. Users are still left to interpret the numbers and trends and assess risk for themselves. Designers should strive to directly communicate risk to the user in addition to providing the raw data and statistical trends. Risk can be communicated with visual cues such as trend lines, warning signs, and different colors and icons signaling escalation and de-escalation of the threat (much like what was done to communicate security risks or like what The New York Times (The New York Times, 2020) dashboard uses to communicate pandemic risk). Further, emergency public health websites can exploit crowdsourcing technologies better and combine and leverage government efforts with citizen driven voluntary activities in sharing information about risks and threats by enabling features such as social media feeds that allow for public display of shared information. (Ludwig et al., 2017).

8.2. Design recommendation 2: health emergency websites must include data dashboards that convey the state of urgency to the end user

Data dashboards became more prominent after this COVID-19 pandemic and has been a useful feature for tracking disease trends. We recommend a few additional design considerations for dashboards based on our evaluation. Dashboards can include a composite assessment indicator for the state of urgency on a map for locations or next to the map in table form to facilitate risk communication. We recommend...
that dashboards include interactive features that allow users to select
detailed location information (i.e., county, city, state) and provide a
state of urgency based on the location, to facilitate travel considerations
during national health emergencies. On the map, we recommend
communicating the state of urgency using conventional urgency colors
instead of non-standard color schemes (i.e., red, yellow, green, for high,
moderate, and no urgency levels, respectively) so that the colors directly
respond to user mental models representing danger, moderate
danger, and no danger. In addition to using colors, including patterned
or texture areas could address accessibility concerns for color-blind
users.

8.3. Design recommendation 3: health emergency websites must strive to
include data dashboards with minimalist aesthetics by reducing non-data
ink
Users perceive and trust that a dashboard can convey critical infor-
mation. However, designers of data dashboards need to balance what a
dashboard must include with the amount of information they include in
them. Designers should avoid excessive non-data ink that distracts users
and does not offer them any additional useful information by elimi-
nating any unnecessary data, figures, graphs, and maps. While data
dashboard technologies now available conveniently mine vast amounts
of raw data and make it easy and automatic to populate the dashboards,
it is still the designer’s burden to filter out non-data ink and to present
the most appropriate and useful information to users.

8.4. Design recommendation 4: health emergency websites must be
designed for easy viewing on mobile devices
With the prevalence of smartphones and tablets, many users rely on
mobile devices exclusively to look for and review information on the
internet. Therefore, it has become a necessity to design web pages to be
responsive to mobile screens. Although, many websites use responsive
design features, in an emergency health website, some web pages such
as dashboards need to be designed to be available and responsive to
screen sizes commonly found in mobile devices.

8.5. Design recommendation 5: health emergency websites must use
responsive design and attempt to use fluid layouts for webpage design
Designers should design web pages to automatically adjust based on
the screen size and resolution. This is an important design consideration
particularly in web pages that include images, or in dashboards con-
taining tables and charts, where the content may not render and display
properly and would not fit all screen sizes. We agree with Cosgrove’s
(Cosgrove, 2018) recommendation that making health emergency
websites responsive will not only increase access to diverse populations
but also allow use of alternative devices for accessing these websites.
Fluid grids, flexible image use, and provisions for media queries and
social media embeds, all promote responsiveness of the websites (Cos-
grove, 2018).

8.6. Design recommendation 6: health emergency websites must provide
end users options to view pages in a language other than English
Public health emergencies affect people from diverse cultures who
may write and speak different languages. Our findings agree with that of
Venkatesh et al. (2017) that government run health websites must
consider different educational backgrounds and literacy of users. From a
readability and accessibility perspective, designers can include
embedded translation features or design the entire web page to convey
information in languages other than English. At a minimum, designers
can embed the freely available Google Translate or similar application
programming interfaces (APIs) to render pages in different languages for
users.

8.7. Design recommendation 7: health emergency websites must strive for
clarity in organizing the website content
During public health emergencies, people need to gain situational
awareness and learn about the different actions they can take and that
may be mandated by health agencies (e.g., quarantine guidelines). Users
can comprehend information and navigate that information more
effectively, when information is organized in some meaningful and
logical sequence, and when that logic is made apparent to the user.
Studies researching (Cockburn et al., 2017) the sequence of how people
interact with computers show that user preferences for a sequence of
tasks can often conflict with the sequences a designer may build into a
system for users. Therefore, designers should exploit effects such as
recency and primacy (Cockburn et al., 2017) to organize the content in
all web pages to clearly communicate the purpose of providing the
content, emergency messages, and guidelines, and to sequence user
interaction tasks that accentuate their experiences. At a minimum, the
content should be organized to provide general information about the
situation, the risk factors, and the appropriate course of actions
(guidelines) for users. The content should contain meaningful titles, sub-
titles, a logical hierarchy, and formatting features to demarcate it
clearly and easily. Websites can begin to incorporate smart, assistive
features that infer user intention and that reformulate a user’s input
requirements to match that intention (Cockburn et al., 2017). When
possible, some degree of sharing of information and hence standardi-
zation of the information architecture for health emergency website
content between national, state, county, and city websites will help
alleviate usability problems (Cosgrove, 2018).

8.8. Design recommendation 8: health emergency websites must optimize
the density of the information displayed
Optimizing information display density can help users scan and
obtain important and relevant health information quickly, especially
when information updates tend to be voluminous and change daily, as
would be the case during public health emergencies. Additionally,
optimizing information display density on a webpage would alleviate
the cognitive burden on users to have to search and sort through volu-
minous amounts of information. When users are worried about a public
health emergency and how the emergency might impact them, they
need to be able to find critical and relevant health information without
much effort – users may not only be visiting these websites for its news
value, but they may also be active users of the information such as pa-
tients and health care workers who rely on this information as an official
source for educating themselves and for caring for their patients. One
way designers can organize the content, for instance, is by displaying
similar and related information close to each other using interface
proximity compatibility principles (Proctor and Van Zandt, 2018).

8.9. Design recommendation 9: health emergency websites must strive to
use images to facilitate end user learning
Given the particularly information-dense nature of public health
emergency webpages, using digitally designed images and icons can
provide additional streams of information and let users map images and
icons with the text and learn by seeing what they must do in real world
scenarios and situations. The images should communicate the central
message associated with any text clearly and not mislead users into
actions that may compromise their safety. An important consideration
when using images is their size – larger images take longer to load and
may discourage and drive away users from that page. Therefore, de-
signers need to balance messages conveyed via text and messages
conveyed with images. Also, designers must include text alt attributes
to make the images accessible and inclusive.
8.10. Design recommendation 10: health emergency websites must designate visited links with appropriate color changes

Designers should use a different color for visited links to distinguish these links from unvisited links. Although designating used links might seem like a conventional design criterion, it can play an important role in providing a positive user experience when users need to navigate several pages of a public health website to find useful health information efficiently. During a public health emergency, the quantity and criticality of the guidelines published by the health agencies can increase and change very quickly, resulting in webpages having many new hyperlinks to these guidelines, but also continuing to contain old, no longer relevant links. Given the amount and mix of new and old links on these pages, if visited links are not distinguished, users may spend more time and feel frustrated clicking on and navigating through links that they may have already reviewed and did not intend to review again. This is particularly important because a user's perception of how long they take to complete a task correlates with their experience of using and interacting with websites and factors into their assessment of usability (Trukenbrod et al., 2020). Designating visited links can easily be done using an appropriate color selection when coding the page or by considering additional cues (lighter versus darker or bold versus normal fonts) for color-blind users.

8.11. Design recommendation 11: health emergency websites must format information for both reading online and for printing

Some users may prefer to read online, and some may prefer to print content to review at their convenience or to archive and access it anytime. Therefore, websites must support both online and printed readability. We recommend that designers plan to display all content for both sets of users and not design information that is only available for online reading. All web browsers have printing options – many render a PDF print file. Therefore, for users who choose to print a webpage from a browser print function, designers must ensure that the webpage content and format are appropriately reflected in the PDF. Additionally, including a printing icon may still be useful for some users. If a webpage conveys information that users may tend to print out (e.g., emergency, prevention, guidelines, vaccination information such as an address, what to expect, or tips upon arrival), designers should ensure that the page supports standard paper sizes (like Letter and A4). Designers should avoid using unnecessary whitespaces, redundant information on any blank spaces, overlapping headings, and any unusual widths for pages that can clutter the content. When website pages include PDF files as hyperlinks with information embedded in them, care should be taken not to convey important information only through the PDF document – designers can use the principle of redundancy to ensure users do not miss important information by providing a piece of short-length information on the webpage about what that PDF file contains, and whether users should print or save that file.

A usable website has the potential to facilitate rapid, credible, and effective communication and dissemination of vital health information, particularly during chaotic, high consequence, and information overloaded emergency events like a pandemic. But for a website to be usable, understanding users, their limitations and their experiences is also important (Law and Abraham, 2014). In addition to users, designers' background knowledge and training play important roles in making emergency health websites effective. As Alonso-Virgós et al. (2019) show, the degree to which web developers comply with usability guidelines can vary with their knowledge and training in usability. While we did not survey web developers in this study, in future work, we must attempt to understand how web developers think about incorporating usability guidelines when they design websites from scratch, especially for public health websites designed during health emergencies – do they find it easy to include certain usability elements than others even without any usability knowledge or training? What assumptions do they make about users? How closely do those assumptions match user expectations about what they can do with a website? Do the developers incorporate easy to design elements into a website more than difficult to design elements such as elements that may require more coding effort, or elements for which technology resources might be unavailable? In conclusion, our usability evaluation findings show that most state health agency webpages are usable overall with a few critical usability problems, but we still want to underscore the need for public health agencies and website designers to continue to find ways to improve the usability of these websites and leverage these sites as a medium to communicate critical public health information.

Author statement

All authors participated in the conceptualization, development and design of methodology, and formal analysis. Amirmasoud Momenipour and Brandon Murphy designed the visual illustrations in the project, and contributed to quantitative statistical analyses.

Arunkumar Pennathur and Priyadarshini Pennathur wrote the original draft of the paper.

All authors contributed to the review of the original draft.

During the revision stage after initial review of the manuscript by journal reviewers, Amirmasoud Momenipour, Salvador Rojas-Murillo, Priyadarshini Pennathur and Arunkumar Pennathur discussed the reviewer comments, prepared content in response to the reviewer comments, and contributed to revising the final revised document. In addition, Amirmasoud Momenipour provided additional visual illustrations, and wrote the design recommendations section. Salvador Rojas-Murillo strengthened the research literature section. Arunkumar Pennathur and Priyadarshini edited the final version of the document. Arunkumar Pennathur provided oversight and leadership responsibility, and Priyadarshini Pennathur led the administration of the project.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

APPENDIX 1. The table below lists criteria our study excluded from Shneiderman et al. (2006) and Youngblood and Youngblood (2018)

| Heuristics Category | Criteria | Relative Importance |
|---------------------|----------|---------------------|
| Design Process and Evaluation | 2: Establish user requirements | 5 |
| | 4: Involve Users in Establishing User Requirements | 5 |
| | 5: Set and State Goals | 5 |
| | 6: Focus on Performance Before Preference | 4 |
| | 7: Consider Many User Interface Issues | 4 |
| | 9: Set Usability Goals | 3 |
| | 10: Use Parallel Design | 2 |
| | 11: Use Personas | 1 |

(continued on next page)
Heuristics Category | Criteria | Relative Importance
--- | --- | ---
2: Optimizing user experience | 13: Do Not Require Users to Multitask While Reading | 3
4: Hardware and software | 5: Design for Commonly Used Screen Resolutions | 3
5. The homepage | 9: Attend to Homepage Panel Width | 2
6: Page layout | 13: Use Frames when Functions Must Remain Accessible | 1
7: Navigation | 10: Use Site Maps | 2
8: Scrolling and Paging | 11: Use ‘Glosses’ to Assist Navigation | 1
9: Headings, titles, and labels | 12: Breadcrumb Navigation | 1
10: Links | 9: Ensure that Embedded Links are Descriptive | 3
11: Text appearance | 10: Use ‘Pointing-and-Clicking’ | 2
12: Lists | 9: Color-Coding and Instructions | 2
13: Screen-based controls (widgets) | 7: Use Headings in the Appropriate HTML Order | 3
 | 8: Provide Users with Good Ways to Reduce Options | 2
 | 10: Do Not Make User-Entered Codes Case Sensitive | 4
 | 11: Screen-based controls (widgets) | 5: Label Data Entry Fields Clearly | 4
 | 12: Partition Long Data Items | 3
 | 13: Use a Single Data Entry Method | 3
 | 14: Prioritize Pushbuttons | 3
 | 17: Do Not Limit Viewable List Box Options | 3
 | 18: Display Default Values | 3
 | 19: Place Cursor in First Data Entry Field | 2
 | 20: Ensure that Double-Clicking Will Not Cause Problems | 2
 | 21: Use Open Lists to Select One from Many | 2
 | 22: Use Data Entry Fields to Speed Performance | 2
 | 23: Graphics, Images, and Multimedia | 2
 | 24: Provide Auto-Tabbing Functionality | 2
 | 25: Minimize Use of the Shift Key | 1
14: Graphs, images and multimedia | 3: Ensure that Images Do Not Slow Downloads | 4
 | 9: Limit the Use of Images | 3
 | 12: Introduce Animation | 2
 | 14: Use Thumbnails Images to Preview Larger Images | 2
 | 16: Using Photographs of People | 1
15: Writing web content | 8: Limit Prone Text on Navigation Pages | 3
 | 9: Use Active Voice | 3
 | 10: Write Instructions in the Affirmative | 3
17: Search | 3: Make Upper- and Lowercase Search Terms Equivalent | 4
 | 6: Allow Simple Searches | 3
 | 8: Include Hints to Improve Search Performance | 3
 | 9: Provide Search Templates | 3
18: Findability | 6: Presence of legacy sites | 2

**References**

Afacan, Y., Erbug, C., 2009. An interdisciplinary heuristic evaluation method for universal building design. Appl. Ergon. 40 (4), 731–744. https://doi.org/10.1016/j.apergo.2008.07.002.

Alonso-Virgo, L., Espada, J.P., Crespo, R.G., 2019. Analyzing compliance and application of usability guidelines and recommendations by web developers. Comput. Stand. Interfac. 64, 117–132. https://doi.org/10.1016/j.csi.2019.01.004.

Bevan, N., 2001. International standards for HCI and usability. Int. J. Hum. Comput. Stud. 55 (4), 533–552.

Bevan, N., 2005. Guidelines and standards for web usability. Proceedings of HCI International 2010, 5.

Brown, A., Lopez, G., Lopez, M.H., 2016. Digital Divide Narrows for Latinos as More Spanish Speakers and Immigrants Go Online. Pew Research Center, Washington, DC.

Bucy, E.P., Lang, A., Potter, R.F., Grabe, M.E., 1999. Formal features of cyberspace: Bevan, N., 2001. International standards for HCI and usability. Int. J. Hum. Comput. Stand. Interfac. 64, 117–132. https://doi.org/10.1016/j.csi.2019.01.004.

Falotico, R., Quatto, P., 2015. Fleiss’ kappa statistic without paradoxes. Qual. Quantity 49 (2), 463–470.

Fogg, B.J., Sooboo, C., Danielson, D.R., Marable, L., Stanford, J., Tauber, E.R., 2003. How do users evaluate the credibility of Web sites? A study with over 2,500 participants. Proceedings of the 2003 Conference on Designing for User Experiences 1–15.

Gebauer, J., Tang, Y., Baimai, C., 2008. User requirements of mobile technology: results from a content analysis of user reviews. Inf. Syst. E Bus. Manag. 6 (4), 361–384. https://doi.org/10.1007/s10257-007-0074-9.

Georgsson, M., Weir, C., Staggers, N., 2014. Revisiting Heuristic Evaluation Methods to Improve the Reliability of Findings. MISQ2014.

Gebauer, J., Tang, Y., Baimai, C., 2008. User requirements of mobile technology: results from a content analysis of user reviews. Inf. Syst. E Bus. Manag. 6 (4), 361–384. https://doi.org/10.1007/s10257-007-0074-9.

Hertzum, M., 2010. Images of usability. Int. J. Hum. Comput. Interact. 26 (6), 567–584. https://doi.org/10.1080/10447311003781300.

Ha, L., James, E.L., 1998. Interactivity reexamined: a baseline analysis of early business application systems. J. Broadcast. Electron. Media 42 (4), 457–474.

Hertrum, M., 2010. Images of usability. Int. J. Hum. Comput. Interact. 26 (6), 567–600. https://doi.org/10.1080/10447311003781300.

Huang, Z., Brooks, L., Chen, S., 2009. The assessment of credibility of e-government: users’ perspective. Symposium on Human Interface—Part 11: usability: Definitions and concepts. https://www.iso.org/obp/ui/#iso:std:iso:9241:1:ed-2:v1:en.

Jimenez, C., Rusu, C., Roncagliolo, S., Inostroza, R., Rusu, V., 2012. Evaluating a methodology to establish usability heuristics. In: 2012 31st International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 1815–1820.
of the Chilean Computer Science Society, pp. 51–59. https://doi.org/10.1109/SCCC.2012.14.
King, B.A., Youngblood, N.E., 2016. E-government in Alabama: an analysis of county voting and election website content, usability, accessibility, and mobile readiness. Govern. Inf. Q. 33 (4), 715–726.
Krug, S., 2009. Rocket Surgery Made Easy: the Do-It-Yourself Guide to Finding and Fixing Usability Problems (New Riders).
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174. https://doi.org/10.2307/2529310.
Law, E.L.C., Al враhão, S., 2014. Interplay between User Experience (UX) evaluation and system development. Int. J. Hum. Comput. Stud. 72 (6), 529–525. https://doi.org/10.1016/j.ijhcs.2014.03.003.
Lechner, B., Fruhling, A., 2014. Towards public health dashboard design guidelines. International Conference on HCI in Business 49–59.
Ludwig, T., Kotheaus, C., Reuter, C., Dongen, S. van, Pipek, V., 2017. Situated crowdsourcing during disasters: managing the tasks of spontaneous volunteers through public displays. Int. J. Hum. Comput. Stud. 102, 103–121. https://doi.org/10.1016/j.ijhcs.2016.09.008.
Ludwig, T., Reuter, C., Siebigteroth, T., Pipek, V., 2015. CrowdMonitor: mobile crowd sensing for assessing physical and digital activities of citizens during emergencies. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 4083–4092. https://doi.org/10.1145/24702123.2702265.
Marcotte, E., 2017. Responsive Web Design: A Book Apart N 4. Edition Eyrolles.
Mator, J.D., Lehman, W.E., McManus, W., Powers, S., Tiller, L., Unverricht, J.R., Still, J. D., 2020. Usability: adoption, measurement, value. Human factors. The Journal of the Human Factors and Ergonomics Society, 00187208199509. https://doi.org/10.1177/00187208199509.
Merchant, R.M., Elmer, S., Lurie, N., 2011. Integrating social media into emergency-preparedness efforts. N. Engl. J. Med. 365 (4), 289–291.
Mital, A., Pennathur, A., 2000. Perspectives on designing human interfaces for automated systems. In: Handbook of Industrial Automation. CRC Press, pp. 749–792.
Nielsen, J., 1999. User interface directions for the web. Commun. ACM 42 (1), 65–72.
Nielsen, J., 2000. Designing Web Usability. New Riders Publishing.
Nielsen, J., 2001. Ten usability heuristics. In: ACHI 2011: the Fourth International Conference on Advances in Computer-Human Interactions.
Nielsen, J., 2005. Ten usability heuristics. http://www.nngroup.com/articles/ten-usability-heuristics/.
Nielsen, J., Tahir, M., 2001. Homepage Usability: 50 Websites Deconstructed. New Riders Publishing.
Palmer, J.W., 2002. Web site usability, design, and performance metrics. Inf. Syst. Res. 13 (2), 151–167.
Park, J., Han, S.H., Park, J., Park, J., Kwak, J., Lee, M., Jeong, D.Y., 2018. Development of a web-based user experience evaluation system for home appliances. Int. J. Ind. Ergon. 67, 216–228. https://doi.org/10.1016/j.ergon.2018.05.017.
Pew Research Center, 2017. Mobile fact sheet. https://www.pewresearch.org/internet/fact-sheet/mobile/.
Proctor, R.W., Van Zandt, T., 2018. Human Factors in Simple and Complex Systems. CRC press.
Rusu, C., Ronzagliolo, S., Rusu, V., Collanzo, C., 2011. A methodology to establish usability heuristics. 4. In: ACHI 2011: the Fourth International Conference on Advances in Computer-Human Interactions.
Sayyad Abdi, E., Partridge, H., Bruce, C., 2016. Web designers and developers experiences of information literacy: a phenomenographic study. Libr. Inf. Sci. Res. 38 (4), 353–359. https://doi.org/10.1016/j.lisr.2016.11.008.
Scott, J.K., 2005. Assessing the quality of municipal government Web sites. State Local Govern. Rev. 37 (2), 151–165.
Shneiderman, B., Lewitt, M., others, 2006. Research-based Web Design & Usability Guidelines. Department of Health and Human Services, Washington DC.
Singh, S.N., Dalal, N., Spears, N., 2005. Understanding web home page perception. Eur. J. Inf. Syst. 14 (3), 288–302.
Solano, A., Collanzo, C.A., Rusu, C., Fardoun, H.M., 2016. Combinations of methods for collaborative evaluation of the usability of interactive software systems. Advances in Human-Computer Interaction, 2016 1–16. https://doi.org/10.1155/2016/4089520.
The New York Times, 2020. The New York Times (nytimes.com).
Trukenbrod, A.K., Backhaus, N., Thomaschke, R., 2020. Measuring subjectively experienced time in usability and user experience testing scenarios. Int. J. Hum. Comput. Stud. 138, 102399. https://doi.org/10.1016/j.ijhcs.2020.102399.
USDA Geographical Regions, 2020. USDA regions. https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/docs/regions/.
Venkatesh, V., Hoehle, H., Aljafari, R., 2017. A usability study of the Obamacare website: evaluation and recommendations. Govern. Inf. Q. 34 (2), 199–210. https://doi.org/10.1016/j.giq.2017.01.003.
Virzi, R.A., 1992. Refining the test phase of usability evaluation: how many subjects is enough? Hum. Factors 34 (4), 457–468.
Youngblood, S.A., Youngblood, N.E., 2018. Usability, content, and connections: how county-level Alabama emergency management agencies communicate with their online public. Govern. Inf. Q. 35 (1), 50–60. https://doi.org/10.1016/j. giq.2017.12.001.