Do Advisories, Warnings, or Color Coding Matter to Risk Perception and Precautionary Decisions?

GALA GULACSIK,[a] SUSAN L. JOSLYN,[a] JOHN ROBINSON,[a] AND CHAO QIN[a]

[a] *University of Washington, Seattle, Washington*

(Manuscript received 26 April 2021, in final form 10 February 2022)

ABSTRACT: There are lingering questions about the effectiveness of the watch, warning, and advisory system (WWA) used to convey weather threats in the United States. Recently there has been a shift toward alternative communication strategies such as the impact-based forecast. The study reported here compared users' interpretation of a color-coded impact-based prototype designed for email briefings, to a legacy WWA format. Participants, including emergency managers and members of the public, saw a weather briefing and rated event likelihood, severity, damage, and population affected. Then they recommended emergency response actions. Each briefing described the severity of the weather event and the degree of impact on population and property. In one condition a color-coded impacts scale was added to the text description. In another, an advisory and/or warning was added to the text description. These were compared with the text-only control. Both emergency managers and members of the public provided higher ratings for event likelihood, severity, damage, and population affected and recommended a greater response for higher impact levels regardless of format. For both groups, the color-coded format decreased ratings for lower-impact events. Among members of the public, the color-coded format also led to increases for many ratings and greater response at higher levels relative to the other two conditions. However, the highest ratings among members of the public were in the WWA condition. Somewhat surprisingly, the only effect of the WWA format on emergency managers was to *reduce* action recommendations, probably because of the inclusion of the "advisory" in some briefings.

KEYWORDS: Communications/decision-making; Decision-making; Decision support

1. Introduction

In the United States, since 1966 the National Weather Service (NWS) has communicated severe weather risk using the categories of the watch, warning, and advisory (WWA; Corfidi 2010). Each category is intended to indicate a unique combination of the severity and likelihood of the event. A watch indicates a *potential* for a dangerous hazard. A warning indicates an *imminent/occurring* life-threatening hazard. An advisory indicates an *imminent/occurring* less serious hazard (NWS 2017).

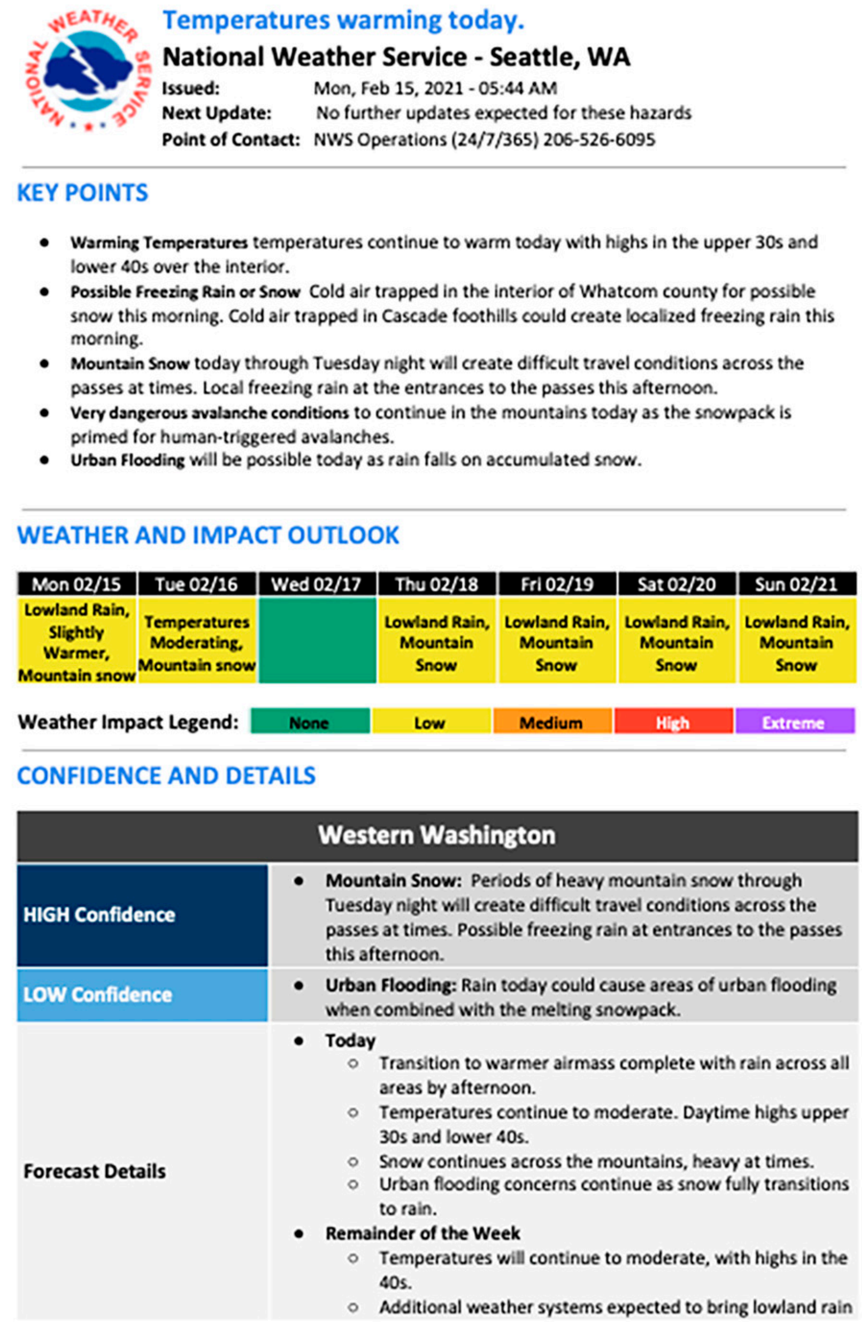
However, for the last several years there has been an ongoing discussion about the usefulness of the WWA system due to persistent misinterpretations and confusions among members of the public as well as emergency managers (NWS 2014, 2018; Morss et al. 2016; Weaver et al. 2017). Indeed, the current plan is to phase out the "advisory" category by 2024 (NWS 2021). Moreover, some research suggests that interpretations of the precise levels of event likelihood indicated by the WWA categories tend to differ between members of the public and the forecasters who issue them (Morss et al. 2016). However, the same research indicated that when informational context (e.g., quantity and rate of rainfall; storm track) accompanied flash-flood warnings, the majority of respondents anticipated taking similar protective action (evacuating or assessing risk). In addition, an assessment of the tornado warning system following the deadly 2011 Joplin tornado found that most people sought additional information outside the weather service before taking shelter (NWS Central Region 2011). This suggests that the situational context may be at least as important to users as the informational terminology, if not more so. In summary, it is clear that the WWA terminology remains confusing for many and may be less important for weather-related decisions than other kinds of information.

Many believe an impacts-based approach to risk communication, framing risk communication in terms of the impact to end users, will improve both understanding and user decisions (NWS 2014; Williams et al. 2017). In response, in recent years the National Weather Service has made an effort to communicate impact levels as well as to distinguish levels of urgency between events (NWS Central Region 2011; NWS 2016), although there is as yet no standardized format. One instantiation of this policy is the Impact-Based Decision Support Services (IDSS) email briefing format employed by the National Weather Service Western Region office (NWS 2020). The briefing consists of three sections (see Fig. 1). At the top, the *key points* section summarizes the threat. In the center, the *weather and impact outlook* section includes a 7-day outlook table. It uses five levels, color coded as green, yellow, orange, red, and purple and accompanied by adjectives to indicate the degree of impact (e.g., low, medium, high, extreme). Just below, a *confidence and details* section categorizes forecast events as low, moderate, or high forecaster confidence. This is followed by a description that conveys the severity of the weather event, the impacts (e.g., power outages, property damage, and temporary impairments to roads) and the geographic area involved.

The impact-level email weather briefings are intended to provide tailored operational decision support for emergency managers and other key partners. However, the effect of the

Corresponding author: Chao Qin, robertqc@uw.edu

DOI: 10.1175/WCAS-D-21-0070.1



Temperatures warming today.

National Weather Service - Seattle, WA

Issued: Mon, Feb 15, 2021 - 05:44 AM
Next Update: No further updates expected for these hazards
Point of Contact: NWS Operations (24/7/365) 206-526-6095

KEY POINTS

- **Warming Temperatures** temperatures continue to warm today with highs in the upper 30s and lower 40s over the interior.
- **Possible Freezing Rain or Snow** Cold air trapped in the interior of Whatcom county for possible snow this morning. Cold air trapped in Cascade foothills could create localized freezing rain this morning.
- **Mountain Snow** today through Tuesday night will create difficult travel conditions across the passes at times. Local freezing rain at the entrances to the passes this afternoon.
- **Very dangerous avalanche conditions** to continue in the mountains today as the snowpack is primed for human-triggered avalanches.
- **Urban Flooding** will be possible today as rain falls on accumulated snow.

WEATHER AND IMPACT OUTLOOK

Mon 02/15	Tue 02/16	Wed 02/17	Thu 02/18	Fri 02/19	Sat 02/20	Sun 02/21
Lowland Rain, Slightly Warmer, Mountain snow	Temperatures Moderating, Mountain snow		Lowland Rain, Mountain Snow	Lowland Rain, Mountain Snow	Lowland Rain, Mountain Snow	Lowland Rain, Mountain Snow

Weather Impact Legend: None | Low | Medium | High | Extreme

CONFIDENCE AND DETAILS

Western Washington	
HIGH Confidence	• **Mountain Snow:** Periods of heavy mountain snow through Tuesday night will create difficult travel conditions across the passes at times. Possible freezing rain at entrances to the passes this afternoon.
LOW Confidence	• **Urban Flooding:** Rain today could cause areas of urban flooding when combined with the melting snowpack.
Forecast Details	• **Today** ○ Transition to warmer airmass complete with rain across all areas by afternoon. ○ Temperatures continue to moderate. Daytime highs upper 30s and lower 40s. ○ Snow continues across the mountains, heavy at times. ○ Urban flooding concerns continue as snow fully transitions to rain. • **Remainder of the Week** ○ Temperatures will continue to moderate, with highs in the 40s. ○ Additional weather systems expected to bring lowland rain

FIG. 1. National Weather Service IDSS email weather briefing.

color-coded IDSS email briefing on users' understanding of the risk posed is at present an open question.[1] Here we define risk as a combination of outcome severity and event likelihood (Weber and Milliman 1997). The simple and uniform organization of the IDSS email briefings is likely be helpful as it allows users to build a schema that could facilitate comprehension (Bartlett 1932). However, one of the most salient aspects of the new format is the color coding. At the time during which this experiment was conducted, the color-coded scales were provided to recipients without accompanying definitions beyond the adjectives described above, although more complete definitions were available on the NWS website (see appendix A). Thus, users' understanding of the risk level

[1] The previous version of the email briefings issued by National Weather Service Western Region included similar components without the color-coded header.

intended by the color coding is an important and open question because behavioral research suggests that there can be issues with understanding color coding. The majority of evidence demonstrates inconsistencies in how people rank/order color-coded risk (Chapanis 1994; Wogalter et al. 1995; Rashid and Wogalter 1997; Mayhorn et al. 2004; Bryant et al. 2014) suggesting that colors mean different things to different individuals. In addition, warnings with color are perceived as more hazardous overall (Braun et al. 1995). However, as far as we are aware, little research investigates the precise level of risk conveyed by color coding. For instance, a user may understand that orange conveys a greater risk than green, however, the same person may overestimate the likelihood conveyed by orange (e.g., 80%) as compared with the intended likelihood (40%). As a result, people may perceive the risk to be different than what is intended or systematically greater than what is intended when color coding is added. This could, in turn, affect the decisions that people make based on the forecast.

The two studies reported here were conducted to determine whether there are differences in risk perception and emergency decisions using the color-coded IDSS email format versus the legacy WWA. In both experiments, participants read a series of email briefings describing weather events. Some also included the color-coded impacts scale while others also included an advisory or a warning. Both conditions were compared with the text-only description that served as the control. In experiment 1 a group of emergency managers served as participants. In experiment 2, in an effort to test a larger sample and increase the power for inferential analyses, participants were members of the public. This also allowed us to compare understanding across these two levels of expertise.

2. Experiment 1: Emergency managers

a. Method

1) PARTICIPANTS

Participants ($n = 17$) were recruited via email and onsite at the Washington Emergency Managers Association 2019 Annual Meeting. Participants included public safety officials from local, state, and federal government and ranged in experience from 1 to 20 years with an average of 10.6 years. Participants ranged in age from 27 to 64 years old (mean $M = 48.2$; standard deviation SD = 11.3), and 18% of them were females.

2) PROCEDURE

The experiment was conducted on Qualtrics, a web-based survey platform, in November 2019, accessed through an anonymous link on a computer.[2] After providing informed consent, participants were instructed to assume the role of a consultant to an emergency management organization in western Washington. The goal was to provide advice to coordinate emergency responses for communities under threat

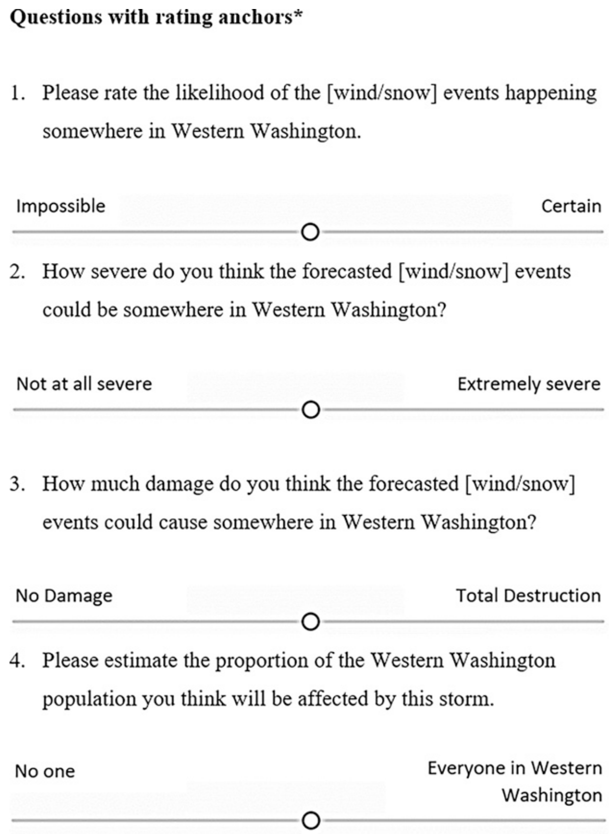


FIG. 2. Questions gauging understanding of weather briefings for both experiments 1 and 2.

of severe weather. Background information was provided, including maps of local counties and major cities as well as population density (see appendix B).

Next, participants received a series of weather briefings that were described as occurring weeks apart. To gauge their understanding of the key components of weather risk that were included in all of the briefings, participants rated the *likelihood*, *severity*, resulting *damage* of the event described, and the *proportion of the population* that might be affected. Ratings were made by dragging a handle to the desired position on and unmarked line between two anchors. This is known as a visual analog scale (VAS; see Fig. 2).

On the same page, participants decided whether a response was necessary for the event described in the email briefing. If the participant selected "Yes, I recommend the following actions," a list of four possible actions[3] was shown (see Table 1). Instructions indicated that items lower on the list were appropriate for more serious events but involved greater cost in terms of time and resources. Therefore, they should only recommend actions they believed were necessary. They could choose as many actions as they deemed appropriate. For each action,

[2] At the outset, the website stated that the survey was not intended for use on a smartphone.

[3] Actions were drawn from an unpublished interview study conducted on emergency managers' decision-making process with respect to a 2016 wind event in Washington State.

TABLE 1. Action recommendations for both experiment 1 and experiment 2.

	Action for recommendation
Lowest cost	Share information on social media
↓	Share information with local organizations (schools, hospitals, shelters, news media, private companies, etc.)
↓	Put local response departments on alert (fire department on standby, road/plow crews on standby, etc.)
Highest cost	Activate the EOC or other active monitoring situation

participants clicked on a radio button labeled "No, I do not advise this action" or "Yes, I advise this action." Other actions not listed could be added in an open-ended text box, although none differed substantially from those provided and will not be discussed further. If participants selected "No, I do not want to recommend any action," the trial ended, and the next weather briefing was presented. After participants completed the decision task, they were asked to rate how well they understood the email briefings on a VAS from *not at all* to *completely*.

3) STIMULI

Participants saw a sequence of 16 email briefings each describing a wind or a snow event. Wind and snow briefings were equivalent in terms of word count, impact levels, and position in trial order. In the context of the experiment, each briefing indicated only one day during which an event was expected, occurring within 24 h of issuance,[4] although time frame and lead time can vary in actual briefings. The briefings used here were based on actual emails distributed by National Weather Service Seattle office between February 2018 and 2019. Each participant received two briefing formats—text only, used as a control, and one of two experimental formats described below. The first eight briefings were in the text-only format. They included verbal descriptions of the weather event organized into a simplified version of the email briefing described above (Fig. 1) comprising two sections labeled "key points" and "details." The key-points section described the severity of the hazard and the geographic extent of its impact. The details section included information about the severity, impacts and onset of the weather event at specific locations (see Figs. 3a,c).

Note that the text-only format included Figs. 3a and 3c; the advisory/warning format included Figs. 3a, 3c, and 3d; and the outlook format included Figs. 3a–c.

For the remaining eight briefings, participants were randomly assigned to one of two experimental conditions: the advisory/warning format or the outlook format. For both, the description of the event was identical to that in the text-only condition.

For the advisory/warning format in addition to the two sections described above (key points and details) there was a statement at the bottom of the briefing with the advisory or warning designation (Fig. 3d).[5] To reduce the number of trials required per participant, only events with the designations "advisory" and "warning" were used here.

Participants in the outlook condition saw the same two sections (key points and details) with a color-coded 7-day outlook band across the top (see Fig. 3b) showing Thursday–Wednesday from left to right. Thursday was designated as "today," the day upon which the briefing was issued. The day for which the weather event was expected was color-coded to indicate impact level, yellow indicating low, orange indicating medium, red indicating high or purple indicating extreme. A key, describing these pairings was provided on the briefing. Fourteen events occurred on "Friday (tomorrow)" and two on "Thursday (today)," so that all would have similar time frames. The days preceding and following impact days were colored green to indicate that no impacts were expected during those times.

The event descriptions were classified according to the impact-level definitions (see appendix A) used by the NWS Western Region, although the impact-level itself was shown only in the outlook condition. Descriptions included relevant impacts such as power outages, property damage, and temporary impairments to roads. Those classified as low and medium impact corresponded to an advisory in the advisory/warning condition. All high-impact events corresponded to a warning in the advisory/warning condition. However, one event in the high-impact category also included an advisory for a location expected to be less affected. All extreme events (purple) corresponded to a warning in the advisory/warning condition. There were four briefings in each impact level.

The descriptions of 16 unique weather events were divided into two sets of email briefings (referred to as set 1 and set 2). Both sets included two low, two medium, two high, and two extreme level briefings. At every impact level, half were snow events and the other half were wind events.[6] The briefings were placed in a fixed, semirandom order that did not allow for consecutive stimuli of the same hazard and impact-level combination (e.g., no consecutive moderate snow) to encourage viewing them as independent events. See appendix C for a list of the briefings shown in the order displayed. Each set was counterbalanced such that participants were randomly assigned to see set 1 first and set 2 second ($n = 11$) or vice

[4] In fact, such briefings can be sent out several days in advance and have multiple updates; however, variations in lead time have been shown to impact perceptions of likelihood (Joslyn and Savelli 2010) so time frame was held constant here to avoid that extraneous variable, which could obscure effects resulting from format.

[5] These were similar in format to those used prior to the introduction of the IDSS color-coded format tested here.

[6] A comparison between snow and wind events on participants' ratings (likelihood, severity, damage, and proportion of population affected) showed no difference in response due to different weather events.

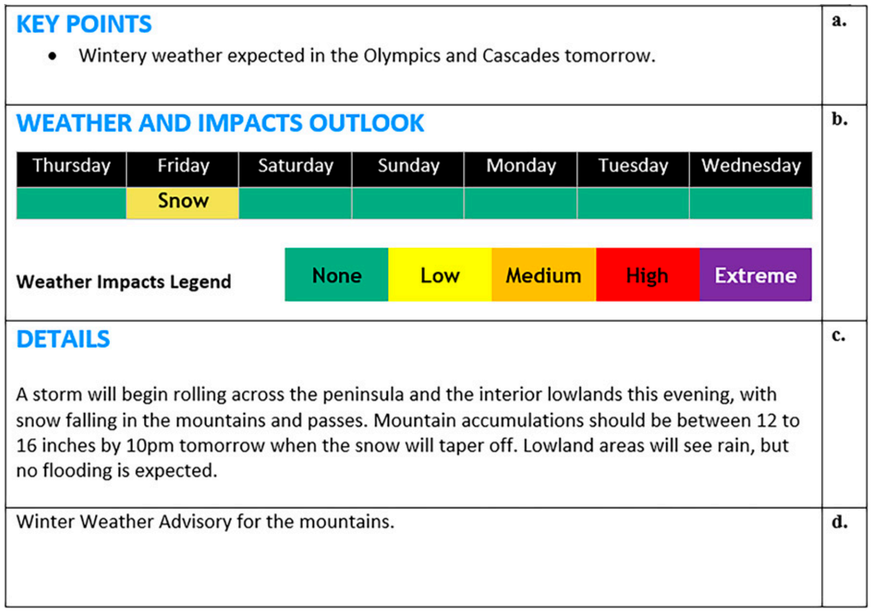


FIG. 3. (a)–(d) Example of weather briefing in all formats for both experiment 1 and experiment 2.

versa ($n = 6$), regardless of the experimental format shown in set 2. The first set was always shown in the text-only format. Thus, all events were shown in all formats.

4) DESIGN

Experiment 1 employed a within-groups design. There were two independent variables: Briefing format and impact level. Briefing format had three levels: advisory/warning, outlook, and text only. All participants saw the control (text only) and one of the two experimental formats (outlook, advisory/warning). Impact level was manipulated within groups and had four levels: low, medium, high, and extreme.

b. Results

To determine the impact of communication format on risk perception and decision-making, a series of factorial repeated measures analyses of variance (ANOVA) were conducted on participants ratings (likelihood, severity, damage, and proportion of population affected), and action recommendations, with two independent variables: impact level (within: low, medium, high, extreme) and briefing format (within: text only vs advisory/warning or outlook). Ratings were summarized as the percentage of the VAS line between the left anchor and the position to which participants moved the handle (see Fig. 2). Effect sizes for ANOVAs were reported as partial eta-squared values. Because each participant saw the text-only format and one of the two experimental formats, all comparisons with the text-only format were made within groups and the two experimental formats were not compared directly with one another. Thus, there are two ANOVAs for each dependent variable, one for the group that saw the advisory/warning format ($n = 7$) and another for the group that saw the outlook format ($n = 10$). Where appropriate, P values were corrected (more conservative) for violations of sphericity by adjusting the degrees of freedom associated with the F statistic according to the Greenhouse–Geisser estimate of sphericity. Understanding (self rating) of forecasts was analyzed between groups using an independent t test.

1) LIKELIHOOD RATING

First, we conducted an ANOVA on event likelihood ratings including the advisory/warning format. Likelihood ratings increased with impact level (see Fig. 4) and the main effect was marginally significant [$F(1.56, 9.36) = 4.06$, $p = 0.061$, and $\eta_P^2 = 0.40$]. The effect of format [$F(1, 6) = 0.13$, $p = 0.73$, and $\eta_P^2 = 0.02$] and the interaction between format and impact level [$F(3, 18) = 0.54$, $p = 0.662$, and $\eta_P^2 = 0.08$] failed to reach significance. See Fig. 4 for mean likelihood ratings.

Next, we conducted an ANOVA on event likelihood ratings with the outlook format. Here, likelihood ratings increased significantly with impact level (see Fig. 4) [$F(3, 27) = 21.07$, $p < 0.001$, and $\eta_P^2 = 0.70$]. There was also a main effect of

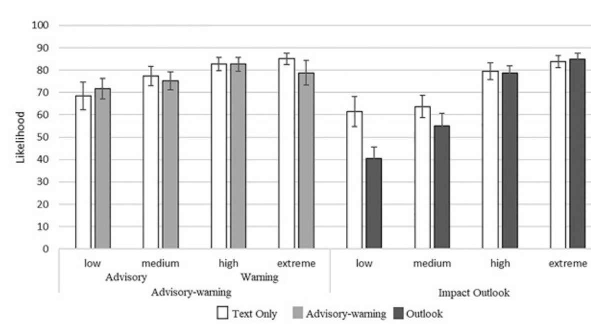


FIG. 4. Likelihood ratings in the text only + advisory/warning and text only + outlook conditions for experiment 1.

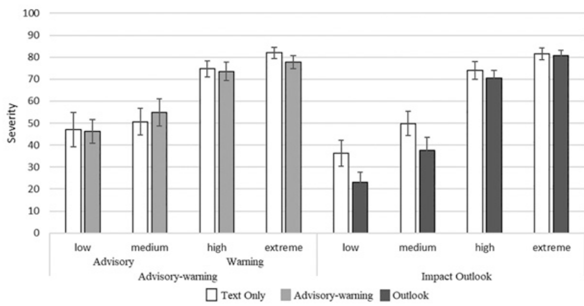


FIG. 5. Severity ratings in the text only + advisory/warning and text only + outlook conditions for experiment 1.

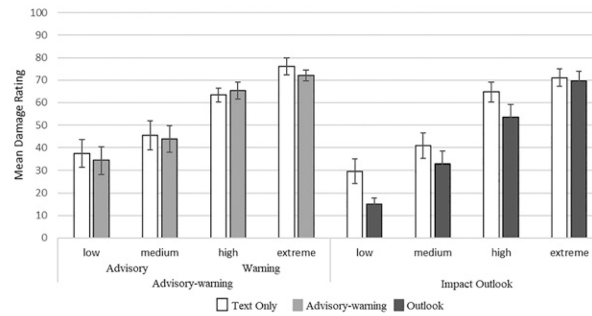


FIG. 6. Mean damage ratings in the text only + advisory/warning and text only + outlook conditions for experiment 1.

format [$F(1, 9) = 5.80$, $p = 0.039$, and $\eta_P^2 = 0.39$]. Mean likelihood rating was significantly lower in the outlook format ($M = 64.76$; SD = 26.22) than in the text-only format ($M = 72.37$; SD = 26.22). In addition, there was a significant interaction between impact level and format such that likelihood ratings in the outlook format were lower than in the text format at the low and medium impact levels but not at the high and extreme levels where the two formats were rated similarly [$F(3, 27) = 3.15$, $p = 0.041$, and $\eta_P^2 = 0.26$]. Thus, although neither the effect of impact level nor format reached significance in the advisory/warning analysis, both were significant in the outlook condition, which reduced likelihood ratings at the lower levels relative to text-only control.

2) SEVERITY RATING

For the ANOVA on severity ratings that included the advisory/warning format, ratings increased significantly with impact level (see Fig. 5) [$F(3, 18) = 27.41$, $p < 0.001$, and $\eta_P^2 = 0.82$]. However, the effects of format [$F(1, 6) = 0.02$, $p = 0.89$, and $\eta_P^2 = 0.00$] and the interaction between impact level and format [$F(3,18) = 0.29$, $p = 0.83$, and $\eta_P^2 = 0.05$] failed to reach significance.

For the ANOVA including the outlook format, severity ratings also increased significantly with impact level (see Fig. 5) [$F(3, 27) = 72.91$, $p < 0.0001$, and $\eta_P^2 = 0.89$]. In addition, the outlook format led to significantly lower severity ratings ($M = 53.01$; SD = 29.96) than did text-only format ($M = 60.41$; SD = 27.72), with $F(1, 9) = 5.17$, $p = 0.049$, and $\eta_P^2 = 0.36$. The interaction between impact level and format failed to reach significance [$F(2.07, 18.63) = 2.47$, $p = 0.11$, and $\eta_P^2 = 0.22$], although the trend was similar to the likelihood analysis with greater reduction at lower impact levels. Thus, for severity ratings, the effect of impact level was seen in both formats, although only in the outlook condition did format matter. The outlook format reduced severity ratings relative to text-only control.

3) DAMAGE RATING

For the ANOVA on damage ratings that included the advisory/warning format, damage ratings increased significantly with impact level (see Fig. 6) [$F(3, 18) = 33.89$, $p < 0.001$, and $\eta_P^2 = 0.85$]. Neither the effect of format [$F(1, 6) = 0.25$, $p = 0.64$, and $\eta_P^2 = 0.04$] nor the interaction between impact level and format [$F(3,18) = 0.20$, $p = 0.89$, and $\eta_P^2 = 0.03$] reached significance.

For the ANOVA including the outlook format, again, damage ratings increased significantly with impact level (see Fig. 6) [$F(1.38, 12.42) = 58.01$, $p < 0.001$, and $\eta_P^2 = 0.87$]. In addition, damage ratings in the outlook format ($M = 42.74$; SD = 29.36) were significantly *lower* than in the text format ($M = 51.58$; SD = 27.42), with $F(1, 9) = 5.67$, $p = 0.041$, and $\eta_P^2 = 0.39$. The interaction between impact level and format failed to reach significance [$F(3, 27) = 2.27$, $p = 0.10$, and $\eta_P^2 = 0.20$]. In sum, the effect of impact level on damage was seen in both conditions; however, the effect of format was seen only in the outlook condition, which had reduced ratings relative to text-only control.

4) PROPORTION OF POPULATION AFFECTED

For the ANOVA on the proportion of the population affected that included the advisory/warning condition, proportion ratings increased significantly with impact level (see Fig. 7) [$F(3, 18) = 25.32$, $p < 0.0001$, and $\eta_P^2 = 0.81$]. Neither the effect of format [$F(1, 6) = 0.18$, $p = 0.69$, and $\eta_P^2 = 0.03$] nor the interaction between impact level and format [$F(3, 18) = 0.36$, $p = 0.79$, and $\eta_P^2 = 0.06$] reached significance.

For the ANOVA including the outlook format, again, proportion ratings increased significantly with impact level (see Fig. 7) [$F(3, 27) = 31.92$, $p < 0.001$, and $\eta_P^2 = 0.78$]. The proportion ratings in the outlook format were lower ($M = 0.47$; SD = 0.25) than in the text-only format ($M = 0.55$;

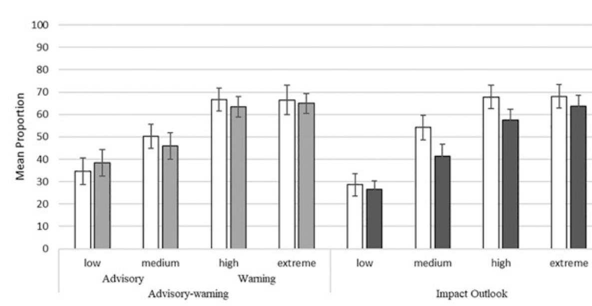


FIG. 7. Mean proportion of population affected in text only + advisory/warning and text only + outlook conditions for experiment 1.

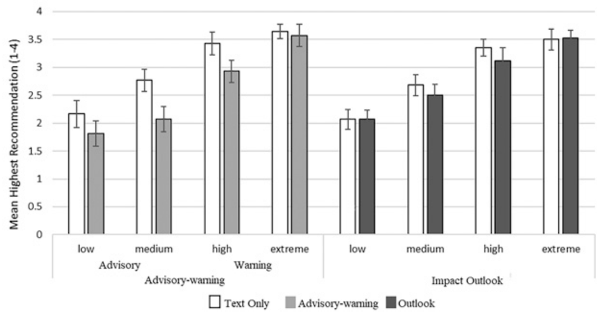


FIG. 8. Mean response cost (of recommended action) in the text only + advisory/warning and text only + outlook conditions for experiment 1.

SD = 0.28), but the difference failed to reach significance [$F(1, 9) = 3.94$, $p = 0.08$, and $\eta_P^2 = 0.30$]. The interaction between impact level and format failed to reach significance [$F(1.41, 12.69) = 58$, $p = 0.52$, and $\eta_P^2 = 0.06$]. Thus, as with likelihood and severity ratings, the effect of impact level on proportion of the population was seen in both conditions.

5) ACTION RECOMMENDATIONS

Next, we examined the effect of format and impact level on participants' action recommendations. For this analysis, recommended actions were categorized from 1 (lowest cost action) to 4 (highest cost action). A value of 0 was assigned to decisions not to act at all (see Table 1 in the methods section, section 2a). For every email briefing the highest number chosen by the participant (highest response cost) was assigned as the score for that trial. Then, means were calculated across trials by impact level and format. A pair of factorial repeated measures ANOVAs were conducted on mean action scores with impact level (low, medium, high, and extreme) and format (control, advisory/warning, outlook) as the within-groups independent variables.

For the ANOVA including the advisory/warning format, action recommendations increased significantly with impact level (see Fig. 8) [$F(3, 18) = 30.30$, $p < 0.001$, and $\eta_P^2 = 0.83$]. For this analysis, mean action recommendations in the advisory/warning format were significantly lower ($M = 2.50$; SD = 1.16) than in the text-only format ($M = 2.88$; SD = 1.11), with $F(1, 6) = 13.50$, $p = 0.01$, and $\eta_P^2 = 0.69$. The interaction between impact level and format failed to reach significance [$F(3, 18) = 1.27$, $p = 0.32$, and $\eta_P^2 = 0.17$] (see Fig. 8).

For the ANOVA including the outlook format, again action recommendations increased significantly with impact level (see Fig. 8) [$F(3, 27) = 40.68$, $p < 0.001$, and $\eta_P^2 = 0.82$]. There was also a significant effect of format such that action recommendations in the outlook format ($M = 2.28$; SD = 1.45) were significantly *lower* than in the text-only format ($M = 2.74$; SD = 1.19), with $F(1, 9) = 6.35$, $p = 0.03$, and $\eta_P^2 = 0.41$. The interaction between impact level and format failed to reach significance [$F(3, 27) = 2.12$, $p = 0.12$, $\eta_P^2 = 0.19$]. Thus, action recommendations increased with impact level regardless of format. However, unlike the previous dependent variables, the effect of format on action recommendations was observed in both format conditions. Surprisingly, both the color coding in the outlook condition and the addition of an advisory or a warning *reduced* action recommendations on average.

6) EFFECT OF IMPACT LEVEL

The previous analyses suggest that the effect of impact level was conveyed by the text alone. To verify this, we examined all ratings (likelihood, severity, damage, and percent of population affected) as well as action recommendation in the text-only condition. Indeed, all five ANOVAs showed a significant effect of impact level with higher ratings for higher levels [*likelihood*: $F(1.83, 29.28) = 12.07$, $p < 0.001$, and $\eta_P^2 = 0.43$; *severity*: $F(3, 48) = 50.27$, $p < 0.0001$, and $\eta_P^2 = 0.76$; *damage*: $F(3, 48) = 45.84$, $p < 0.001$, and $\eta_P^2 = 0.74$; *proportion of population affected*: $F(3, 48) = 25.22$, $p < 0.001$, and $\eta_P^2 = 0.61$; *action recommendations*: $F(1.68, 26.88) = 55.27$, $p < 0.001$, and $\eta_P^2 = 0.78$]. This suggests that the text alone conveyed the increase in impact levels.

7) UNDERSTANDING (SELF) RATING

Ratings for (self reported) understanding of email briefings were compared between the two experimental conditions. Although ratings did not differ by format, participants' rated their own understanding as high. Understanding in the advisory/warning condition ($M = 80.14$; SD = 11.14) was slightly but not significantly lower than in the outlook condition ($M = 81.3$; SD = 15.5), with $t(15) = -0.17$ and $p = 0.87$.

c. *Discussion*

These results suggest that risk perception among emergency managers was influenced mainly by the descriptions in the text rather than by the inclusion of either a warning or advisory designation. Ratings of likelihood, severity, damage, and proportion of the population, all increased significantly and systematically with the designated levels in the text-only format. Somewhat surprisingly, the effect of the color-coded outlook format was mainly to reduce risk perception, primarily at the lower levels. This contradicts the literature indicating that color coding increases perceived hazardousness (Wogalter et al. 2002) and risk perception overall (Braun et al. 1995). Our result suggests instead that the precise impact of color may depend on the context. The significant interaction between format and impact level for likelihood ratings, revealed that the observed reduction was mainly at the lower levels, color-coded yellow and orange. Thus, there may be an advantage for the color coding in that it allows for greater differentiation between the lower and higher likelihood levels.

Also, somewhat surprisingly, the only effect on action recommendations of either format was to reduce recommendations. Although the interaction failed to reach significance in this small sample, the reduction appears to be again mainly at the lower impact levels where the yellow color or the advisory was shown, perhaps signaling less urgency to these experienced participants.

To determine whether these results would hold in a larger sample, and whether there were differences between expert and nonexpert decision-makers, we conducted a similar study among a group of online nonexpert participants.

3. Experiment 2: Members of the public

a. Method

1) Participants

Participants ($n = 117$), all U.S. residents, were recruited via Amazon Mechanical Turk (Mturk), a data-collection domain administered by Amazon. Participants ranged in age from 19 to 69 years old ($M = 37.7$; SD = 11.42), and 41% were females.

2) Procedure

The experiment, conducted in November 2019, was similar to experiment 1 in terms of platform and procedure, including the same instructions and background information. However, there was additional explanation about the function of emergency management and emergency operations centers (EOCs) because these were nonexpert participants (see appendix B). In addition, Mturk participants received an orientation to the layout and content of the email briefings prior to the experimental task because, unlike emergency managers, they had no previous exposure to this format. For the same reason, in the experimental conditions, participants were provided with definitions of advisory and warning, or the table of impact-level definitions that they could review at any point (appendix A). Participants in experiment 2 were asked the same questions (Fig. 2) as the emergency managers in experiment 1. In addition, after each of the 12 trials, attention check questions were shown asking participants to name the hazardous weather event or quantity (e.g., wind speed) described in the previous briefing.

3) Stimuli

Participants saw a sequence of 12 email briefings (a subset of the 16 briefings in experiment 1) each one for either a wind or snow event.[7] Each participant received email briefings in either the control (text only), advisory/warning, or outlook (Figs. 3a–c) format.

As with experiment 1, the email briefings were divided into two sets of 12. In experiment 2, participants were randomly assigned to one of three format groups, the 1) control, 2) advisory/warning, or 3) outlook format. They were also randomly assigned to see text from either set 1 or set 2. Thus, both briefing text sets, were paired with each format, although the pairing differed by participant. Both sets included two low-, four medium-, four high-, and two extreme-level briefings. The briefings were placed in a fixed semirandom order that did not allow for consecutive stimuli of the same hazard and impact-level combination (e.g., no consecutive moderate snow) to avoid the impression that events were related. See appendix C for the order of briefings.

[7] Participants' ratings were significantly higher for snow than wind for severity [$t(232) = 2.65$; $p = 0.01$], damage [$t(232) = 1.43$; $p = 0.15$], and proportion of population affected [$t(232) = 2.58$; $p = 0.01$].

4) Design

The experiment employed a 3×4 full factorial design. Briefing format was manipulated between groups and had three levels: advisory/warning, outlook, and text-only control. Impact level was a within-groups factor and had four levels: low, medium, high, and extreme.

b. Results

To determine the impact of communication format on risk perception and decision-making, a series of ANOVA was conducted on participant ratings (likelihood, severity, damage, and percent of population affected), and action recommendations, with impact level (low, medium, high, and extreme) as the within-groups independent variable. Most also included briefing format (control, outlook, and advisory/warning) as the between-groups independent variables. Planned contrasts were corrected for familywise error using the Bonferroni correction ($\alpha = 0.0167$).

To determine whether the effect of impact level was conveyed by the text alone, as had been the case among emergency managers in experiment 1, we first examined all ratings (likelihood, severity, damage, and percent of population affected) as well as action recommendations in the text-only control condition. Indeed, all ANOVAs showed a significant effect of impact level with higher ratings for higher levels [*likelihood*: $F(2.28, 109.44) = 31.38$, $p < 0.001$, and $\eta_P^2 = 0.40$; *severity*: $F(2.49, 119.52) = 118.45$, $p < 0.0001$, and $\eta_P^2 = 0.71$; *damage*: $F(3, 144) = 73.48$, $p < 0.001$, and $\eta_P^2 = 0.60$; *percent of population affected*: $F(2.4, 115.2) = 45$, $p < 0.0001$, and $\eta_P^2 = 0.22$; *action recommendations*: $F(2.19, 105.12) = 58.78$, $p < 0.001$, and $\eta_P^2 = 0.55$]. This suggests that, as with the small group of emergency managers, the text alone conveyed the increase in impact levels to members of the public.

1) Likelihood rating

Next, we asked whether the differences in email briefing format influenced any of these same dependent variables and/or interacted with impact levels. For the ANOVA on likelihood rating, neither the main effect of format [$F(2, 114) = 0.19$, $p = 0.83$, and $\eta_P^2 = 0.00$] nor the interaction between format and impact level [$F(4.26, 242.82) = 0.44$, $p = 0.79$, and $\eta_P^2 = 0.01$] reached significance (see Fig. 9). There was a significant main effect of impact level (see Fig. 9) [$F(2.13, 242.82) = 73.42$, $p < 0.0001$, and $\eta_P^2 = 0.39$]. Thus, although participants indicated higher likelihood with higher impact levels neither color-coded outlook format nor advisory/warning label influenced likelihood ratings.

2) Severity rating

In the ANOVA on severity ratings, there was a significant main effect of format [$F(2, 114) = 3.34$, $p = 0.04$, and $\eta_P^2 = 0.06$]. The mean severity rating in the advisory/warning condition ($M = 58.07$; SD = 25.07) was significantly higher than in the control condition ($M = 49.72$; SD = 26.43), with $t(86) = -2.73$ and $p = 0.008$. Neither differed significantly from the outlook condition ($M = 54.03$; SD = 28.50), which

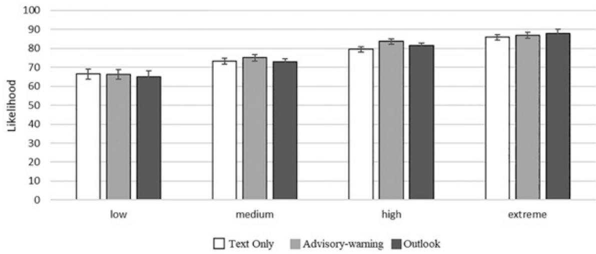


FIG. 9. Likelihood ratings in all formats for experiment 2.

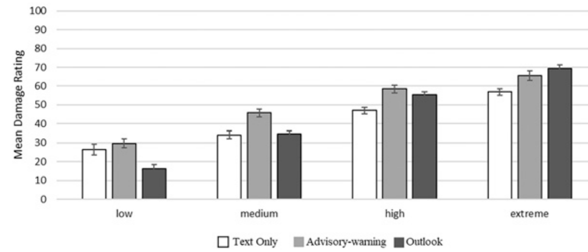


FIG. 11. Damage ratings in all formats for experiment 2.

fell in between. There was also a significant interaction between format and impact level [$F(4.26, 242.82) = 11.11$, $p < 0.001$, and $\eta_P^2 = 0.16$]. The greatest differentiation in impact levels was seen in the outlook condition: Mean ratings were lowest at the low impact level ($M = 19.07$; SD = 17.69), and the highest ratings were in the extreme impact level ($M = 84.50$; SD = 28.50) (see Fig. 10). There was also a significant main effect of impact level on severity (see Fig. 10) [$F(2.13, 242.82) = 354.99$, $p < 0.001$, and $\eta_P^2 = 0.76$]. Thus, although all participants indicated higher severity with higher impact levels, the advisory/warning format was perceived as indicating greater severity overall and the outlook format gave rise to the greatest differentiation between levels.

3) DAMAGE RATING

In the ANOVA on damage ratings there was a significant main effect of format [$F(2, 114) = 4.0$, $p = 0.021$, and $\eta_P^2 = 0.07$]. As with severity, the mean damage rating was significantly higher for the advisory/warning format ($M = 50.61$; SD = 25.85) than for the text-only control ($M = 40.94$; SD = 25.04), with $t(86) = -2.93$ and $p = 0.004$, but not significantly different than the outlook format ($M = 44.20$; SD = 27.14), with $t(66) = -1.81$ and $p = 0.07$, which fell in between. Mean damage rating in the outlook condition was not significantly different from the control, $t(76) = -0.98$, with $p = 0.33$.

There was also a significant interaction [$F(4.38, 249.66) = 7.80$, $p < 0.001$, and $\eta_P^2 = 0.12$], such that the greatest differentiation was seen in the outlook condition: Mean ratings were the lowest at the low impact level ($M = 25.00$; SD = 23.26) and the highest at the extreme impact level ($M = 62.89$; SD = 19.56). There was also a significant main effect of impact level (see Fig. 11) [$F(2.19, 249.66) = 262.06$, $p < 0.001$, and $\eta_P^2 = 0.70$]. Thus, although participants indicated greater damage with higher impact level, the advisory/warning was perceived as indicating greater damage overall, as with the previous two analyses, there was greater differentiation in levels with the outlook format.

4) PROPORTION OF POPULATION AFFECTED

In the ANOVA on proportion of population affected ratings, both the main effect of format [$F(2, 114) = 0.85$, $p = 0.43$, and $\eta_P^2 = 0.01$] and the interaction between format and impact level [$F(5.22, 297.54) = 1.46$, $p = 0.20$, and $\eta_P^2 = 0.02$] failed to reach significance. There was a significant main effect of impact level (see Fig. 12) [$F(2.61, 297.54) = 116.38$, $p < 0.001$, and $\eta_P^2 = 0.51$]. Thus, in all formats, ratings increased with impact level.

5) ACTION RECOMMENDATION

Next, we examined the effect of briefing format and impact level on participants' decisions to recommend actions in response to the event. In the ANOVA on the mean action recommendations score, although the effect of format failed to reach significance [$F(2, 114) = 0.39$, $p = 0.68$, and $\eta_P^2 = 0.01$], there was a significant interaction [$F(4.2, 239.4) = 4.13$, $p = 0.002$, and $\eta_P^2 = 0.07$]. The mean action recommendations score in the outlook condition was the lowest at the low impact level ($M = 0.72$; SD = 1.18) and the highest at the extreme level ($M = 3.38$; SD = 0.95). There was also a significant main effect of impact level (see Fig. 13) [$F(2.1, 239.4) = 215.99$, $p < 0.001$, and $\eta_P^2 = 0.65$]. Thus, although in all formats, ratings increased with impact level, there was greater differentiation in levels with the outlook format.

6) UNDERSTANDING RATING

Ratings for (self reported) understanding of email briefings were analyzed using a one-way ANOVA with briefing format

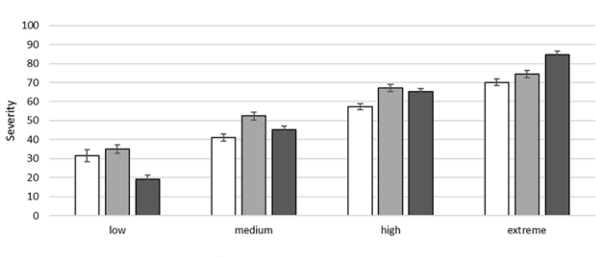


FIG. 10. Severity ratings in all formats for experiment 2.

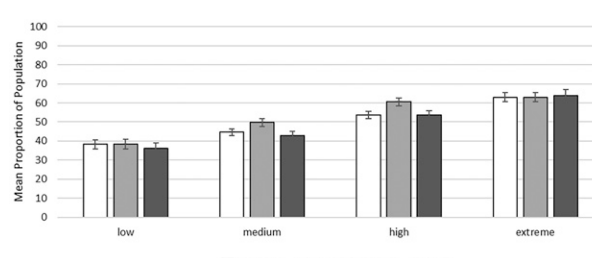


FIG. 12. Ratings of proportion of population affected in all formats for experiment 2.

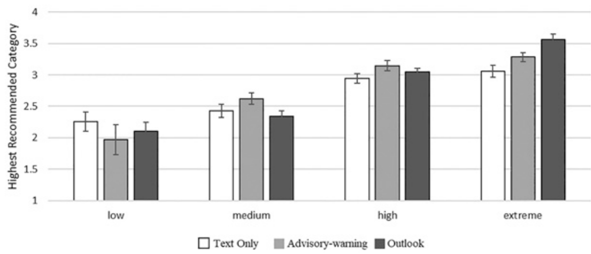


FIG. 13. Recommended action in all formats for experiment 2.

(text only, advisory/warning, and outlook) as the independent variable. As with emergency managers, although understanding ratings were generally high, there was no significant difference between formats (*advisory/warning*: $M = 80.14$; SD = 11.85 and *outlook*: $M = 81.3$; SD =15.5), with $F(2114) = 1.83$, $p > 0.05$, and $\eta_P^2 = 0.03$.

c. Discussion

As with emergency managers the text description alone allowed nonexpert participants to distinguish between impact levels in terms of likelihood, severity, damage, percent of population, as well as to recommend increasing degrees of response (action recommendations). However, there were important differences due to format. Some of these were similar to those seen among emergency managers. The color-coded impact level tended to increase the differentiation between the highest (purple) and lowest (yellow) levels for severity and action recommendations. However, for this nonexpert user group, unlike emergency managers, including the additional terms "advisory" or "warning" tended to increase the perception of severity and damage overall. Surprisingly however, this was not true of action recommendations, which, as with the outlook format, tended to be similar to the text-only control condition. Thus, although there are some clear similarities in the two studies, there are important differences as well.

4. Conclusions

The results of two studies, one with the small group of emergency managers and the other with members of the public, suggest that the verbal descriptions of weather events, taken from actual NWS email briefings, are sufficient to communicate differences in the degree of event likelihood, severity, damage, and proportion of population affected. Ratings for all of these dependent variables increased significantly with impact level based on the text description alone. In addition, more serious and costly responses were recommended for higher impact levels. This aligns with previous research suggesting that context, in this case provided by the text descriptions, is critical for such judgements (Morss et al. 2016).

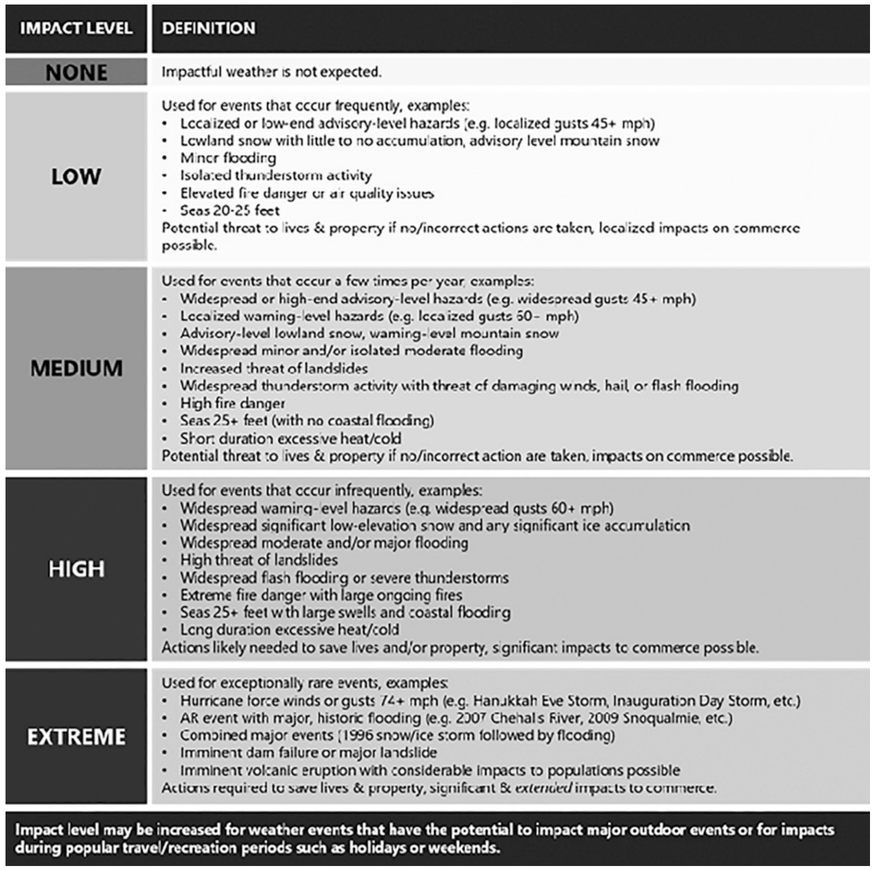
This is not to say, however, that format did not matter. In both groups the color-coded outlook format allowed for greater differentiation of likelihood [among emergency managers (EMs)], severity, damage, and action recommendations (among members of the public). In other words, for these dependent variables, the difference in ratings between the low and high levels was greater when color coding was added. For severity and damage among EMs, although trending in the same direction, the interactions failed to reach significance, likely because of the small sample size and lack of power in the experiment.[8] Taken together, this suggests that similar formats incorporating color coding, may allow users to better differentiate between high and low impact events.

In addition, the effect of the color-coded outlook format, when it had an overall effect (as with likelihood, damage, severity, and action recommendations among EMs), was to *lower* ratings overall. This is somewhat surprising because previous research had suggested the opposite effect of color coding, that it implies greater hazard (Braun et al. 1995). A possible explanation for the difference in the research reported here is the rich context provided in the task that participants completed in these experiments, including the realistic goal, background information and details of the weather events provided in "key points" and "details" in each briefing. This may have reduced the role of color coding in the decision process relative to tasks in which color coding is one of few sources of information. Availability of contextual information may have functioned to lower ratings overall when combined with the effect of greater differentiation provided by color, which lowered responses for low impact events. Thus, more research, employing realistic stimuli and contexts such as those employed in the research reported here, is required to fully understand this effect.

Important was that there were many similarities between the expert and nonexpert user groups. Both appeared to be relying mainly upon the text description to judge the risk posed by the events described in the email briefings. Moreover, for both groups, the major effect of the color-coded format was to encourage greater differentiation between high and low impact levels. This suggests that color coding might be a useful approach when communicating to both professionals as well as to members of the public when such distinctions are crucial.

The main differences between the expert and nonexpert groups were in the advisory/warning format. Among nonexperts the legacy advisory/warning format tended to increase the perceived risk of the described events, increasing severity and damage ratings relative to the text alone; however, it had almost no effect on these variables for emergency managers. It is possible that emergency managers' familiarity with the advisory/warning terms may have neutralized their impact on risk perception. Among members of the public with less direct experience, the advisory/warning terms may have elevated perceived risk or heightened awareness to a dangerous developing situation, perhaps because they are less familiar, relative to the text alone. Whether or not this is advantageous,

[8] For damage ratings, an n of 27 is required to detect an interaction effect of $\eta_P^2 = 0.12$ and power of .95. For severity ratings, an n of 21 is required to detect an interaction effect of $\eta_P^2 = 0.16$ on severity ratings.



IMPACT LEVEL	DEFINITION
NONE	Impactful weather is not expected.
LOW	Used for events that occur frequently, examples: • Localized or low-end advisory-level hazards (e.g. localized gusts 45+ mph) • Lowland snow with little to no accumulation, advisory level mountain snow • Minor flooding • Isolated thunderstorm activity • Elevated fire danger or air quality issues • Seas 20-25 feet Potential threat to lives & property if no/incorrect actions are taken, localized impacts on commerce possible.
MEDIUM	Used for events that occur a few times per year, examples: • Widespread or high-end advisory-level hazards (e.g. widespread gusts 45+ mph) • Localized warning-level hazards (e.g. localized gusts 50+ mph) • Advisory-level lowland snow, warning-level mountain snow • Widespread minor and/or isolated moderate flooding • Increased threat of landslides • Widespread thunderstorm activity with threat of damaging winds, hail, or flash flooding • High fire danger • Seas 25+ feet (with no coastal flooding) • Short duration excessive heat/cold Potential threat to lives & property if no/incorrect action are taken, impacts on commerce possible.
HIGH	Used for events that occur infrequently, examples: • Widespread warning-level hazards (e.g. widespread gusts 60+ mph) • Widespread significant low-elevation snow and any significant ice accumulation • Widespread moderate and/or major flooding • High threat of landslides • Widespread flash flooding or severe thunderstorms • Extreme fire danger with large ongoing fires • Seas 25+ feet with large swells and coastal flooding • Long duration excessive heat/cold Actions likely needed to save lives and/or property, significant impacts to commerce possible.
EXTREME	Used for exceptionally rare events, examples: • Hurricane force winds or gusts 74+ mph (e.g. Hanukkah Eve Storm, Inauguration Day Storm, etc.) • AR event with major, historic flooding (e.g. 2007 Chehalis River, 2009 Snoqualmie, etc.) • Combined major events (1996 snow/ice storm followed by flooding) • Imminent dam failure or major landslide • Imminent volcanic eruption with considerable impacts to populations possible Actions required to save lives & property, significant & *extended* impacts to commerce.

Impact level may be increased for weather events that have the potential to impact major outdoor events or for impacts during popular travel/recreation periods such as holidays or weekends.

FIG. A1. Impact-level definitions and advisory and warning criteria used by the National Weather Service (retrieved from https://www.weather.gov/sew/briefing). Note that the image that was provided to participants was colored.

may depend on the specific situation. Thus, if these terms continue to be used among members of the public, it might be advisable to include clarifying language.

In addition, the advisory/warning had no effect on resource allocation among members of the public. However, somewhat surprisingly, among emergency managers both the advisory/warning and the outlook formats tended to *decrease* response recommendations relative to text alone. In the advisory/warning condition this may have been because emergency managers interpreted the "advisory" designation, used in 56% of emails, as an indication that they could delay resource allocation until more information was available. A related explanation may account for the reduction in recommendations for the color-coded outlook format, which tended to lead to greater differentiation in risk perception. Greater differentiation among emergency managers may translate into greater effort to preserve resources for more serious emergencies. Although these are interesting and plausible interpretations of the differences between experts and novices, we are reluctant to draw firm conclusions based on the limited sample of experts tested here.

Thus, although much of the relevant risk information appears to be communicated in the text alone, there is a slightly different contribution of each of the formats tested here. The advisory/warning designation may serve as an alert to members of the public that a significant event is developing, while the same terms may be less useful to experienced emergency managers. The color-coded outlook format appears to allow both experts and members of the public to better differentiate between more and less serious events. This is an important advantage in that it may help decision-makers across levels of expertise and experience

TABLE A1. Advisory and warning definitions (NWS 2020).

	Definition	Threat	Action
Warning	Hazard is occurring, is imminent, or is very likely	Threat to life and property	Take protective action
Advisory	Hazard is occurring, imminent, or very likely	Threat of significant inconvenience	Use caution

FIG. B1. Western Washington counties, shown by the dark shades (source: Wikipedia). Note that the image that was provided to participants was colored, and the caption of their version read "highlighted in red."

to rapidly distinguish the situations that warrant more attention.

Acknowledgments. We thank the emergency managers of Washington State who contributed their time and expertise to this study. This research was supported by the National Science Foundation under Award 1559126.

Data availability statement. The data that support the findings of this study are available on request from the corresponding author.

APPENDIX A

Definitions Provided to Participants

Figure A1 shows the impact-level definitions and advisory and warning criteria used by the NWS. Table A1 lists the NWS advisory and warning definitions.

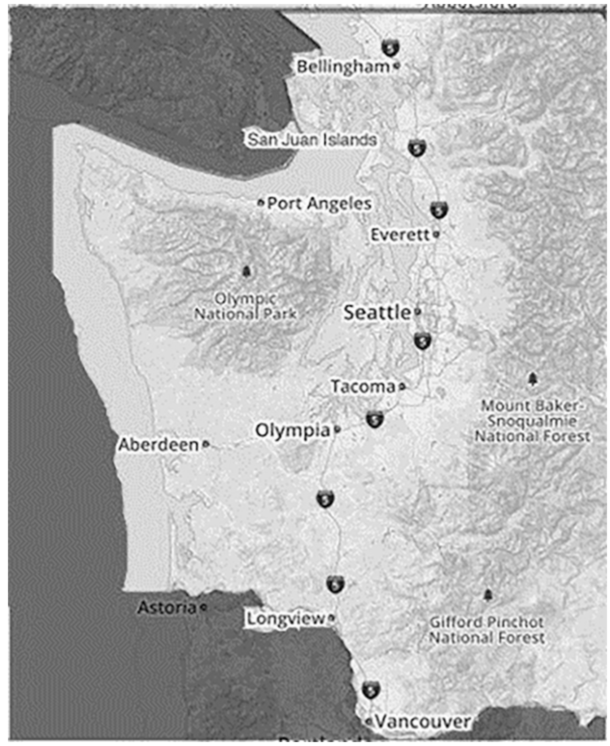


FIG. B3. Major cities in western Washington along with some of the more major roads, including Interstate Highway 5 (I-5). Note that the image provided to participants was colored.

APPENDIX B

Background Information and Instructions for Experiments

The following information was provided to participants prior to the experimental task:

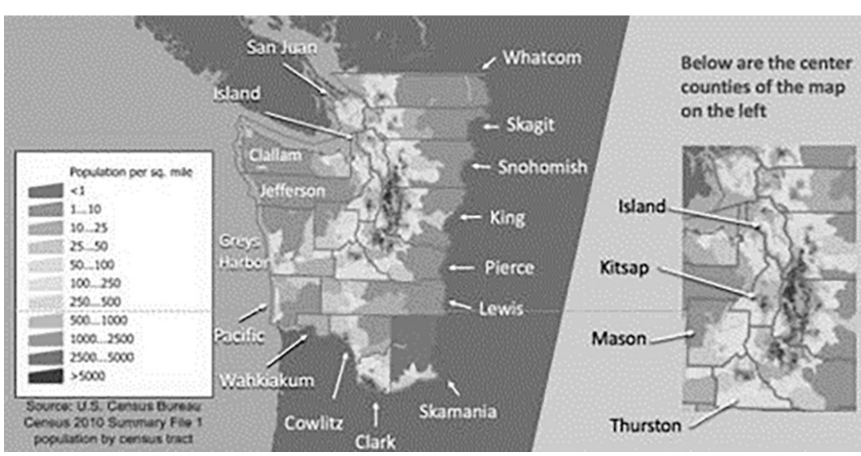


FIG. B2. Population in Washington State (source: Wikipedia). Note that the image provided to participants was colored, and the caption of their version read "Orange and Red indicate a higher population density (more people per square mile), while greens and yellows indicate lower densities."

TABLE C1. Order of stimuli. For emergency managers, both set-1 and set-2 stimuli were shown in outlook (see impact level and color columns) or advisory/warning format (see category column) or with text only. For members of the public, participants saw either set-1 or set-2 stimuli in an experimental format or text only.

	Impact level	Color	Weather	Category	Onset
			For emergency managers		
Set-1 trials					
1	Medium	Orange	Snow	Advisory	Today
2	Extreme	Purple	Snow	Warning	Tomorrow
3	Medium	Orange	Wind	Advisory	Tomorrow
4	Low	Yellow	Wind	Advisory	Tomorrow
5	High	Red	Wind	Warning	Tomorrow
6	Extreme	Purple	Wind	Warning	Tomorrow
7	Low	Yellow	Snow	Advisory	Tomorrow
8	High	Red	Snow	Advisory and warning	Tomorrow
Set-2 trials					
1	High	Red	Snow	Warning	Tomorrow
2	Medium	Orange	Wind	Advisory	Tomorrow
3	Extreme	Purple	Wind	Warning	Today
4	Extreme	Purple	Snow	Warning	Tomorrow
5	Low	Yellow	Wind	Advisory	Tomorrow
6	High	Red	Wind	Warning	Tomorrow
7	Medium	Orange	Snow	Advisory	Tomorrow
8	Low	Yellow	Snow	Advisory	Tomorrow
			For members of the public		
Set-1 trials					
1	Medium	Orange	Snow	Advisory	Today
2	Extreme	Purple	Snow	Warning	Tomorrow
3	Medium	Orange	Wind	Advisory	Tomorrow
4	Low	Yellow	Wind	Advisory	Tomorrow
5	High	Red	Wind	Warning	Tomorrow
6	Medium	Orange	Snow	Warning	Tomorrow
7	High	Red	Wind	Warning	Tomorrow
8	Extreme	Purple	Wind	Warning	Tomorrow
9	Low	Yellow	Snow	Advisory	Tomorrow
10	High	Red	Snow	Advisory and warning	Tomorrow
11	Medium	Orange	Wind	Advisory and warning	Tomorrow
12	High	Red	Snow	Advisory and warning	Tomorrow
Set-2 trials					
1	High	Red	Snow	Warning	Tomorrow
2	Medium	Orange	Wind	Advisory	Tomorrow
3	Extreme	Purple	Wind	Warning	Today
4	Extreme	Purple	Snow	Warning	Tomorrow
5	Low	Yellow	Wind	Advisory	Tomorrow
6	High	Red	Wind	Warning	Tomorrow
7	Medium	Orange	Snow	Advisory	Tomorrow
8	Low	Yellow	Snow	Advisory	Tomorrow
9	Medium	Orange	Wind	Warning	Tomorrow
10	High	Red	Snow	Warning	Tomorrow
11	Medium	Orange	Snow	Advisory	Today
12	High	Red	Wind	Warning	Tomorrow

"In this task, you will be given the role of a consultant to an emergency management organization. You will receive a series of emails with weather forecasts. You will then need to answer several questions about each of these emails, and make suggestions about the actions that should be taken."

"All of these emails will affect areas in Western Washington (see Fig. B1), which includes several large cities including Seattle, Tacoma, Everett, Vancouver, and Olympia (see Fig. B3 [below]). In Western Washington, two kinds of severe weather include high winds and snowfall. In extreme events, both can cause significant damage and potentially loss of life."

"Since your goal is to provide the best advice possible to emergency managers in the forecast areas, you may need to consider the geography and population in the region."

"Interstate 5 (I-5) is one of the major roads that runs north to south through Washington. Approximately 5.2 million residents live in Western Washington, and most live along

I-5, which runs along the Puget Sound, an inlet of the Pacific Ocean. In Fig. B2, the most densely populated areas are colored orange and red. As can be seen in Figs. B2 and B3, most of the population is between Tacoma and Everett focused largely around Seattle."

"Each of these diagrams will be available to view throughout your task."

a. Task instructions for emergency manager participants

Emergency managers were given the following additional instructions:

"In this task, you will act as a consultant. You will need to interpret weather forecasts and suggest actions."

"You will be making decisions today to advise emergency managers ahead of weather events in Western Washington. You will experience 16 weather events. Please regard these as separate events occurring a few weeks apart from one another. Your job will be to read email briefings from the National Weather Service and to answer a few questions about your personal understanding of the email. You will then be asked if you want to advise any actions to the local emergency managers. The actions below are listed in order from lowest cost, appropriate for less serious events to highest cost, appropriate for more serious events. These actions include sharing information on social media (share, retweet, post), sharing information with private organizations or news media (e.g., email, conference call), putting emergency responders on alert to take preparative actions (fire department, police department, road maintenance, utilities, etc.), and activating an emergency operations center (EOC). Because the actions are listed in order of increasing cost, you should only advise actions you believe are truly necessary to protect people, property, and businesses in Western Washington."

b. Task instructions for participants who were members of the public

Members of the public were instead given the following additional instructions:

"In this task, you will act as a consultant. As a consultant, you will need to interpret weather forecasts and suggest actions. Emergency managers are individuals who plan and coordinate community responses to threats like severe weather. Although many are found in government positions, they also exist in many large organizations like hospitals, universities, and companies. Before events they may coordinate resources and notify emergency responders (fire and medic) to be on standby in case they are needed."

"In this task, the types of actions you can suggest are limited. The actions below are listed in order from lowest cost, appropriate for less serious events to highest cost, appropriate for more serious events. These actions include sharing information on social media (share, retweet, post), sharing information with private organizations or news media (e.g., email, conference call), putting emergency responders on alert to take preparative actions (fire department, police department, road maintenance, utilities, etc.), and activating an emergency operations center (EOC). An EOC is a central location for command and coordination during emergency events. Officials and volunteers from multiple areas of government can use the central location to communicate, coordinate, and ultimately manage the response. Each of these actions have [sic] a cost, so suggesting action when it is unnecessary can be costly, but an important part of an emergency managers [sic] job is to manage situations so that they do not spiral out of control."

"You will be making decisions today to advise emergency managers ahead of weather events in Western Washington. You will experience 12 weather events. Please regard these as separate events occurring a few weeks apart. Your job will be to read emails [sic] briefings from the National Weather Service and to answer a few questions about your personal understanding of the email. You will then be asked if you want to advise any actions to the local emergency managers. The actions are listed in order of increasing cost, so you should only advise actions you believe are truly necessary to protect people, property, and businesses in Western Washington."

"Please read the email briefings carefully. There will be questions in each trial about the briefing contents."

TABLE C2. Possible combinations of conditions and stimuli sets.

Stimuli set 1, followed by stimuli set 2
Text only + advisory/warning
Text only + outlook
Stimuli set 2, followed by stimuli set 1
Text only +advisory/warning
Text only + outlook

APPENDIX C

Order of Stimuli

Table C1 lists the order of stimuli for emergency managers and members of the public, and Table C2 lists the possible combinations of conditions and stimuli sets.

REFERENCES

Bartlett, F. C., 1932: *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, 317 pp.

Braun, C. C., P. B. Mine, and N. C. Silver, 1995: The influence of color on warning label perceptions. *Int. J. Ind. Ergon.*, **15**, 179–187, https://doi.org/10.1016/0169-8141(94)00036-3.

Bryant, B., M. Holiner, R. Kroot, K. Sherman-Morris, W. Smylie, L. Stryjewski, M. Thomas, and C. Williams, 2014: Usage of color scales on radar maps. *J. Oper. Meteor.*, **2**, 169–179, https://doi.org/10.15191/nwajom.2014.0214.

Chapanis, A., 1994: Hazards associated with three signal words and four colours on warning signs. *Ergonomics*, **37**, 265–275, https://doi.org/10.1080/00140139408963644.

Corfidi, S., 2010: A brief history of the Storm Prediction Center. NOAA, https://www.spc.noaa.gov/history/early.html.

Joslyn, S., and S. Savelli, 2010: Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteor. Appl.*, **17**, 180–195, https://doi.org/10.1002/met.190.

Mayhorn, C. B., M. S. Wogalter, J. L. Bell, and E. F. Shaver, 2004: What does code red mean? *Ergon. Des.*, **12**, 12–14.

Morss, R. E., K. J. Mulder, J. K. Lazo, and J. L. Demuth, 2016: How do people perceive, understand, and anticipate responding to flash flood risks and warnings? Results from a public survey in Boulder, Colorado, USA. *J. Hydrol.*, **541**, 649–664, https://doi.org/10.1016/j.jhydrol.2015.11.047.

National Weather Service, 2014: National Weather Service Hazard Simplification Project social science research for phase I: Focus groups. NOAA Coastal Services Center Final Rep., 30 pp., https://www.weather.gov/media/hazardsimplification/Haz-Simp-Final%20-Focus-Group%20Report-Phase%20I-TO%20NOAA.pdf.

——, 2016: Impact based warning goals. NOAA, https://www.weather.gov/impacts/goals.

——, 2017: Watch/warning/advisory definitions. NOAA, https://www.weather.gov/otx/Watch_Warning_Advisory_Definitions.

——, 2018: National Weather Service Hazard Simplification: Public Survey. NOAA Final Rep., 161 pp., https://www.weather.gov/media/hazardsimplification/HazSimp%20Public%20Survey%20-%20Final%20Report%20-%2006-01-18.pdf.

——, 2020: NWS Seattle Briefings & Information. NOAA Doc., 2 pp., https://www.weather.gov/sew/briefing.

——, 2021: Planned major change to NWS' hazard messaging headlines no earlier than calendar year 2024. Public Information Statement 21-12, 3 pp., https://www.weather.gov/media/notification/pdf2/pns21-12_haz_simp_headlines.pdf.

National Weather Service Central Region, 2011: NWS Central Region service assessment: Joplin, Missouri, Tornado, 22 May 2011. NOAA Doc., 40 pp., https://repository.library.noaa.gov/view/noaa/6576.

Rashid, R., and Wogalter, M. S., 1997: Effects of warning border color, width, and design on perceived effectiveness. *Advances in Occupational Ergonomics and Safety*, B. Das and W. Karwowski, Eds., IOS Press, 455–458.

Weaver, J. F., L. C. Fast, S. Miller, and R. J. Mazur, 2017: Public response to National Weather Service severe weather watches and warnings. Part I: Severe winter storms. Colorado State University CIRA Societal Impacts Project Tech. Rep. 04, 20 pp.

Weber, E. U., and R. A. Milliman, 1997: Perceived risk attitudes: Relating risk perception to risky choice. *Manage. Sci.*, **43**, 123–144, https://doi.org/10.1287/mnsc.43.2.123.

Williams, C. A., P. W. Miller, A. W. Black, and J. A. Knox, 2017: Throwing caution to the wind: National Weather Service Wind products as perceived by a weather-salient sample. *J. Oper. Meteor.*, **05**, 103–120, https://doi.org/10.15191/nwajom.2017.0509.

Wogalter, M. S., A. B. Magurno, A. W. Carter, J. A. Swindell, W. J. Vigilante, and J. G. Daurity, 1995: Hazard associations of warning header components. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **39**, 979–983, https://doi.org/10.1177/154193129503901503.

——, V. C. Conzola, and T. Smith-Jackson, 2002: Research-based guidelines for warning design and evaluation. *Appl. Ergon.*, **33**, 219–230, https://doi.org/10.1016/S0003-6870(02)00009-1.