FINAL REPORT

An Examination of the SPC's Convective Outlook Products and Risk Category System Among Members of the Public

January 2022

Dr. Castle Williams
Dr. Andrew Grundstein
Dr. Jiyeon So

**Statistical analyses for Study 3 were performed by*
Yu Wang and Dr. Xianyan Chen at the UGA Statistical Consulting Center

NA18OAR4590366
Funding provided by NOAA's Joint Technology Transfer Initiative from the Weather Program Office (WPO) and Office of Atmospheric Research (OAR).

Table of Contents

Executive Summary

Overview:

The National Weather Service Storm Prediction Center (SPC) develops and issues several forecast products that depict both severe and non-severe thunderstorm threats across the contiguous United States (Grams, Bunting, and Weiss, 2014). Among these forecast products are the Day 1 Convective Outlook and the Public Severe Weather Outlook (PWO) which provide categorical risk information that is derived from probability forecasts for tornadoes, damaging winds, and large hail. This categorical forecast information uses numbers (e.g., 4), risk language or risk words (e.g., Moderate), and colors (e.g., red) to graphically communicate an area's overall risk for severe or convective weather (e.g., tornadoes, hail, high winds). The categorical risk levels range from non-severe thunderstorm areas (i.e., Thunder) to five risk types (Marginal, Slight, Enhanced, Moderate, and High).

This five-tier system is relatively new, with the SPC overhauling their risk category system in October 2014. Prior to this change, the SPC used three categorical risk levels (Slight, Moderate, and High) to convey severe weather risk information. Since the change, only a few studies have examined the usability of the Day 1 Convective Outlook and the newly designed risk categories. In particular, the National Weather Service conducted a study that examined the interpretation and use of the Day 1 Convective Outlook among broadcast meteorologists and emergency managers (NOAA 2016). Although the emergency managers and broadcast meteorologists felt as though the new changes were effective for their use, they expressed concern that members of the public many not be able to adequately understand or use the Day 1 Convective Outlook. With broadcast meteorologists more frequently using this product on-air and members of the Weather Enterprise regularly sharing this graphic on social media, the Day 1 Convective Outlook graphic has gained visibility among these audiences. **However, to our knowledge, no study to-date has examined the general public's knowledge, use, or understanding of the Day 1 Convective Outlook or the Public Severe Weather Outlook.**

To address the operational concerns outlined above, a mixed methods approach was used to determine how members of the public interpret, use, and understand (1) the Day 1 Convective Outlook and (2) the SPC's severe thunderstorm risk category system. This report outlines three different studies, explains their complementary research design, and presents each of their individual findings. More importantly, however, this report goes a step further to triangulate the research findings from all three studies to offer research-guided recommendations on how to improve severe weather risk communication among members of the public.

Methodology:

STUDY 1

When this research was initially conducted, only a few studies to date had examined the usability of the Day 1 Convective Outlook graphic and the SPC's risk category system. Those studies, however, only examined the use of this product among emergency managers and broadcast meteorologists (NOAA 2016). Therefore, there was a lack of research that explored the usability, interpretation, and comprehension of the Day 1 Convective Outlook graphic and the SPC's risk category system among members of the public. To fill this gap in the literature, we conducted semi-structured interviews to gauge general public knowledge, use, and understanding of the Day 1 Convective Outlook graphic and risk category system to provide SPC meteorologists with qualitative general public feedback.

STUDY 2

Study 1 offered rich qualitative feedback on how members of the public use, interpret, and understand the SPC's Day 1 Convective Outlook, and accompanying risk category system. However, the goals of qualitative research are more exploratory in nature. Therefore, follow-on studies are often important to explore the patterns that emerged among a larger sample of participants. In particular, the results of Study 1 highlight some usability challenges that may exist within the SPC's risk category system—namely that members of the public frequently interchanged both Marginal and Slight, as well as Enhanced and Moderate. The aim of Study 2, then, was to deploy an experimental survey that asked a larger sample of participants to complete an identical usability exercise to examine the patterns that emerged and compare them with the results from Study 1. This provides insight on whether the interpretation challenges that occurred in Study 1 are more prevalent and generalizable to the U.S. population.

STUDY 3

Study 1 and Study 2 similarly highlighted that members of the public experienced usability challenges when attempting to navigate the SPC's risk category system using risk words alone. A new research article, by Ernst et al. (2021), came to the same conclusion. Across all three studies, members of the public frequently interchanged the positions of the words Marginal and Slight, as well as Moderate and Enhanced when using the SPC's risk category system. With these convergent findings, the research team decided that the next area worth exploring was whether there were any changes that could be made to improve the usability of the SPC's risk category system. Therefore, drawing on the methodologies of the 2019 Severe Weather and Society Survey (Silva et al. 2020) and a study by MacLeod and Pietravalle (2017), Study 3 asked a representative sample of the public to quantitatively rank commonly used risk words on an experimental survey to determine if there are any suitable alternatives that may help alleviate some of the severe weather interpretation challenges that were found in previous studies.

Triangulated Findings & Research-Guided Recommendations:

After looking across all three studies, the following section will provide some triangulated findings and research-guided recommendations on how to improve severe weather risk communication among members of the public. The conclusions and research-guided recommendations that are provided below are listed in order based on how often they appeared across the three studies.

Conclusion #1: There is strong evidence that the words used in the SPC's risk category system are not easily understood by members of the public. The results of Study 1, Study 2, and the research by Ernst et al. (2021) all show that members of the public frequently interchanged both Marginal and Slight, as well as Enhanced and Moderate. Similarly, Study 3 provides additional insight suggesting that both Marginal and Enhanced are likely the most challenging words to interpret, and may lead to variability in severe weather risk interpretation based on their large interquartile ranges (IQR). Therefore, in addition to participants interchanging both Marginal and Enhanced, Study 3 also suggests that these risk words likely have different meanings to people when communicating severe weather risk information.

- **Recommendation: Additional social science research is needed to make the current risk category system more intuitive for *all* end users.** Although Study 3 provided some insight on possible risk category alternatives for Enhanced and Marginal, additional social science research is needed to experimentally evaluate new risk category system prototypes. Not only that, but specific metrics must be identified ahead of the prototype evaluation to determine how best to conceptualize "improvement" between the current risk category system and the prototype risk category system(s). Identifying and documenting these metrics will provide support for any operational changes or decisions that are made. These efforts should also include *all* end users. Given the ongoing diversity, equity, inclusion, and accessibility efforts, the risk category prototypes must consider language translation and ensure that any new system considerations are thoroughly tested among a diverse set of end users.

- **Recommendation: Consult emergency managers, broadcast meteorologists, and other core partners on any proposed operational changes to the risk category words.** Although most of the research to-date on the SPC's risk category system has involved emergency managers, broadcast meteorologists, and other core partners, it will be incredibly important to include them when evaluating any changes to the SPC's risk category system. Not only will it be important to similarly evaluate the experimental prototypes with these users, but it will also be necessary to qualitatively explore this topic with core partners to obtain feedback on how any changes will affect their products, processes, and policies.

- **Recommendation: Risk levels represent a promising alternative to using risk category words.** Although the current studies did not explore the use of risk levels (i.e., Level 1) as a replacement for risk category words, findings from Study 3 suggest that there are likely no risk words that can be used to uniquely communicate five levels of risk. Based on the hierarchical cluster analysis, a variety of risk words that are used in practice today only differentiate into three levels—low, medium, and high. Therefore, if the SPC aims to continue using risk category words to communicate severe weather risk, they will struggle to find five different words that convey those five levels of risk. This is why using numerical risk levels might be a promising alternative. However, more social science research is needed before this can be implemented operationally.

- **Recommendation: Reevaluate whether the 'Thunder' category should be included in the Convective Outlook graphic and, if so, consider changing its name.** In Study 1, many participants expressed concerns with the 'Thunder' category. In particular, individuals were confused by the purpose and/or meaning of this category. Therefore, it is recommended that the SPC reconsider the value of placing this category on the Convective Outlook graphic. If it provides valuable information, then a name change is recommended. Because individuals questioned whether this would be the only zone that could expect thunder, it is recommended that the category be renamed to "Thunderstorms" or "Non-Severe." This would provide additional clarity on the threats associated with this category, but also what makes it unique in comparison to the other risk categories.

Conclusion #2: Some participants had difficulties interpreting magenta in the context of the other risk category colors. Although less prominent in Study 2, both the results from Study 1 and Study 2 showcase the challenges that participants faced when trying to rank magenta against the other risk category colors. In the qualitative study of the SPC risk category system, a majority of participants (n = 19, 63%) thought that red was a more threatening color than magenta and that "pink is not a threatening color." A similar pattern emerged in Study 2 with a larger sample of participants, however this time, it was a smaller portion of participants that perceived red to be more threatening than magenta (n = 436, 40%). Even though it was not a pattern favored by a majority of participants, it is still a large enough proportion to be noteworthy. Not only that, but a closer look at the color rankings in Study 2 (Table 7, pg. 30) reveal that participants simply interchanged and moved Magenta to various spots within the five color ranking. This adds to the evidence that some participants had difficulty interpreting magenta in the context of other colors. This is concerning considering this misinterpretation is occurring at the high end of the risk category scale.

- **Recommendation: Additional social science research is needed to determine if there is a color alternative that would better communicate high end weather threats.** There is not enough social science evidence at this time to recommend that the SPC should alter their color scheme. However, it is suggested that additional social science research explore color alternatives to determine if they can better communicate high end weather threats. In particular, this social science research might also inform

NOAA/NWS' ongoing consistent depiction of risk efforts. Therefore, prior to finalizing the policies surrounding those efforts, additional social science research is warranted that provides theoretical advancements on how to effectively use color for scales, indices, categories, and/or risk risk levels.

Conclusion #3: There is emerging evidence that combining colors and risk category words does improve the intuitiveness of the SPC's risk category system. Recall that when the new SPC risk categories became operational, SPC's leadership made a conscious decision to use colors, numbers, and risk category words to provide multiple cues to help communicate severe weather risk. Therefore, the research team evaluated whether combining colors *and* risk categories would improve the intuitiveness of the SPC's risk category system among members of the public. The small qualitative sample in Study 1 revealed mixed results when presenting participants with two distinct cues, however in Study 2, it became obvious that combining both colors and risk category words does improve the intuitiveness of the SPC's risk category system. This was seen in Study 2 as a large majority of participants (67%) improved their risk category ranking when color was added as a secondary cue. Not only that, but almost 40% of people improved their risk category ranking *and* matched the official risk category word ranking used by the Storm Prediction Center when color was included as a secondary cue. That means there was a 35% increase in individuals correctly matching the SPC's official ranking when color was also used. However, this does not mean that the previously documented challenges with the SPC's risk category words simply disappear. A closer look reveals that the word misinterpretations still exist among a smaller portion of the sample.

- **Recommendation: Continue using colors, numbers, and risk category labels in all SPC risk communication with the general public.** Although the SPC currently uses multiple cues in their severe weather risk messaging and graphical products, this research showcases the value of using color in addition to risk category words when communicating severe weather risk. This also means that the use of color and numbers should continue, even when researchers explore risk category alternatives and experimental prototypes.

Conclusion #4: The qualitative study revealed that participants experienced several interpretation challenges while attempting to use and interpret the Day 1 Convective Outlook graphic. Although a majority of participants were able to navigate the Day 1 Convective Outlook and obtain information that they felt was helpful for making decisions, there were several interpretation challenges that resulted in participants incorrectly interpreting information that is not necessarily being communicated by the graphical product. For example, when asked to interpret their severe weather risk, some individuals interpreted the graphic in terms of likelihood that their location might experience severe weather whereas others described it in terms of severity. A closer look revealed that some participants were using color as a proxy for severity information. Another common misinterpretation was the assumption that a storm's movement or motion would affect the risk category of their location. Therefore, a lot of participants looked downstream from their location to determine their severe weather potential. After examining the results a bit closer, the research team questioned whether participants

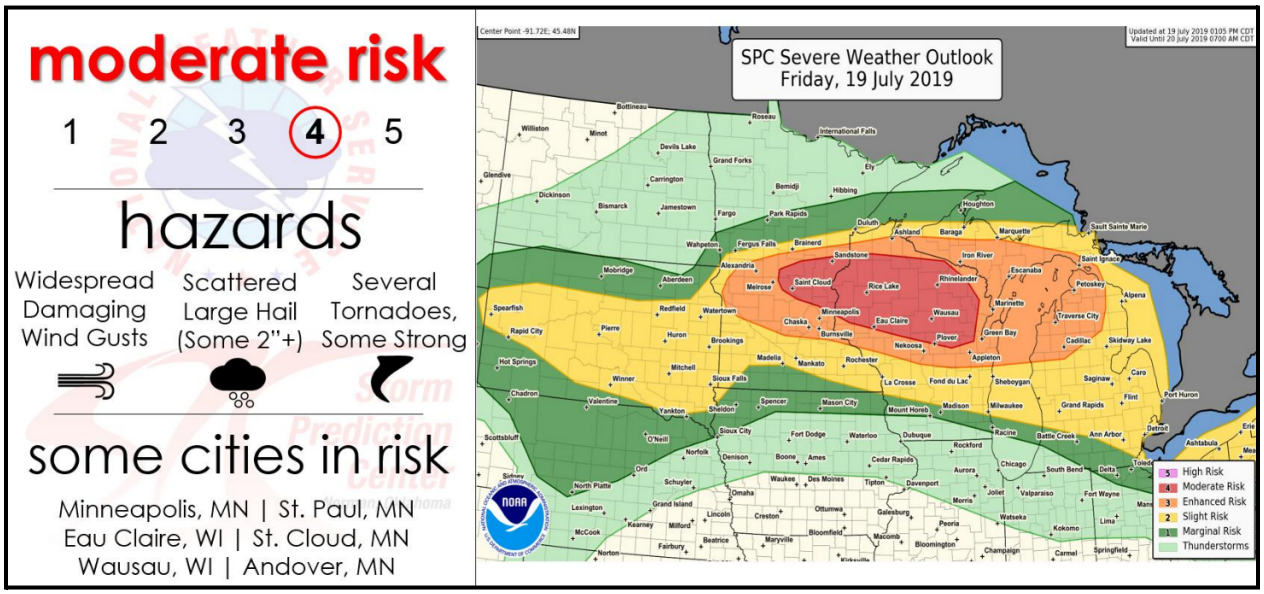
might be connecting or comparing the Day 1 Convective Outlook with the graphical output of a weather radar. This misinterpretation may also explain why some participants associated color with the intensity of the severe weather potential.

- **Recommendation: When communicating with the public and sharing information on social media, the SPC should use the Public Severe Weather Outlook graphic more frequently.** With more participants preferring the PWO in Study 1, and several commenting that it was "a lot cleaner, crisper, and easier to follow" it is recommended that the design of the National SPC Categorical Outlook graphic be used less frequently for public communication. Not only that, but participants struggled to understand and interpret the abbreviations in the legend of the National SPC Categorical Outlook graphic (e.g., SLGT = Slight). These graphics can still be used for more expert audiences; however, when publishing graphics to social media and other channels frequently used by the public, it is recommended that the SPC use the PWO design (see Figure 2). *Note:* This does not mean that the SPC should stop sharing national graphics. Instead, the SPC should promote national graphics that use the PWO's streamlined design.

- **Recommendation: Emphasize the use of "today," time updated, and valid time until on Public Severe Weather Outlook graphics.** In an attempt to overcome the perception that storm movement plays a large role in an individual's risk perception, it is recommended that the SPC consider using "today" in the title of the graphic or the date, the time the graphic was updated, and the time that the graphic is valid until on the PWO graphic. Currently, if a PWO graphic contains these pieces of information, they are often tiny and underemphasized. Perhaps increasing their visibility could better convey that this is a static graphic. However, as an enterprise, meteorologists have sought to convey that the atmosphere is always in motion. As such, many weather products can be put into motion. Based on the findings from this study, it is anticipated that the pervasiveness of the weather radar and weather products may be negatively affecting an individual's ability to consider this graphic in a static state. Beyond these recommendations, other ways to communicate that an individual's severe weather risk lasts throughout the day should be considered. For example, the SPC might consider adding timing information. This would provide additional context clues and help the user understand when they are most at risk for severe weather.

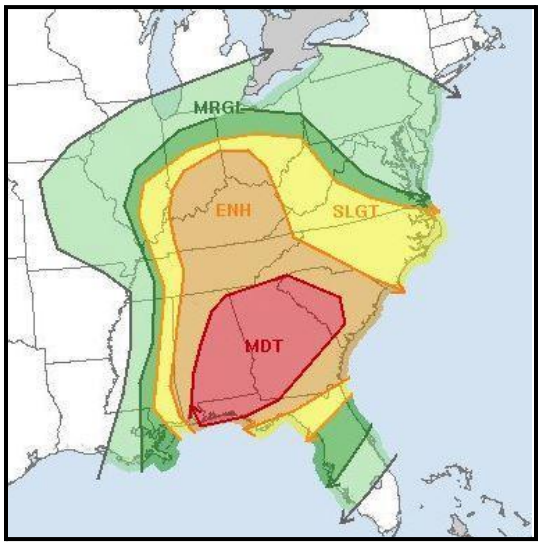
- **Recommendation: Emphasize the threat or hazards being depicted by the graphic, and/or change the title of the PWO graphic to "Severe Storm Outlook."** Many participants in Study 1 struggled to identify and/or understand the hazard that was threatening their area. Therefore, it is recommended that PWO graphics include information about the specific hazards that are threatening the risk area. For example, some recent SPC graphics have provided this information on the side of the graphic (Figure 11). This should provide more emphasis on the threats directly impacting individuals in these areas. Another possibility is to change the name or title of the PWO graphic. Throughout the interviews, individuals were unsure what encompassed "severe weather" and usually attributed various types of weather as "severe." Participants

described the Day 1 Convective Outlook as depicting excessive rain, snow, hurricanes, thunderstorms, and even extreme heat. Therefore, the SPC might consider changing the title of the graphic to “Severe Storm Outlook.” This would emphasize the threat for “storms,” and eliminate the possibility of other weather hazards. A combination of a name change and the addition of the threat information would provide more clarity to the end users.



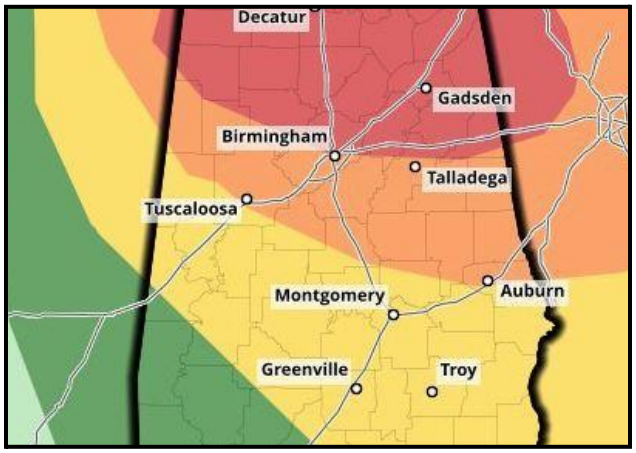
See Figure 11: New SPC Day 1 Convective Outlooks that emphasize hazard-related information.

- **Recommendation: Remove the arrows on the SPC National Categorical Outlook graphic.** Although it was recommended that the SPC National Categorical Outlook graphic be used less frequently when communicating with members of the public, it is important to note that individuals overly emphasized the arrows in the SPC National graphic and frequently used them to infer storm direction. Therefore, to reduce confusion, it is recommended that these arrows be removed from the risk boundaries (see Figure 3).



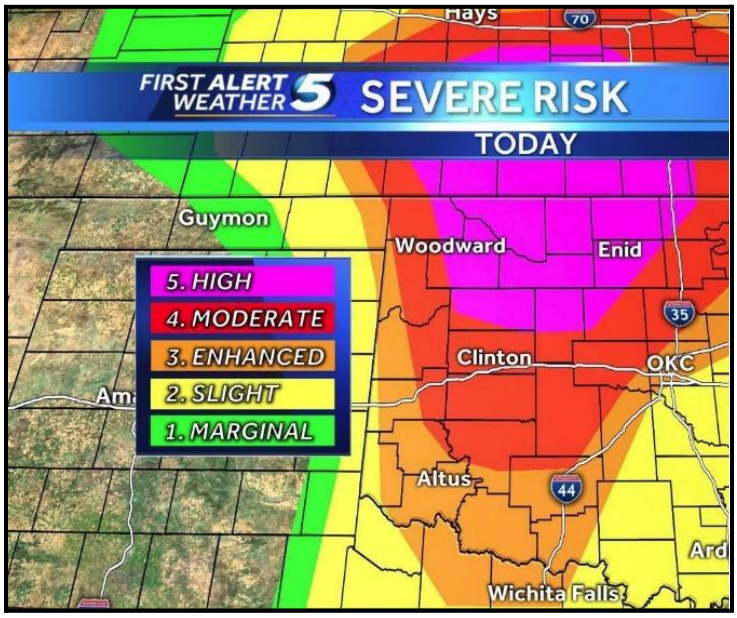
See Figure 3. Arrows on the National Convective Outlook Graphic.

- **Recommendation: The risk boundaries should be completely removed, or made less prominent on the PWO graphic.** With the risk boundaries creating many different interpretations of risk and uncertainty in Study 1, it is recommended that they be completely removed or made less prominent on the PWO graphic. By emphasizing the boundaries, the participants focused on them and attempted to understand their meaning. Therefore, making them less prominent on the PWO graphic, should lead to more individuals focusing on the categorical information and make transitioning between categories more fluid. To bolster the generalizability of these findings, further social science research is needed to explore the implications of removing the risk boundaries from the PWO graphic (see Figure 4)..



See Figure 4: Tuscaloosa is on a risk boundary between a 'Slight' and 'Enhanced' risk.

- **Recommendation: The legend on the PWO graphic should be more strategically placed and graphically modified to improve its usability.** In viewing a variety of Convective Outlook graphics, participants in Study 1 were able to examine various legends. This provided the opportunity for individuals to offer graphical preferences and recommendations for the PWO graphic. In particular, individuals thought the legend should be strategically placed so that it was closer to the risk areas. This would draw the eye more to the risk areas and the legend at the same time. Sometimes strategically placing a legend can interfere with the usability of the graphic, especially if it covers up a city, town, or region. Therefore, legends should be strategically placed when possible. Other recommendations for the legend include (1) making it vertical, (2) putting the entire risk category word in a long-colored box (see Figure 5), and (3) adding the word "Risk" to the end of



See Figure 5. Broadcaster graphic preferred by many because they were drawn to the vertical legend.

each risk category word. This offers more context clues in describing the risk categories. *Note:* Newer SPC Day 1 Convective Outlooks have started making the legend vertical and adding the word “Risk” to the end of each risk category (see Figure 9). This practice should be repeated when possible.

Conclusion #5: Social and physical scientists working on similar NWS products, services, policies, or processes should be connected when possible. In addition to the results of these three studies offering relevant conclusions and recommendations to the NWS and Storm Prediction Center, it is also important to highlight the collaborations that were fostered by NWS and OAR’s Weather Program Office that informed our research process. During 2020, the NWS, SPC, and Weather Program Office helped connect researchers that had been funded by NOAA to improve, alter, and/or change the SPC’s Convective Outlook graphic. This resulted in collaborations between researchers at different institutions, the sharing of knowledge across research projects, and also knowledge sharing efforts for the broader weather enterprise. For example, our collaboration with researchers at the University of Oklahoma resulted in an entire American Meteorological Society session on social science findings associated with the SPC’s Convective Outlook graphic. Not only that, but our fruitful collaboration also resulted in a review of all social science research that had been conducted on the Convective Outlook graphic (see Krocak et al. 2021). Therefore, by connecting with researchers that had similar goals and research questions, we were able to brainstorm together, build complementary research projects, and work together to improve NWS products, services, policies, and processes.

Introduction and Overview

The National Weather Service Storm Prediction Center (SPC) develops and issues several forecast products that depict both severe and non-severe thunderstorm threats across the contiguous United States (Grams, Bunting, and Weiss, 2014). Among these forecast products is the Day 1 Convective Outlook (Figure 1) which provides categorical risk information that is derived from probability forecasts for tornadoes, damaging winds, and large hail. This categorical forecast information uses numbers (e.g., 4), risk language or risk words (e.g., Moderate), and colors (e.g., red) to graphically communicate an area's overall risk for severe or convective weather (e.g., tornadoes, hail, high winds). The categorical risk levels range from non-severe thunderstorm areas (i.e., TSTM) to five risk types (Marginal, Slight, Enhanced, Moderate, and High; Figure 1).

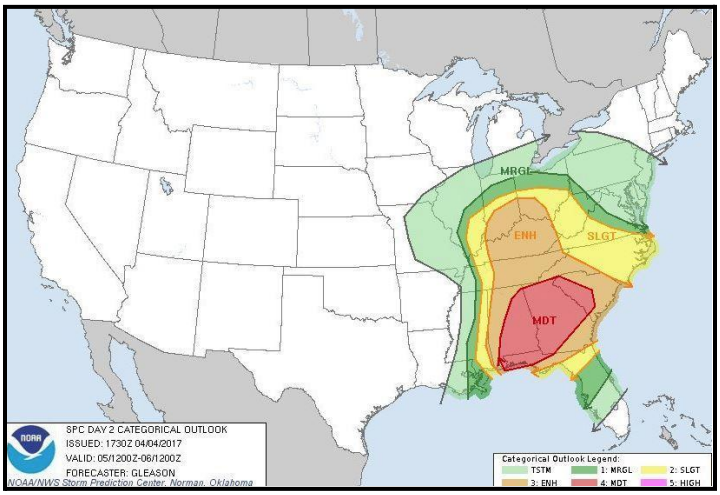


Figure 1: Day 1 Convective Outlook graphic.

This five-tier system is relatively new, with the SPC overhauling their risk category system in October 2014. Prior to this change, the SPC used three categorical risk levels (Slight, Moderate, and High) to convey severe weather threat information. This change, from a three-tiered risk category system to a five-tiered risk category system, was the result of an internal observation that SPC meteorologists were using the 'Slight' category more often than the other categories (Dr. Patrick Marsh, personal communication, November 29, 2018). Not only that, but the 'Slight' category was also exhibiting higher variability between events. To address these concerns, the SPC sought to break out the range of probabilities that were inherently lumped into the 'Slight' risk category by creating a new risk category above 'Slight' - known internally as 'Enhanced-Slight.' Therefore, by adding both Marginal and Enhanced, the SPC was able to expand the meteorological and probabilistic information that already existed in one risk category into two. At the same time, the SPC wanted to drop the 'See Text' label and replace it with a stand-alone category. Similar to the 'Slight' category, none of the underlying probabilities associated with this product changed - simply the name. However, the next challenge was to determine the best risk language or risk words to use for the newly proposed risk categories.

In search of answers, the SPC conducted some informal survey work by adding a few items to a Customer Satisfaction survey that sought feedback from individuals who frequently visited the SPC's website and used their products and services (Dr. Russ Schneider, personal communication, November 28, 2018). These survey items asked participants to rank newly proposed risk language against the legacy risk categories. Unfortunately, the results from the

survey provided little help in determining the best risk language or risk words to use for the newly proposed risk categories. In particular, these survey data revealed that:

- There were more opportunities to insert new risk words between ‘Moderate’ and ‘High.’
- Only a small gap existed to insert new risk words between ‘Slight’ and ‘Moderate.’
- There were very few risk words that conveyed a severe weather risk below ‘Slight.’

These results proved problematic, given the desire to create a new risk category below and above ‘Slight.’ After considering the survey feedback, there was not an obvious choice for a risk word that fit below ‘Slight;’ however, after careful consideration ‘Marginal’ appeared to be the best fit and was selected to replace the ‘See Text’ label. Similarly, the survey data did not reveal an obvious choice for the newly proposed ‘Enhanced-Slight’ category. Because the length of the newly proposed label (i.e., ‘Enhanced-Slight’) was rather long, it proved difficult when depicting it on a map and navigating in the SPC forecasting software. As a result, it was shortened to simply ‘Enhanced.’ Before making these operational changes, a small team of social and behavioral scientists were consulted about the results of the survey. Together, they recommended using colors and numbers, in addition to the risk words, to (1) provide other cues beyond the risk words that help communicate severe weather risk information and (2) make it easier to change the risk language when/if the time comes.

Since the change, only a few studies have examined the usability of the Day 1 Convective Outlook and the newly designed risk categories. In particular, the National Weather Service conducted a study that examined the interpretation and use of the Day 1 Convective Outlook among broadcast meteorologists and emergency managers (NOAA 2016). Most, if not all, of the emergency managers and broadcast meteorologists that were interviewed were able to successfully interpret the information provided in the Day 1 Convective Outlook and appreciated the various types of information that it provides (NOAA 2016). In terms of the newly developed risk words, most emergency managers like and understand the two recent additions; however, some emergency managers felt as though the ‘Enhanced’ category “suggests more significant risk than the term ‘Moderate” (NOAA 2016) In terms of the broadcast meteorologists, some were very concerned with the change from three risk categories to five and others felt it better communicated the range of possible severe weather risk. In particular, the terms ‘Marginal,’ ‘Slight,’ and ‘Enhanced’ produced the most concern. Many noted that both ‘Marginal’ and ‘Enhanced’ are very vague terms, and that ‘Enhanced’ can be confusing in that it may communicate to many people a higher risk than the word ‘Moderate.’ In all, the emergency managers and broadcast meteorologists felt as though the new changes were effective for their use; however, they expressed concern that members of the public may not be able to adequately understand or use the Day 1 Convective Outlook.

Although the Day 1 Convective Outlook was not originally designed for members of the public to use when making weather-related decisions, with broadcast meteorologists more frequently using this product on-air and members of the Weather Enterprise regularly sharing this graphic on social media, it has gained visibility among these audiences. As a result, it is now more likely than ever that members of the public encounter the Day 1 Convective Outlook and this severe

weather categorical information. To meet the needs of these users, the SPC uses a public-friendly version of the Day 1 Convective Outlook. This product, called the Public Severe Weather Outlook (PWO; Figure 2), provides a simplified and less technical alternative to the Day 1 Convective Outlook. **However, to our knowledge, no study to-date has examined the general public's knowledge, use, or understanding of the Day 1 Convective Outlook, the Public Severe Weather Outlook, or the SPC's risk category system.**

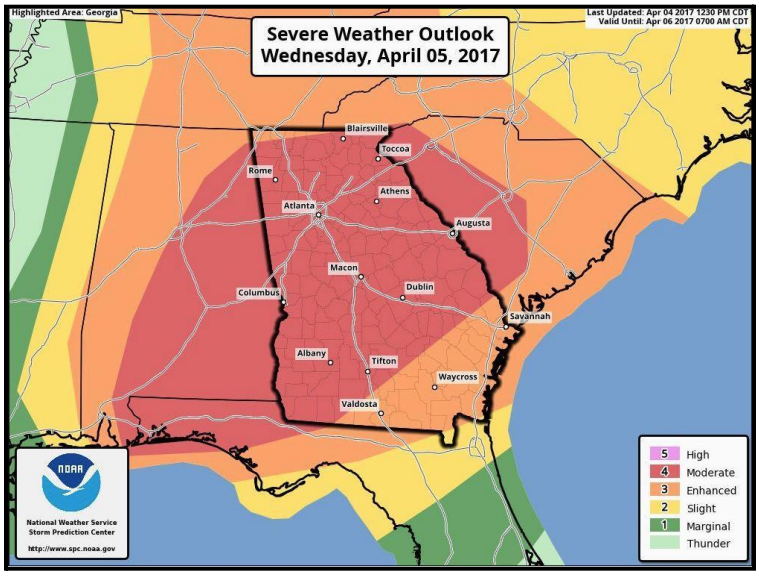


Figure 2: Public Severe Weather Outlook graphic.

To address the operational concerns outlined above, a mixed methods approach was used to determine how members of the public interpret, use, and understand (1) the Day 1 Convective Outlook and (2) the SPC's severe thunderstorm risk category system. This report outlines three different studies, explains their complementary research design, and presents each of their individual findings. More importantly, however, this report goes a step further to triangulate the research findings from all three studies to offer research-guided recommendations on how to improve severe weather risk communication among members of the public. A short description of each study is presented below.

Study 1: A qualitative study to better understand the general public's knowledge, use, and understanding of the SPC Convective Outlook graphic and the SPC's risk category system.

Study 2: A follow-up usability study that identifies how intuitive the SPC's risk category system is among a larger sample of participants—specifically students at a southeastern university.

Study 3: A study to investigate suitable risk category word alternatives that may improve the intuitiveness of the SPC's risk category system among members of the public.

STUDY 1

When this research was initially conducted, only a few studies to date had examined the usability of the Day 1 Convective Outlook graphic and the SPC's risk category system. Those studies, however, only examined the use of this product among emergency managers and broadcast meteorologists (NOAA 2016). Therefore, there was a lack of research that explored the usability, interpretation, and comprehension of the Day 1 Convective Outlook graphic and the SPC's risk category system among members of the public. To fill this gap in the literature, we conducted semi-structured interviews to gauge general public knowledge, use, and understanding of the Day 1 Convective Outlook graphic and risk category system to provide SPC meteorologists with general public feedback.

Methodology

Interviews with Weather Forecast Providers and Emergency Managers

On November 28, 2018, a member of the research team traveled to Norman, Oklahoma to meet with SPC personnel. The purpose of the meeting was to understand, in more detail, the history of the Day 1 Convective Outlook graphic, its intended use when communicating severe weather information, and to familiarize SPC personnel with the proposed research project. After speaking with SPC meteorologists and learning about their internal needs, it was collectively decided that it would be beneficial to interview weather forecast providers and emergency managers before speaking with members of the public. The goal of these interviews was to refine question wording and improve the validity of the general public interview instrument by (1) better understanding the Day 1 Convective Outlook and operational needs of these particular expert end users, (2) obtaining their perspective on the general public's knowledge, use, and understanding of the Day 1 Convective Outlook, and (3) asking these expert end users about various visual inconsistency concerns when graphically depicting the SPC's Day 1 Convective Outlook. During December 2018 and January 2019, nine telephone and in-person semi-structured interviews were conducted with three emergency managers, two National Weather Service forecasters, and four broadcast meteorologists. This methodological approach allowed for flexibility during each interview; therefore, the interviews frequently diverged from the interview instrument to explore relevant topics that arose. For more information on the semi-structured interview instruments used with weather forecast providers and emergency managers, please contact the authors.

It should be noted that an in-depth exploration of the interpretation and use of the SPC's Day 1 Convective Outlook, among weather forecast providers and emergency managers, has already been conducted (see NOAA 2016). Therefore, interview data was not collected for research purposes. However, these interviews revealed unique insights regarding visual inconsistencies

and the use of graphical design when communicating weather-related risk in the Weather Enterprise. Therefore, it is highly suggested that future research should qualitatively examine message inconsistency and, particularly, visual inconsistencies among these expert user groups.

Interviews with Members of the Public

Recruiting Participants

Individuals from Athens-Clarke County, Georgia and the surrounding counties were solicited and incentivized ($25 dollar Walmart gift card) to participate in the interviews. Due to the range of socioeconomic status in Athens-Clarke County and the surrounding counties, it was determined that this would be an optimal location to obtain a diverse sample of participants. A flyer was created and placed in public spaces (e.g., public libraries, local coffee shops, local businesses, child daycares, local organizations, local community centers, etc.) to advertise the study. Additionally, the research team drafted an email that was sent to various neighborhood and community listservs advertising the study. The research team was deliberate in the public spaces that were selected in order to encourage diversity in the participant sample. It is important to remember that a qualitative study, such as the one described in this report, does not typically target a statistically representative sample. Instead, the goal of qualitative research is to prompt an in-depth discussion and conversation with a few individuals, in hopes that their thoughts and perceptions will mirror *some* of the more meaningful concerns found in the larger population.

Interview Instrument and Procedure

A scenario-based interview instrument was developed to explore both the usability of the Day 1 Convective Outlook and to assess any message consistency concerns associated with using different visual designs. The interview instrument was piloted with six individuals with different socioeconomic and educational backgrounds. Using both this pilot study and the interviews with weather forecast providers and emergency managers, the interview instrument was further refined to improve clarity, remove irrelevant items, and add a few interview questions.

The interview began with several general questions relating to their weather information habits. This was used to initiate the conversation and make the participant feel more comfortable, but also provide context on their frequency and use of weather information (e.g., source, frequency, etc.). Participants were then asked to complete several card-sorting tasks that were used to evaluate the colors and risk language commonly associated with Convective Outlook graphics. A card-sorting task is a common method used in the field of psychology, where participants are given a set of index cards and asked to arrange them based on a given set of criteria (Psytoolkit 2019). To gain insight into people's understanding of the risk category system used by the SPC, a prompt was adapted from a study that similarly used a card-sorting task to evaluate the Homeland Security Advisory System (Mayhorn, Wogalter, and Shaver 2004). As a result, participants were asked to rank colors, risk categories, and a combined set of colors/risk

categories to further evaluate the intuitiveness of the SPC's current categorical system. After finishing the card-sorting tasks, the participants transitioned to the scenario-based portion of the interview.

To evaluate both the usability of the Day 1 Convective Outlook and the message consistency concerns associated with the graphic, four scenarios or vignettes were developed to evaluate and assess different aspects of the Day 1 Convective Outlook. After reviewing the conflicting information literature, it was determined that this would be an effective method because it has been used previously to evaluate message consistency or conflicting information concerns (Elstad, Carpenter, Devellis, and Blalock 2012). Therefore, based on conversations with SPC personnel and previous efforts in the Weather Enterprise (Klockow and Jasko 2016; NWA 2017a; NWA 2017b; AMS 2018; Williams et al. 2019; Williams and Eosco 2020), four scenarios or vignettes were designed to prompt members of the public to encounter visual similarities and differences in the Day 1 Convective Outlook graphic that meteorologists describe as either consistent and/or inconsistent[1].

- **Scenario 1:** This scenario informed people that they are at home on the weekend and come across two different graphics from the National Weather Service. The graphics used uniform colors, risk language, and spatial risk. However, one of the two graphics was the Public Severe Weather Outlook graphic and was more public-friendly. This scenario was used to elicit an initial response to their understanding and interpretation of the Day 1 Convective Outlook, was used to tease out whether members of the public preferred the PWO graphic, and was important in establishing why people thought (or did not think) the two graphics were consistent.

- **Scenario 2:** This scenario informed people that they were traveling and that they had come across two different graphics, one from the Storm Prediction Center and the other from a local Weather Forecast Office. The graphics used uniform colors and language; however, the spatial risk differed between the two graphics. This scenario was designed to evaluate the SPC's interpretation of a consistent message, whereby the exact placement of the lines or risk boundaries can be different between the SPC and a local WFO. When members of the public did not view these two graphics as consistent, this scenario also provided further insight on the implications of conflicting information between two NWS sources. Finally, this scenario placed participants at a location straddling a risk boundary. Therefore, it was designed to elicit the participants' interpretation of the graphic when on a risk boundary.
- **Scenario 3:** This scenario informed people that they were traveling and that they had come across two different graphics, one from a local broadcast meteorologist and the other from the Storm Prediction Center. The graphics used uniform colors, risk language, and spatial risk; however, the basic design of the graphics was different. This scenario was designed to understand whether these two graphics would be perceived as

[1] It should be noted that all of the graphics used in the scenarios were taken from previous severe weather events. Therefore, each of these graphics were previously created and shared by operational meteorologists.

consistent, even though they are from different sources and differ in basic graphic design. This scenario also varied in that it provided a local broadcast meteorologist's perspective first, and then moved to the forecast graphic provided by the SPC.

- **Scenario 4:** This scenario informed people that they were at home on the weekend and had come across two different graphics, one from the Storm Prediction Center and one from a local broadcast meteorologist in the Atlanta area. The graphics used uniform spatial risk; however, the language and colors differed between the two graphics. This scenario was designed to evaluate whether changing the colors and risk language in the Day 1 Convective Outlook affects the consistency between the two graphics. If members of the public did not view these two graphics as consistent, this scenario also provided further insight on the implications of conflicting information between two expert sources in the Weather Enterprise. Finally, this scenario placed participants in a location straddling a risk boundary. Therefore, it was designed to elicit the participants' interpretation of the graphic when on a risk boundary.

Within each scenario, participants were first shown a single graphic and then asked several questions about their familiarity of the graphic (Kain and Smith 2010), interpretation of the graphic (Kain and Smith 2010; Demuth, Lazo, and Morss 2012a), perceived susceptibility and perceived severity based on the graphic (So, Kuang, and Cho 2019) and their behavioral response to the graphic (Kain and Smith 2010; Demuth, Lazo, and Morss 2012a). After, participants were informed that they looked for more information and came across a second graphic. Participants were asked identical questions about this graphic, and then shown both graphics side-by-side. After giving them some time to process the two graphics side-by-side, individuals were asked: "Do you think these forecast graphics are communicating the same message?" (Backhaus 2004). After providing their answer, the researcher conducting the interview spent additional time clarifying and understanding why the participant believed or did not believe they were communicating the same message. Additional questions were asked relating to whether the participant believed there was any conflicting information between the two graphics (Elstad et al. 2012) and questions about graphical trust. This process continued through each of the four scenarios.

At the end of the interview, participants were shown all of the Day 1 Convective Outlooks that they had seen over the course of the interview and asked several questions to better identify qualities and aspects of these graphics that they preferred. Finally, participants were asked to specifically comment on the SPC's PWO graphic and to offer recommendations or ways that it could be improved to better meet their needs. After the interview, participants completed a demographic questionnaire and were thanked for taking part in the study.

The interview process began on February 18, 2019 and ended on March 27, 2019. On average, the interviews lasted between an hour and an hour and a half. After receiving informed consent, each interview was recorded. These recordings were then transcribed, analyzed, and explored via a content coding analysis (Hsieh and Shannon 2005). After the transcripts from the interviews had been examined, the themes for each question were collected and further connections were made between the responses. A final set of content codes was determined

after several iterations of collapsing the thematic categories. The responses were then reanalyzed and assigned a content code from the final set of thematic categories. For additional information on the interview instrument and/or the interview procedure, please see Appendix A.

It should be noted that the interview instrument and procedure was designed to be transferable; therefore, it is our hope that it will be adapted to assess and evaluate other public-facing graphics within NOAA and the NWS (e.g., WPC's Excessive Rainfall Outlook).

Description of Sample

A total of 30 members of the public were interviewed. Participants' ages ranged from 22 to 84, with age being represented fairly equally across all age groupings. However, the sample consisted of slightly more adults 55 and older (n = 17, 56%). The sample contained more females (n = 17, 56%), and many of the participants identified as White or Caucasian (n = 21, 70%). Participants were asked to provide their highest degree earned, with a bachelor (n = 7, 23%) or master's degree (n = 10, 33%) being the most represented categories in the sample. As a result, the sample was more educated. In terms of annual household income, the sample was almost evenly distributed between the low (less than $10,000 to $39,999), middle ($40,000 to $79,999), and high ($80,000 to $150,000+) income categories. Finally, participants were asked to provide information on their family structure. Overall, 13 of the 30 individuals reported having children. For more information on the demographic information, please see Table 1.

We also asked participants about their weather information habits. Most of the respondents reported seeking out weather information multiple times per day (n = 6, 20%) or daily (n = 17, 56%); however, the remaining participants mentioned that they search for weather information a few times a week up to once per month. Participants indicated that they gathered weather information most frequently from smartphone applications (n = 19, 63%), Internet websites (n = 10, 33%), and television (n = 9, 30%).

Table 1. Basic Demographic Information for Study 1

Variable	*n*	%		Variable	*n*	%
Age:				**Education:**		
18 to 24	1	3%		No HS Diploma	1	3%
25 to 34	5	17%		HS Graduate	3	10%
35 to 44	6	20%		Some College	2	7%
45 to 54	1	3%		College Graduate	7	23%
55 to 64	10	33%		Master's Degree	10	33%
65 to 74	6	20%		Doctorate or Professional	5	17%
75 to 84	1	3%				
				Hispanic or Latinx?		
Ethnic Identification:				Yes	2	7%
White	1	3%		No	28	93%
Black or African American	5	17%				
American Indian or Alaska Native	6	20%		**Children?**		
Asian	1	3%		Yes	13	43%
				No	15	50%

Gender:						
Male	13	43.3%		**Income:**		
Female	17	56.7%		Low (<$39,000)	10	33%
				Middle ($40,000 - $79,999)	9	30%
				High ($80,000 - $150,000)	9	30%

Results and Findings

Usability of Day 1 Convective Outlook

Current Risk Category System:

With a previous NWS study (NOAA 2016) and our own expert interviews acknowledging that the SPC's risk category terminology may not be correctly interpreted by the lay public, it was first important to evaluate the usability of the SPC's risk category system among members of the public. To do this, individuals completed three card-sorting tasks that asked them to rank the colors, risk categories, and a combined set of colors/risk categories currently used by the SPC to communicate severe weather risk information. These results provide information relating to the intuitiveness of the SPC's current system - including the colors they use, the risk category language used, and the effectiveness of combining multiple cues when communicating severe weather risk information. Although individuals were only asked to complete the card-sorting tasks, throughout this process, participants also brought up several additional concerns relating to the current risk category system used by the SPC. Therefore, their thoughts and comments are also included in this section.

Colors

The SPC currently uses five colors to communicate severe weather risk information: green, yellow, orange, red, and magenta. When asked to rank these five colors from least threatening to most threatening, most participants (*n* = 25, 83.4%) deviated from the color sequence currently used by the SPC. A closer look at those deviations, reveal three interesting patterns:

- 30% of individuals (*n* = 9) thought the color yellow was less threatening than green.
- 63% of the participants (*n* = 19) thought that red was the most threatening color.
- 66% of the participants (*n* = 20) thought that magenta was less threatening overall.

During this card-sorting task, individuals also talked about and discussed their associations with each color. Although not prompted, this provided additional insight into their interpretations of these five colors - both broadly and in meteorological terms. Overall, almost half of the individuals interviewed mentioned the use of the "traffic light colors" when laying out their index cards. These individuals usually laid green, yellow, and red down first and then tried to integrate orange and magenta into their color sequence. When describing their associations with specific colors, green was most often associated with "happiness," "being safe," "everything being fine," and "rain." When it came to the color yellow, many people associated it with sunshine. This

could be the reason why some participants thought that yellow was less threatening than the color green. Red was almost always associated with danger, and finally magenta brought up associations with winter weather, ice, and snow. Orange was not commented on as frequently as the other colors; however, a few participants had trouble identifying it and labeled it as brown. This proved difficult when trying to rank brown against the other colors.

Risk Categories

The SPC currently uses five risk categories to communicate severe weather threat information: marginal, slight, enhanced, moderate, and high. When asked to rank these risk categories from least threatening to most threatening, almost all of the participants (*n* = 29, 97%) deviated from the word pattern used by the SPC. A closer look at those deviations, reveal two interesting patterns:

- 76% of individuals (*n* = 23) thought that Slight was less threatening than Marginal.
- 83% of the participants (*n* = 25) thought that Enhanced was more threatening than Moderate.

It is important to note that three individuals that were interviewed struggled or had trouble pronouncing some of the risk categories. In particular, 'Marginal' and 'Enhanced' were the most difficult for them to pronounce. Similar to the color card-sorting task, individuals also provided associations, thoughts, and feedback for each risk category. Most of the comments refer to a concern or question about the SPC's current risk category system; therefore, they will be presented in a separate section below.

Combination of Colors and Risk Categories

As described in the introduction, when the new risk categories became operational, the SPC made a conscious decision to use colors, numbers, and risk category words to provide multiple cues to help communicate severe weather risk information. Therefore, the research team was interested in evaluating whether combining colors and risk category words would improve the intuitiveness of the SPC's current risk category system among members of the public. A third card-sorting task was conducted, and revealed that presenting two cues (i.e., colors and risk category words) offered mixed results.[2]

- 38% of participants (*n* = 11) improved their risk category ranking when two cues were presented.
- 4 individuals improved their risk category ranking and successfully matched the risk category sequence used by the SPC when two cues were presented.
- 48% of participants (*n* = 14) kept their risk category ranking the same when two cues were presented.
- The risk category ranking of 5 individuals (17%) got worse when two cues were presented.

[2] Note: These results are out of 29 individuals, because one individual successfully matched the risk category sequence used by the SPC during the risk category card-sorting task.

Concerns about the Current Risk Category System

Throughout each of the three card-sorting tasks, individuals brought up and addressed concerns that they had with the risk category sequence currently used by the SPC. After completing all three card-sorting tasks, some participants were shown the risk category sequence currently used by the SPC and asked to comment on it. Some individuals did not have a problem with the current sequence and thought that it was "just something I'll have to learn." However, a majority of the participants had concerns about the current sequence used by the SPC. The following are the most common concerns heard throughout the card-sorting tasks, and are ranked based on frequency.

Trouble relating 'Enhanced' to the other risk categories. Out of all the concerns, the most common comment about the current sequence, colors, or risk categories used by the SPC involved the 'Enhanced' risk category. Broadly, participants felt that "Enhanced [was] not a good word for weather." In particular, individuals felt as though it was a vague word that has several different connotations. When pushed to describe why 'Enhanced' was not a great adjective to describe severe weather risk, many participants had trouble comparing or relating 'Enhanced' to the other risk categories. For example, one individual describes this frustration as: "To me, Enhanced can mean anything. I'm trying to find a base [or reference] point for 'Enhanced' and it's difficult to see where that would be, because it could be 'Enhanced' past 'High' or it could be 'Enhanced' past 'Moderate,' it can be 'Enhanced' past 'Marginal' or 'Enhanced' past 'Slight.' So I just don't like that." In other words, it seems individuals were unclear where 'Enhanced' should fall in the sequence of risk categories, or as one person put it: "There's this level of obscurity behind 'Enhanced,' I could see it being behind 'High' or in front of 'High' depending on the context." This is an interesting finding considering 'Enhanced' was initially named 'Enhanced-Slight.' Therefore, the interviewees seem to be picking up on the vagueness that may have resulted from dropping 'Slight' from the original risk category name.

Mismatch between colors and words. Another common concern throughout the interviews was the feeling that the colors used with each of the risk categories did not match. For example, one individual describes this concern as: "I assign the color red with greater than 'Moderate' risk. To me, orange would be 'Moderate.' I'm having a hard time with the colors and the level of intensity they are portraying." In other words, some individuals did not feel like the colors and the risk category words used to describe the risks correspond appropriately.

Concerns that magenta is not a threatening color. During the card-sorting tasks, several individuals brought up their concerns with magenta as a color being used to convey severe weather risk information. While some people "did not really have any feelings about pink" and/or where it should go in the color sequence, others felt that "pink is not a threatening color." As a result, the placement of magenta or pink varied significantly between the participants.

Too many risk categories. Some individuals felt the SPC's current risk category system uses "...too many categories." Whenever this comment came up during the interviews, it was often followed by the need to cut, trim, or get rid of one or more risk categories. In particular, many respondents suggested cutting the risk category system down to three categories: "Get rid of

‘Enhanced’ and ‘Marginal,’ get rid of that all together, and have ‘High, Moderate, and Slight.’ Other people provided similar suggestions, but phrased it in a way that reveals their current perception of the risk categories and provides more insight behind their decision to reduce the number of categories: “You know, I’m discounting and not even giving any weight to those first two categories” and “The sort of basic improvement is to [reduce] it down from six categories to three or four. Maybe ‘Thunder’ and ‘Marginal’ are all the same risk, and ‘Slight’ and ‘Enhanced’ are all the same risk category.”

‘Slight’ and ‘Marginal’ are the same thing. A few individuals felt that there was little difference between ‘Slight’ and ‘Marginal’ and struggled when asked to place them in order from least to most threatening. In fact, one individual did not put these two in a sequence and instead stacked them on top of one another, stating that “I think they mean the same thing. Slight seems like a word that would be more understandable and less nerdy.” In short, a few interviewees struggled to determine which one communicated a more threatening risk.

Risk category order is not intuitive. The remaining concerns dealt with the SPC’s risk category order and felt that it was not as intuitive as it could be. In particular, individuals had two major comments that align with some of the deviations found in the previous sections describing the results of the card-sorting tasks: (1) ‘Marginal’ and ‘Slight’ should be reordered and (2) ‘Enhanced’ and ‘Moderate’ should be reordered. Although many of the participants felt like ‘Marginal’ and ‘Slight’ should be reordered, most could not give a reason why. When pressed, a few felt that ‘Marginal’ was simply a “stronger word.” When considering the reordering of ‘Enhanced’ and ‘Moderate,’ many felt that ‘Moderate’ was such a central word that it fit better in the middle of the risk category system: “Moderate sounds middle of the road, whereas Enhanced sounds like it’s stronger.”

Interpretation of the Day 1 Convective Outlook:

After evaluating the usability and intuitiveness of the SPC’s current risk category system, the participants were then introduced to different Day 1 Convective Outlooks across four scenarios. Therefore, the sections that follow will provide results and findings relating to their interpretation of the Day 1 Convective Outlook graphic.

Familiarity with Day 1 Convective Outlook

The first scenario included two graphics from the SPC: The SPC’s National Categorical Outlook graphic (Figure 1) and the Public Severe Weather Outlook graphic (PWO; see Figure 2). This allowed the research team to evaluate the participants’ familiarity with the Day 1 Convective Outlook and whether members of the public use this graphic when making weather-related decisions. Overall, the participants felt as though this graphic was familiar to them because it used a national map and colors to convey weather information. However, none of the participants had ever seen or used the National Categorical Outlook graphic or the Public Severe Weather Outlook graphic produced by the SPC. One participant did report seeing a similar graphic from his favorite broadcast meteorologist. As a result, most of the respondents in the sample had never seen or used the Day 1 Convective Outlook graphic.

Interpretation of Risk Categories

Throughout the scenarios, individuals were hypothetically placed under a variety of severe weather risk areas. Specifically, the hypothetical scenarios put individuals in a 'Slight, Enhanced, and Moderate' risk area. Given this breadth of experience, it was interesting to note their interpretations of the risk categories and what it meant to be in each of those risk categories. A majority of the participants expressed little concern when their city or town was located in a 'Thunder, Marginal, or Slight' risk area because, as one person put it, "it just doesn't feel like it would be anything really intense that I would have to worry about. I just know there's probably going to be rain somewhere." The 'Enhanced' risk category resulted in a few people becoming concerned; however, 'Moderate' and 'High' were by far the most concerning risk categories. In fact, when asked "which risk category would you start to worry about the severe weather threat?" Most of the participants indicated it would need to be either 'Moderate' or 'High' before they would start to be concerned or worried about the potential for severe weather. When asked to explain further, one individual in this group stated that: "[The possibility of] tornadoes doesn't sound moderate to me. That, to me, falls into the high area." Therefore, when evaluating the Day 1 Convective Outlook, a majority of the individuals in this sample began to take notice and develop concern for severe weather when their area was either in a 'Moderate' or 'High' risk area. It should be noted that several individuals were uneasy because the definitions of the risk categories were not given to them on the graphic; therefore, they felt that it forces the end user to "depend on [their] perception of what a Moderate [category] is and is not."

Because the SPC risk categories inherently incorporate both severe weather likelihood *and* intensity information (Dr. Russ Schneider, personal communication, November 28, 2018), the research team was interested in how members of the public described or interpreted these risk categories. More individuals described these risk categories as depicting the likelihood that their location would receive severe weather. Although individuals sometimes described the severity or intensity of the severe weather when interpreting the risk categories, others equated a risk category's color with the expected severity or intensity of the weather. To investigate this further, the research team purposefully asked participants to comment on both the likelihood that their location might experience severe weather (i.e., perceived susceptibility) *and* the seriousness or intensity of the severe weather that their location might experience while looking at the graphic (i.e., perceived severity). When asked about the severity of the severe weather, several individuals thought that the color was being used to convey severity. For example, one individual stated that: "The use of different colors, makes [the severe weather] sound more severe." In fact, participants also thought that the "intensity" or the use of "vibrant" colors also affected the severity message that the graphic was communicating: "The coloring implies that the weather's going to be significantly more intense in one than the other." Therefore, this may hint at the importance of color, or the perceived importance of color, when communicating graphical weather-related information.

Interpretation Challenges

Although not interpreted perfectly, a majority of the individuals navigated the Day 1 Convective Outlook graphic and obtained information that they felt was helpful to them when making

weather-related decisions. However, there were several noteworthy interpretation challenges that arose throughout the interviews. The following are the most common interpretation challenges, and are ranked based on frequency.

Risk perception was influenced by storm movement. Out of all the interpretation challenges observed, the most common misinterpretation was the assumption that storm movement would affect the risk category of their city or town. When interpreting the Day 1 Convective Outlook, a lot of the participants looked downstream from their location to determine their severe weather risk. For example, one individual asked: “I wonder if this purple area three hours from now is going to be over us? Will it go away? Will we be in the orange later tonight?” Instead of interpreting it as a static graphic that represents their risk for the entire day, many individuals were concerned that the storm’s movement may impact their severe weather potential.

The threat or hazard being depicted by the graphic is unclear. Another challenge that arose while interpreting the Day 1 Convective Outlook, was a struggle for individuals to identify and/or understand the hazard that was threatening their area. Participants described the Day 1 Convective Outlook as depicting a variety of hazards, including excessive rain, snow, hurricanes, thunderstorms, and even extreme heat. Although the scenarios and interviewer specifically mentioned that this graphic is used to depict severe weather potential, some individuals were still unclear what was meant by “severe” or “convective” weather. After providing more hazard information (i.e., that the graphic provides information on the potential for hail, high winds, and tornadoes), participants still wanted more information about the specific hazards that their location would likely experience. For example, one individual stated that: “I have no idea [what this graphic is showing me.] Am I looking at strong winds? Am I looking at the likelihood of hail? I’m unclear as to what kind of weather they’re trying to present me with.”

The presence or absence of a ‘High’ risk category lead to different perceptions of risk. When a ‘High’ category was present on a Day 1 Convective Outlook, most participants were drawn to that area immediately. Even though their hypothetical location was not in a ‘High’ risk area, these individuals felt that a ‘High’ risk category was an indicator that the weather was going to be more intense. Put another away, when a Day 1 Convective Outlook did not have a ‘High’ risk zone, people took notice: “When the map doesn’t have a ‘High’ zone on them, I think that communicates that it’s overall a less threatening severe weather [event].” Others consider the opposite to be true. They thought that the presence of a ‘High’ risk category devalued the other risk categories: “If there was no purple on here at all, and this giant area was a moderate risk, it would feel different. But, because there’s this [‘High’ risk] bullseye over here, it makes the orange, the enhanced risk zone, seem so much lesser because of that.” Finally, some individuals thought that the absence of a ‘High’ risk category upgraded the meaning of the other risk categories. For example, one individual noted that: “Not seeing any purple or pink on the map makes me feel like I’m in the [highest risk zone]. Because in my mind, red becomes the highest one.” Although the ‘High’ risk category was interpreted in a variety of ways, a majority of the respondents considered the first perspective to be true - when a ‘High’ risk category is not on a Day 1 Convective Outlook they were overall less concerned with their location’s severe weather risk.

Graphical elements were used to infer storm movement. Like the emphasis on storm direction and movement described above, storm movement information was so critical to their interpretation of the Day 1 Convective Outlook that they used other details and graphical elements (albeit incorrectly) to obtain it. For example, many participants used the arrows that outline the different risk boundaries in the SPC's National Convective Outlook graphic to infer storm direction (Figure 3): "I can't tell whether or not the storm is moving toward me or not, because the arrows indicate that it's going around in a circle. So maybe that's the wind direction?" Although these arrows are only used to demarcate different risk boundaries, the participants frequently used them to depict storm motion. However, if individuals did not use the arrows to infer movement information, then they simply "did not understand the arrows." One individual used the town names, or lack thereof, on a graphic to interpret storm motion: "I mean clearly the storm is moving south because all of the southeastern locations are listed. The storm is moving south to southeast because of those locations being named." Therefore, the need to know the storm's direction is once again a prominent theme among the participants when interpreting the Day 1 Convective Outlook.

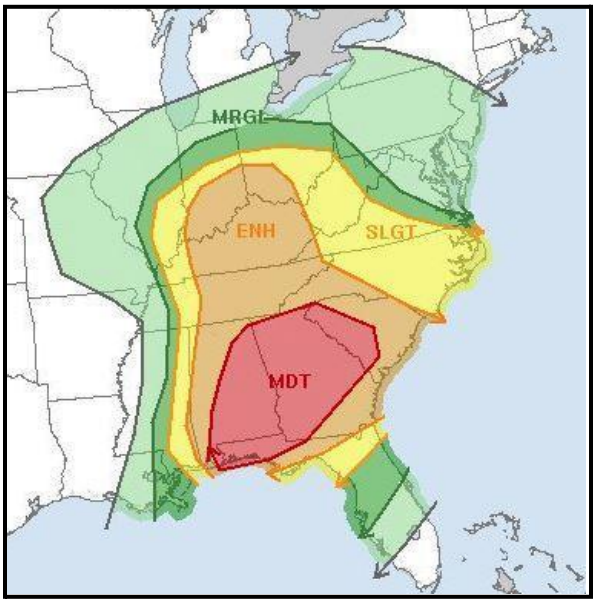


Figure 3: Arrows on the National Convective Outlook graphic.

The 'Thunder' category was hard to understand and interpret. Although not as common as the other misinterpretations, several individuals questioned the meaning and purpose of the 'Thunder' category. One participant describes it best: "I don't really like the 'Thunder' zone. Like, would there not be thunder in the rest of [the categories]? Or is this the only place to expect thunder? Can I expect rain in the 'Thunder' zone? Is that where I'm hearing thunder? Shouldn't there be lightning too?" This individual addresses several of the concerns brought up throughout the interviews: (1) Many participants questioned whether this would be the only zone that should expect thunder, (2) several individuals thought that lightning would be the more appropriate hazard, and (3) a few respondents wanted to know how 'Thunder' was related to the other risk categories. In short, the participants did not understand that the 'Thunder' zone was meant to convey areas that would likely experience non-severe thunderstorms.

Some of the abbreviations in the SPC's National Convective Outlook were unclear. While looking at the SPC's National Convective Outlook graphic, several individuals had difficulty interpreting the abbreviations or acronyms used in the legend. For example, one individual noted that "without doing the [card-sorting tasks]... I mean I know what HIGH means. And I

guess I could guess Slight [SLGT], but I wouldn't know that MGRL means Marginal or that MDT means Moderate, or I wouldn't know that ENH is Enhanced." Although it appears the card-sorting tasks were beneficial in helping participants become more familiar with the risk categories, some still noted that they would have difficulties navigating the legend because of the abbreviations. In particular, several individuals noted that they were not sure what 'TSTM' represented because it was not discussed during the card-sorting tasks.

Interpretation on a Risk Boundary or Border

After speaking with SPC personnel and listening to their concerns that members of the public may find it difficult to interpret their severe weather risk when their location is on a risk boundary or border, we thought it was an important element to add to the research project. Therefore, two scenarios were used that placed individuals in hypothetical locations that were located along a risk boundary. For example, Figure 5 shows Tuscaloosa on a risk boundary between the 'Slight' and 'Enhanced' risk categories. After interviewing members of the public, it was clear that being on a risk boundary created many different interpretations of risk and uncertainty. The following are the most common interpretations, and are ranked based on frequency.

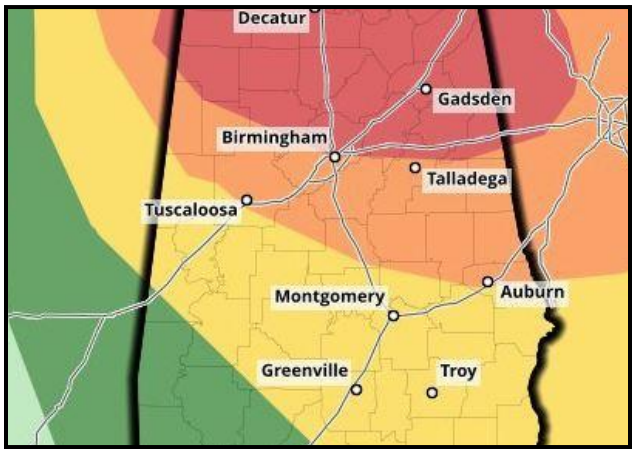


Figure 4: Tuscaloosa is on a risk boundary between a 'Slight' and 'Enhanced' risk.

An emphasis on storm movement. By far the most common interpretation of the risk boundary was feeling as though their risk was heavily influenced by the movement of the storm or the potential change in the storm's location. One participant, for example, described their interpretation of the risk boundary as: "Because I'm like right in between where it changes. So I would be like, where is the storm going? Is it coming from the Gulf or getting pushed down toward me? Where is it going? I would have to figure out where the moderate part is going." Again, the misinterpretation of the storm's movement having an impact on their severe weather risk potential was a prominent theme.

Feeling more uncertain about the weather situation. When their location was on a risk boundary, several individuals felt more uncertain about the severe weather situation. These individuals felt that being on a risk boundary affected their chances of seeing severe weather. In fact, being on a risk boundary or border usually decreased the perceived likelihood that their location would experience severe weather. Some example responses that fit into this category include: "It may be a 50-50 chance of it happening, because it's on the border between being severe and not" and "Well, I would say looking at this one that your chances of having severe weather are probably low because you're right on the edge."

Risk boundaries are part of the risk continuum. Some individual's risk perception was not affected by the risk boundaries. Participants that fell into this category did not think the risk boundaries had an important meaning, and appropriately considered their risk along a continuum. For example, one individual stated that "being on the boundary [between 'Slight' and 'Enhanced'] makes me feel like I'm at the low end of the 'Enhanced' and just barely outside of the 'Slight' risk area." Therefore, these participants interpreted the risk boundaries as the SPC currently uses them.

Fuzziness of the risk boundaries. Although similar to the category described above, a few individuals commented on the fuzziness of the risk boundaries. In other words, they specifically mentioned that the risk boundaries are only used to break apart complicated data into much simpler categories. One individual, for example, stated that: "I'm kind of struck by the way they use heavy lines, for the boundary lines, like they're emphasizing the boundaries. And to me, in reality, those boundaries are quite fuzzy. And so, it's a little bit curious that they would make them distinct, whereas in reality, they're not." This was an interesting finding, considering that this was also a common theme that arose during our discussion with SPC personnel before starting this project.

Prepare for the worst, or hope for the best. When on a risk boundary, some individuals thought that (1) they needed to prepare for the higher risk category or (2) they would be in the lower category. When on a risk boundary, those that prepare for the worst "err on the side that [they're] going to get in the more severe weather and should be prepared for that." On the other hand, those that hope for the best feel optimistic and focus on the lesser category when placed on a risk boundary. One participant describes this as: "The boundaries certainly make me feel like I'll be more hopeful and that I'm not in the Enhanced area. I feel more hopeful that maybe I might be just in the right spot to not catch the worst part of this severe weather." Therefore, when placed on a risk boundary, these individuals chose to either prepare for the worst or hope for the best.

Other interpretations of the risk boundaries included:

- Being on a risk boundary resulted in an individual examining their area more closely to determine their location's risk category. (e.g., "It depends on what side of Tuscaloosa we are on."
- Being on a risk boundary resulted in an individual blending the risk category information (e.g. "I think I'll be experiencing a 3.5 because I'm on the border").
- Being on a risk boundary resulted in an individual believing they will experience different weather types depending on their location (e.g., "It might be between raininess and cloudiness").
- Being on a risk boundary resulted in an individual believing the boundary provides some sort of informational value (e.g., "To me, it seems like the border might mean something important.").

Behavioral Response to the Day 1 Convective Outlook

The SPC's Day 1 Convective Outlook graphic is unique, in comparison to the other products produced by the SPC, because it does not prompt an immediate behavioral response. Instead, it was designed to give people a heads up about their location's severe weather potential and to encourage people to continuously check on the weather throughout the day. Therefore, in response to the Day 1 Convective Outlook, SPC personnel suggested that individuals should (1) continue monitoring the severe weather threat throughout the day and (2) take preparatory actions (e.g., have a severe weather plan, where should you go in the event of a tornado?) in advance of the severe weather threat (Dr. Russell Schneider, personal communication, November 28, 2018). Beyond these two actions, it was not recommended by SPC personnel that people make any other decisions at this point in the forecast chain (Drs. Patrick Marsh and Russ Schneider, personal communication, November 28, 2018). Given these circumstances, the research team was interested in how members of the public would respond to the Day 1 Convective Outlook.

Similar to the recommendations by SPC personnel, the two most popular behavioral responses to the Day 1 Convective Outlook were (1) continued monitoring of the weather throughout the day and (2) preparatory actions in advance of the severe weather threat. In particular, participants were very interested in searching for "more detailed information" about the severe weather threat. They described several types of information they would want after seeing the Day 1 Convective Outlook, including: the direction the storm or weather was moving, timing information, information about what the risk categories mean, specific threat information, probabilities or percentages associated with the severe weather threat, and other Convective Outlook graphics that were more zoomed into their local area.

Beyond the two actions recommended by the SPC, participants described a few other actions that they would take in response to the Day 1 Convective Outlook. Several individuals said that they would plan their day accordingly or alter preexisting plans for the afternoon after seeing the Convective Outlook graphics. Some interviewees stated that they would "stay inside for the rest of the day" or "plan to stay around the house" after seeing the Convective Outlook graphic. A smaller group of individuals noted that they would inform others of the severe weather risk. A few participants admitted that they would take action to "secure outdoor objects," "go cover the car," "bring plants inside," and other preparatory actions before the severe weather arrived. Finally, five individuals stated that they would leave their location in search of a "safe place" after seeing the Day 1 Convective Outlook graphic.

Although the SPC recommended that each risk category should, more or less, prompt the same response, a majority of the interviewees elevated their behavioral response when located in a higher risk category. For example, Scenario 1 and Scenario 4 can be used to compare how people react in their hometown when under a 'Slight' risk versus a 'Moderate' risk (Table 2). In a 'Slight' risk, participants either did nothing, continued to monitor the weather, or used the information to plan their day. On the other hand, when their location was in a 'Moderate' risk zone, more participants were interested in preparing for the weather, informing others of the risk, and changing their plans for the day. Further, a few participants mentioned they would stay

inside and take action to protect their property in a ‘Moderate’ risk zone. It makes sense that participants behave differently across the various risk categories, because there are different tiers of preparatory behaviors. In fact, this provides additional evidence that the risk categories successfully prompt different levels of perceived risk and behavior.

Table 2. Comparing behavioral response between risk categories.

Behavior	Slight Risk *(n)*	Moderate Risk *(n)*
Do nothing	6	0
Monitor the weather	17	5
Prepare	2	6
Change Plans	8	9
Inform Others	0	8
Stay Inside	0	4
Take Action to Protect Property	0	2

Note: *The PWO graphics from Scenario 1 and Scenario 4 were used for this comparison, because they similarly placed participants in their home town and used PWO graphics.*

Graphical Preferences and Suggested Improvements to PWO:

At the end of the interview, participants were shown each of the Day 1 Convective Outlooks that they had seen over the course of the interview and asked several questions to better identify qualities and aspects of these graphics that they preferred. In particular, participants were asked to specifically comment on the SPC’s PWO graphic and to offer recommendations or ways that it could be improved to better meet their needs. The following sections outline the graphical elements that participants preferred when looking across all of the Day 1 Convective Outlook graphics and provide suggestions to improve the Public Severe Weather Outlook (PWO) product.

Graphical Preferences

Over the course of the interview, participants were exposed to eight different graphics from a variety of sources. Therefore, it seemed appropriate to ask members of the public which graphical elements they preferred and which ones that they disliked. This knowledge should help the National Weather Service and the Weather Enterprise develop weather-related graphics that better suit the needs of the general public.

After seeing Scenario 1, the participants were exposed to both the SPC’s National Convective Outlook graphic and the PWO graphic. Therefore, participants were first asked which graphic they preferred. A majority of the participants preferred the PWO because they favored a more zoomed-in map, appreciated that cities/towns were listed, and liked it because it’s “a lot clearer, crisper, and easier to follow.” However, there were a few individuals that preferred the National Convective Outlook graphic. They felt like having a more zoomed-out picture gave them a better perspective on “how far the whole threat goes." Finally, there were several individuals that liked both graphics because they each serve different purposes.

At the end of the interview, participants were shown each of the Day 1 Convective Outlooks that they had seen over the course of the interview and asked to elaborate on their favorite and least favorite graphics. The following graphical elements were highlighted and preferred the most by the study participants:

- Graphics that clearly mark cities and towns, and have them in the correct place.
- Graphics that provide timing information.
- Graphics that are zoomed in and/or provide a more local perspective.
- Graphics that use “bolder,” “brighter,” or “vivid” colors when conveying the risk category areas.
- Graphics that have a legend that is strategically placed next to the threat area.
- Graphics that clearly mark prominent highways and roads.
- Graphics that provide information on storm movement or direction.
- Graphics that *do not* have prominent risk boundaries.
- Graphics that convey specific threat or hazard information.

As an example, many participants preferred the broadcast meteorologist graphic from Scenario 3 (Figure 5). This graphic embodies almost all of the graphical elements outlined above. In particular, participants were drawn to the legend in this graphic. Individuals pointed out that they (1) liked that the legend was strategically placed near the threat area and made it feel like “the legend was immediately important,” (2) appreciated that the “entire [risk category] word [was] in a long colored box,” and (3) preferred a vertical legend because it “draws your eye” and lets you “[look at] the legend all at once.” Beyond these legend preferences, participants also frequently mentioned that they preferred this graphic because it used vivid colors and did not have prominent risk boundaries. In addition to this graphic, the local Weather Forecast Office (WFO) graphic in Scenario 2 was also a favorite among many participants. These individuals noted that this graphic was their favorite because it included timing information and specific threat or hazard-related information. Therefore, graphics that include these elements may better suit the needs of the lay public.

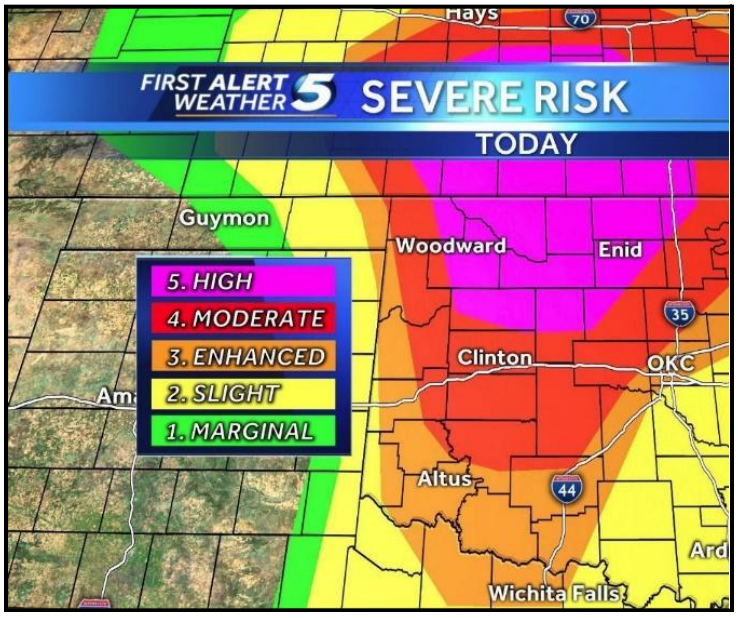


Figure 5: Broadcaster graphic preferred by many because they were drawn to the legend and vivid colors.

Suggested Improvements to Public Severe Weather Outlook (PWO)

After examining each of the Day 1 Convective Outlooks, participants were shown the three Public Severe Weather Outlooks (PWO) that they saw throughout the interview. Participants

were asked to examine these graphics and provide any suggested recommendations or improvements that would better meet their needs. The following suggestions were mentioned the most and are ranked based on frequency:

Make changes to the risk category system. This suggestion represents a variety of recommendations that involve the risk category system currently used by the SPC. However, the three most common suggestions were: (1) Reduce the number of categories used, (2) Remove the 'Thunder' category, and (3) change the colors associated with the risk categories (e.g., "Since you don't see [pink/purple] as much, will people recognize it as the highest level?").

Provide arrows or information about storm movement. Again, participants were very interested in the direction that the storm was moving. Therefore, these individuals recommended that the PWO graphic include arrows or other information about the storm's movement. Additional social science research is needed to better understand how to depict storm movement effectively on the PWO graphic.

Provide timing information. Participants appreciated the timing information that was provided in the Birmingham WFO graphic. Therefore, the individuals recommended that the PWO graphic include additional information about the timing or onset of severe weather conditions at their location.

Make changes to the risk boundaries or borders. Several individuals provided recommendations or suggestions to improve the risk boundaries or borders used in the PWO graphic. The most common suggestions included: (1) Completely remove the risk boundaries, (2) make the risk boundaries thinner or sharper, and (3) make the color of the 'Slight' and 'Enhanced' boundaries more distinct.

Make changes to the legend. As discussed in the previous section, many participants were drawn to the broadcast meteorologist graphic in Scenario 3 (see Figure 5) and preferred a legend that was strategically placed near the threat area. Other legend suggestions included: (1) making it vertical instead of horizontal, (2) reducing the spacing between the categories in the legend, (3) moving it to the top of the graphic, and (4) reordering the risk categories from least to greatest (e.g., Thunder to High vs. High to Thunder).

Other suggestions included:

- Provide additional information that describes each risk category.
- Use vivid colors instead of the pastel colors to convey the severe weather threat areas.
- Put prominent roads/highways on the graphic.
- Make the state lines darker and more prominent.
- Change the title of the graphic.

Summary

Although several interpretation challenges emerged, a majority of the participants in the study were able to navigate the Day 1 Convective Outlook and obtain information that they felt was helpful to them when making weather-related decisions. Three card-sorting tasks were used to evaluate the usability and intuitiveness of the SPC's current risk category system. Most of the participants deviated from both the color (*n* = 25, 83%) and risk category (*n* = 29, 97%) sequences currently used by the SPC. A third card-sorting task was used to evaluate whether it was beneficial to present members of the public with two cues (i.e., colors and risk category words) instead of one; however, it offered mixed results. Throughout the card-sorting tasks, participants brought up concerns about the current risk category system used by the SPC. The most frequent concerns included: (1) they had trouble relating 'Enhanced' to the other risk categories and (2) they did not feel like the colors and words used to describe the risk categories matched.

After evaluating the usability and intuitiveness of the SPC's risk category system, the participants were presented with different Day 1 Convective Outlook graphics. Most of the participants in the sample acknowledged that they had never seen or used the Day 1 Convective Outlook graphic. When asked to interpret their severe weather risk, some individuals interpreted the graphic in terms of the likelihood that their location might experience severe weather whereas others described it in terms of severity. After further investigation, it was discovered that some of the participants were using color as a proxy for severity information. Therefore, this may hint at the importance of color, or the perceived importance of color, when communicating graphical weather-related severity information. Finally, to gauge the risk perception associated with each risk category, participants were asked to comment on the risk category where they would start to take notice of the severe weather threat. Most individuals noted that they would not take notice or worry until their area was in a 'Moderate' or 'High' risk area. In terms of behavior, the two most popular responses to the Day 1 Convective Outlook were (1) continued monitoring of the weather throughout the day and (2) preparatory actions in advance of the severe weather threat. It is also important to note that participants in the sample performed different behaviors across various risk categories.

While examining the Day 1 Convective Outlook graphic, there were several noteworthy interpretation challenges that emerged. The most common misinterpretation was the assumption that a storm's movement or motion would affect the risk category of their city or town. In particular, a lot of participants looked downstream from their location to determine their severe weather potential. In looking across all of the interview data, the influence of storm movement or motion also shows up when discussing the participants' interpretation of a risk boundary and is identified as a preferred graphical element. Therefore, it is possible that participants are connecting or comparing the Day 1 Convective Outlook with the graphical output of a weather radar. This connection or comparison to the weather radar may also help explain why some participants associated color with the intensity of the severe weather.

The remaining results highlighted elements of the Day 1 Convective Outlook graphics that people preferred and provided suggestions for improving the Public Severe Weather Outlook (PWO) graphic to better meet the needs of the general public. Some of the most popular graphical elements among participants included graphics that: clearly mark cities and towns, provide timing information, provide a local/zoomed-in perspective, and use more vivid colors. Some of these preferred graphical elements emerge as suggestions to improve the PWO graphic. In particular, there were five recommendations to improve the PWO graphic that were frequently mentioned in the interviews:

- Make changes to the risk category system.
- Provide arrows or information about storm movement.
- Provide timing information.
- Make changes to the risk boundaries or borders.
- Make changes to the legend.

STUDY 2

Study 1 offered rich qualitative feedback on how members of the public use, interpret, and understand the SPC's Day 1 Convective Outlook, and accompanying risk category system. However, the goals of qualitative research are more exploratory in nature. Therefore, follow-on studies are often important to explore the patterns that emerged among a larger sample of participants. In particular, the results of Study 1 highlight some usability challenges that may exist within the SPC's risk category system—namely that members of the public frequently interchanged both Marginal and Slight, as well as Enhanced and Moderate. The aim of this study, then, was to conduct an identical usability exercise with a larger sample of participants to examine the patterns that emerged and compare them with the results from Study 1. This provides insight on whether the interpretation challenges that occurred in Study 1 are more prevalent and generalizable to the U.S. population.

Methodology

Survey with College Undergraduate Students

This study was reviewed and approved prior to its beginning in October 2019 by the University of Georgia Institutional Review Board (IRB; PROJECT00001227).

Recruiting Participants

Undergraduate students from the University of Georgia were solicited to complete a survey on severe weather graphics and messaging. In particular, undergraduate students were recruited through introductory courses in the Geography department. Introductory courses were selected to capture a more diverse set of students, in hopes that it would more closely represent a general public audience. This recruitment process began with the graduate researcher on the project (Williams) reaching out to all of the instructors that taught introductory Geography courses in Fall 2019. After speaking with each of the faculty members, five instructors (for a total of eight introductory Geography classes) agreed to participate and share the survey with their students. Therefore, those instructors were given an email template with instructions on how to complete the survey and asked to share it with their students. Some faculty members offered their students extra credit as an incentive for completing the survey. However, if a student did not wish to participate in the research study, they were still able to receive extra credit by completing an alternative assignment that was equivalent in effort or duration.

Survey Instrument and Procedure

Although the original intent of the survey was to evaluate the perceived consistency between two severe weather graphics (these results will be discussed in the second report), a secondary goal of the survey was to conduct a larger card sorting task to further explore the SPC's risk

category system among a larger pool of participants. Before being asked to evaluate the consistency of two Convective Outlook graphics, participants were asked to complete several card-sorting tasks that were used to evaluate the colors and risk language associated with the SPC's risk category system. Recall, a card-sorting task is a common method used in the field of psychology, where participants are given a set of index cards and asked to arrange them based on a given set of criteria (Psytoolkit 2019). These survey items were identical to those used in Study 1 (Mayhorn, Wogalter, and Shaver 2004), except participants were asked to sort virtual cards within the online survey. As a result, participants were asked to rank colors, risk categories, and finally, a combined set of colors/risk categories to further evaluate the intuitiveness of the SPC's risk category system.

Before being asked to complete the card-sorting tasks, participants were randomly assigned to one of the SPC risk categories (i.e., Marginal, Slight, Enhanced, Moderate, High). Here, participants were (1) asked to provide a sentence or two that describes what the risk category means to them and (2) asked to indicate the likelihood that they might experience severe weather if they were within the risk category that was assigned to them. This provided an opportunity for participants to provide unbiased, qualitative and quantitative perceptions of the risk categories *before* seeing all of the risk categories in the card-sorting task. These questions were adapted from the 2019 Severe Weather and Society Survey (Silva et al. 2020). Finally, to avoid ordering effects on the card-sorting task that combines both colors *and* words, participants were randomly assigned to complete either the color card-sorting task or the risk category card-sorting task first. For additional information on the survey items used in this study, please refer to Table 3.

The survey was launched on 16 October 2019 by sending an email to all of the instructors that had agreed to share the survey with their introductory Geography classes.The message provided a brief overview of the purpose of the survey and contained a link to the survey, which was administered through the Qualtrics online survey platform. To increase response rates, the instructors were asked to send a reminder email to their students one week after the initial launch of the survey. We closed the survey on 1 November 2019. At the close of the survey we had received 1091 usable (complete) responses. This yielded a survey return rate of 69.5%.

Table 3. Survey items used in Study 2.

Survey Item	Adapted Source
What does it mean if there is a [**RANDOM RISK CATEGORY] RISK** of severe weather in your area tomorrow evening? Please provide a sentence or two interpreting the phrase [**RANDOM RISK CATEGORY] RISK**	Silva et al. 2020 (WX19)
If there is a [**RANDOM RISK CATEGORY] RISK** of severe weather in your area tomorrow evening, how likely is it that severe weather will hit within 25 miles of your residence? Use the slide scale below to indicate the probability as a percent.	Silva et al. 2020 (WX19)

The National Weather Service uses the following phrases to describe the risk of severe weather. We want to know what these phrases mean to you. Please rank them from one (lowest risk; top) to five (highest risk; bottom) below by dragging and dropping them in your desired order.	Mayhorn, Wogalter, and Shaver (2004)
The National Weather Service uses colors to describe the risk of severe weather. We want to know what these colors mean to you. Please rank them from one (lowest risk; top) to five (highest risk; bottom) below by dragging and dropping them in your desired order.	Mayhorn, Wogalter, and Shaver (2004)
The National Weather Service uses **both colors and words** to more clearly describe the risk of severe weather. We want to know what these **colors and words** mean to you when they are used together. Please rank them from one (lowest risk; top) to five (highest risk; bottom) below by dragging and dropping them in your desired order.	Mayhorn, Wogalter, and Shaver (2004)

Description of Sample

A total of 1091 undergraduate students were surveyed (Table 4). Participants' ages ranged from 18 to 30, however, a majority of the sample was aged 18 to 20 (*n* = 968, 89%). The sample contained more females (*n* = 691, 63.3%), and most of the sample identified as White or Caucasian (*n* = 881, 80.8%). Unlike the previous study, the participants were not asked to provide their highest degree earned or their annual household income, as the participants were undergraduate students. Compared to Study 1, this sample was more representative of the general public (U.S Census Bureau 2019). However, compared to population estimates in 2019, this sample does lack (1) diversity of age, (2) individuals that identify as Black or African American, (3) individuals that identify as Hispanic or Latinx, and (4) individuals that identify as male (U.S Census Bureau 2019).

We also asked participants about their weather information habits. Most of the respondents reported seeking out weather information at least once (*n* = 910, 83.4%) or multiple times per day (*n* = 509, 46.7%). The remaining participants mentioned that they search for weather information less frequently--from two or more times a week up to once or twice a month. The survey also asked participants to provide information on the frequency with which they gathered weather information through various channels. You can find more specific information in Table 5 below.

Participants were also asked more specific information about their preferred medium for searching for weather information and the amount of time they spend searching for this information. When asked which medium individuals preferred when searching for weather forecast information, a large majority indicated they prefer getting weather information from smartphone or mobile applications (*n* = 953, 87.4%). In terms of time spent searching for weather information, nearly half of the sample explained that they spent anywhere between 1 to 3 minutes searching for weather information (*n* = 541, 49.6%). However, both 0 to 1 minute (*n* =

380, 34.8%) and 3 to 5 minutes (*n* = 142, 13%), were also popular options. For the sake of comparison, participants were asked the same two questions but this time they were asked to provide details about their severe weather forecast information-seeking tendencies. As previous studies have shown (Sherman-Morris 2005), when severe weather is in the forecast, participants had a higher preference for gathering information via television (*n* = 552, 50.6%) and the Internet (*n* = 216, 19.8%). However, interestingly, getting weather information from smartphone or mobile applications remained somewhat high (*n* = 245, 22.5%). When asked to report on the amount of time spent searching for severe weather forecast information, participants indicated that they spend more time searching when severe weather is in the forecast. One-third of participants indicated that they spend anywhere from 3 to 5 minutes searching for weather information when severe weather is in the forecast (*n* = 398, 36.5%). This is followed by an equal split between individuals reporting that they spend either 5 to 10 minutes (*n* = 262, 24%) or 1 to 3 minutes (*n* = 258, 23.6%) searching for severe weather forecast information. You can find more specific information in Table 5 below.

Table 4. Basic demographic information for participants in Study 2.

Variable	*n*	%		Variable	*n*	%
Age:				**Hispanic or Latinx?**		
18	475	43.5%		Yes	64	5.9%
19	335	30.7%		No	1024	93.9%
20	158	14.7%				
21	77	7.1%		**Ethnic Identification:**		
22	25	2.3%		White	881	80.8%
23	7	0.6%		Black or African American	76	7%
24 - 30	11	1.1%		American Indian/Alaska Native	5	0.5%
				Asian	102	9.3%
Gender:				Native Hawaiian or Pacific Islander	6	0.5%
Male	394	36.1%		Other	21	1.9%
Female	691	63.3%				

Table 5. Weather information habits for participants in Study 2.

Variable	*n*	%		Variable	*n*	%
Weather Source Preference:				**Svr Wx Source Preference**:		
Smartphone Applications	953	87.4%		Smartphone Applications	245	22.5%
Social Media	26	2.4%		Social Media	39	3.6%
Internet Websites	70	6.4%		Internet Websites	216	19.8%
Friends and Family	7	0.6%		Friends and Family	15	1.4%
Watching Television	25	2.3%		Watching Television	552	50.6%
Radio	9	0.8%		Radio	24	2.2%
Time Searching for Wx Info:				**Time Searching for Svr Wx:**		
0-1 Minute	380	34.8%		0-1 Minute	66	6%
1-3 Minutes	541	49.6%		1-3 Minutes	258	23.6%
3-5 Minutes	142	13%		3-5 Minutes	398	36.5%
5-10 Minutes	26	2.4%		5-10 Minutes	262	24%
More than 10 minutes	2	0.4%		More than 10 minutes	107	9.8%

Results and Findings

Perceived Likelihood of SPC Risk Category Words

With previous studies (Ernst et al. 2021) and our own findings from Study 1 acknowledging that members of the public may experience challenges navigating the SPC Risk Category System, it was first important to explore the descriptive statistics associated with participants' perceived likelihood of each of the five risk category words. Remember that participants were randomly assigned to receive one of the five risk category words, and then were asked (1) to provide a sentence or two describing what being in that risk category means to them and (2) to indicate the likelihood that they might experience severe weather (on a scale from 0 to 100) if they were within the risk category that was assigned to them. Due to the large sample size, the research team did not have time to explore the qualitative findings associated with each risk category word; therefore, the following results will highlight the perceived likelihood associated with each of the SPC risk category words.

The descriptive statistics associated with the five risk category words reveal two patterns that similarly emerged during Study 1:

- The perceived likelihood associated with Slight (M = 31.85, SD = 16.71) was lower than Marginal (M = 38.53, SD = 19.79), even though Slight is higher than Marginal in the SPC's current risk category system. The means associated with these two categories were not significantly different.

- The perceived likelihood associated with Enhanced (M = 58.88, SD = 21.89) was significantly higher than Moderate ($p < 0.01$, M = 47.80, SD = 18.79) even though Enhanced is lower than Moderate in the SPC's current risk category system.

When compared against similar items in the 2019 Severe Weather and Society (WX19) survey (Silva et al. 2019), we see that participants in our study provided overall higher means. This could be attributed to the Severe Weather and Society survey's use of "tornado" in their question language, whereas we used "severe weather" to be more encompassing of both severe thunderstorms and tornadoes. For more information on the descriptive statistics associated with perceived likelihood, please refer to Table 6.

Table 6. Comparing our mean perceived likelihood responses to the WX19 Survey items.

SPC Risk Category Word	**Our Survey Items (~200 per condition)**		**WX19 Survey Items (~900 per condition)**	
	Mean	Standard Deviation	Mean	Standard Deviation
Marginal	38.53	19.79	**	**
Slight	31.85	16.71	24.51	21.78
Enhanced	58.88	21.89	**	**
Moderate	47.80	18.79	38.99	22.03
High	65.09	21.91	51.87	27.63
**Indicates that the WX19 Survey did not ask about the Marginal or Enhanced risk categories.				

Public Interpretation of the SPC's Risk Category System

Colors:

When asked to rank the SPC's color scheme for the five risk tiers, from least threatening to most threatening, participants' responses were much closer to SPC's official ranking (Figure 6). However, still over half of the sample (*n* = 586, 55.4%) deviated from the color sequence currently used by the SPC. A closer look at these deviations, reveals two common patterns that were similarly found in Study 1:

- 40% of participants (*n* = 436) thought that red was the most threatening color.
- 55.5% of participants (*n* = 606) thought that magenta was less threatening overall.

Even with these deviations, the most common ordering of the five colors matches the official order that is currently used by the SPC (*n* = 475, 44.6%). The second largest grouping involved participants interchanging the positions of Red and Magenta, identifying red as the most threatening color. The remaining groupings interchanged and moved Magenta to various spots within the five color ranking, indicating that some participants had difficulty interpreting Magenta in the context of the other colors. For more information on color rankings, please see Table 7.

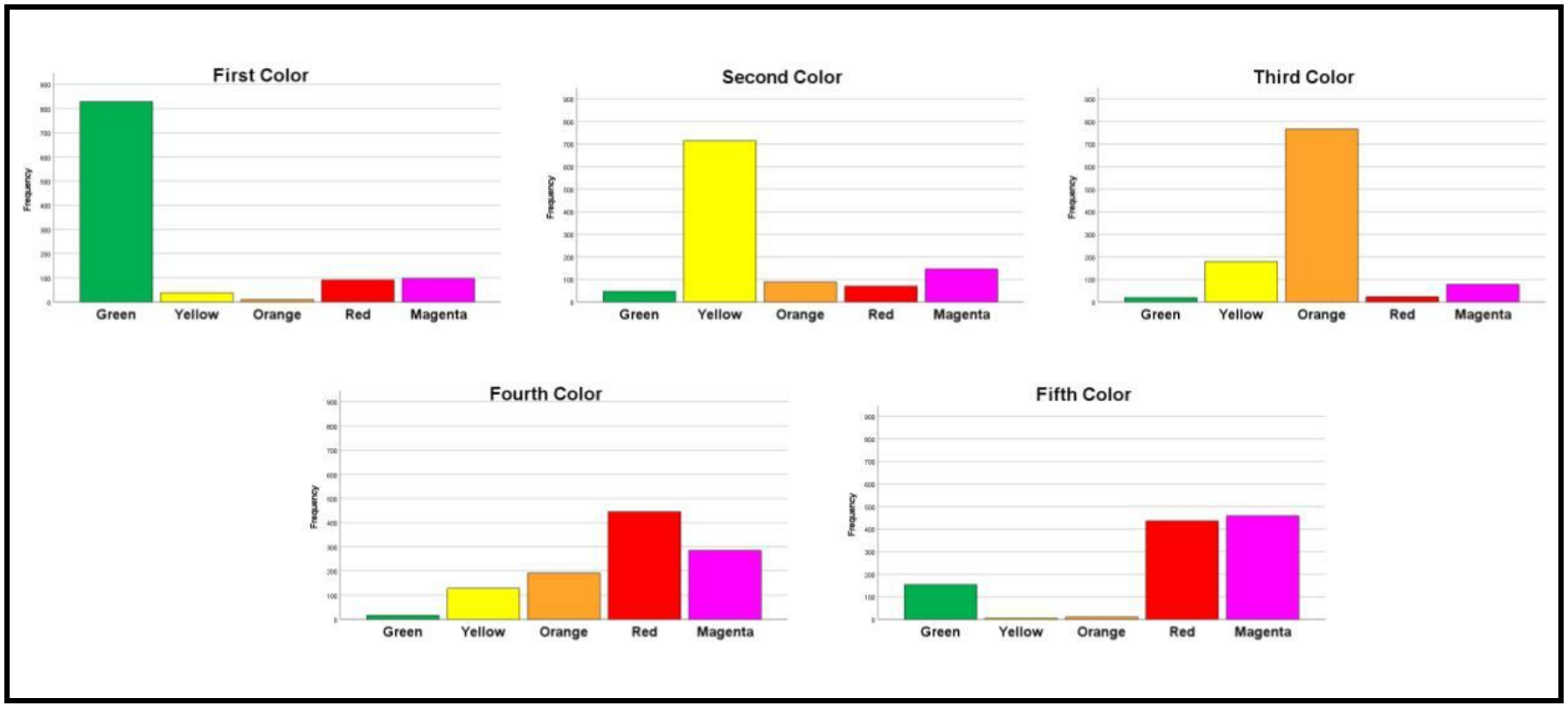


Figure 6: Distribution of colors ranked from least threatening to most threatening.

Table 7. Color rankings when asked to rank from least to most threatening.

First Color	Second Color	Third Color	Fourth Color	Fifth Color	*n*	Percent of participants (%)
Green	**Yellow**	**Orange**	**Red**	**Magenta**	475	44.6%
Green	Yellow	Orange	Magenta	Red	220	20.7%
Green	Magenta	Yellow	Orange	Red	80	7.5%
Green	Yellow	Magenta	Orange	Red	55	5.2%
Magenta	Green	Yellow	Orange	Red	23	2.2%
Note: The bolded words denote the official color ranking used by the SPC.						

Risk Categories:

When asked to rank the SPC's five risk category words from least threatening to most threatening, almost all of the participants (n = 1026, 96.7%) deviated from the word pattern that is currently being used by the SPC. When these distributions are plotted (Figure 7), the two patterns that arose in Study 1, similarly emerge here with a larger sample:

- 63.4% of participants (*n* = 692) thought that Slight was less threatening than Marginal.
- 71.2% of individuals (*n* = 777) thought that Enhanced was more threatening than Moderate.

As such, participants tended to interchange the positions of the words Marginal and Slight, as well as Moderate and Enhanced (Table 8). Thus, the most common ordering of the first four risk words was Slight, Marginal, Moderate, and Enhanced, in that order. Over 50% of the sample believed the word High to be the fifth and highest risk category. This resulted in the most common ordering of the risk categories as "Slight, Marginal, Moderate, Enhanced, High" (*n* = 283, 26.7%). The official ordering of the SPC's risk categories was the eighth largest group of participants (*n* = 35, 3.3%). The groupings that occurred above the official ordering of the SPC's risk categories included interchanging Marginal and Slight, Enhanced and Moderate, and interestingly, Enhanced and High. The second largest grouping included both Marginal and Slight swapping positions, as well as Enhanced and High (*n* = 111, 10.5%). This is rather unique and may offer some insights that Enhanced may be perceived as a more threatening word than initially anticipated. This pattern of switching both Enhanced and High is similarly observed in both the fifth and sixth largest groups as well. For more information on the risk category rankings, please see Table 8.

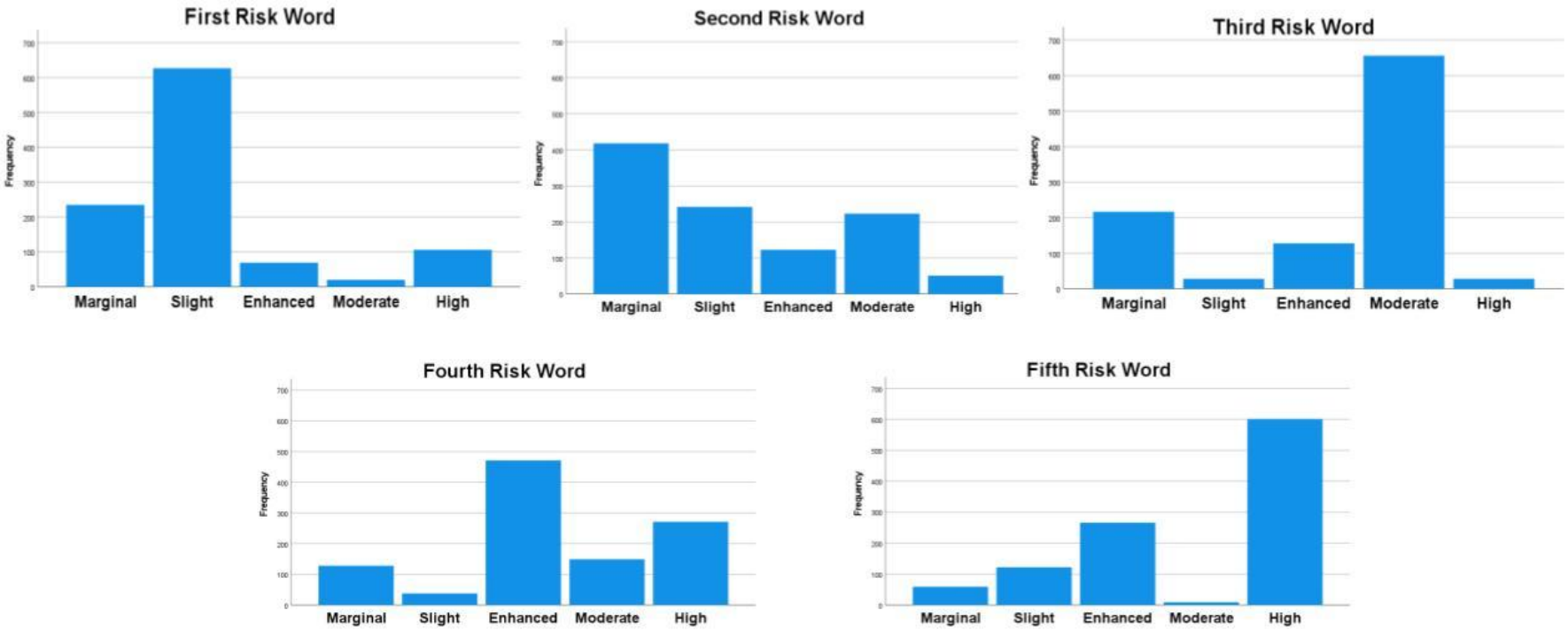


Figure 7: Distribution of risk words ranked from least threatening to most threatening.

Table 8. Risk word rankings when asked to rank from least to most threatening.

First Word	Second Word	Third Word	Fourth Word	Fifth Word	*n*	Percent of participants (%)
Slight	Marginal	Moderate	Enhanced	High	283	26.7%
Slight	Marginal	Moderate	High	Enhanced	111	10.5%
Marginal	Slight	Moderate	Enhanced	High	103	9.7%
Slight	Moderate	Marginal	Enhanced	High	95	9%
Marginal	Slight	Moderate	High	Enhanced	75	7.1%
Slight	Moderate	Marginal	High	Enhanced	63	6%
Slight	Marginal	Enhanced	Moderate	High	47	4.4%
Marginal	**Slight**	**Enhanced**	**Moderate**	**High**	35	3.3%
Note: The bolded words denote the official risk category ranking used by the SPC.						

Combination of Colors and Risk Categories:

Recall that when the new SPC risk categories became operational, SPC's leadership made a conscious decision to use colors, numbers, and risk category words to provide multiple cues to help communicate severe weather risk to a variety of end users. Therefore, in Study 1, the research team evaluated whether combining colors *and* risk categories would improve the intuitiveness of the SPC's current risk category system among members of the public. The small qualitative sample of Athens, GA residents revealed mixed results when presenting participants with two distinct cues. As a result, the research team was interested in conducting a similar online card sorting task with a larger sample of participants. When performed on a larger sample of participants, this card sorting task revealed that combining both colors *and* risk category words did improve the intuitiveness of the SPC's risk category system:

- 67% of participants (*n* = 697) improved their risk category ranking with the addition of color as a secondary cue.
- 38.7% of participants (*n* = 422) improved their risk category ranking with the addition of color as a secondary cue AND matched the official ranking currently used by the SPC.
- 26.7% of participants (*n* = 291) kept their risk category ranking the same when color was added as a secondary cue.
- The risk category ranking of 68 individuals (6.2%) got worse when color was added as a secondary cue.

Compared to the qualitative findings in Study 1, it is very clear from these findings that including both colors and risk category words is important for improving the intuitiveness of the SPC's risk

category system. A quick look at the groupings (Table 9) reveals that even with the addition of color, some of the issues with the risk category words still persist. The second and third most common groupings, for example, showcase that participants are still interchanging both Marginal and Slight, as well as Moderate and Enhanced. Even though adding color does improve the intuitiveness of the SPC's risk category system, those word perceptions still remain. For more information on the color *and* risk category rankings, please see Table 9.

Table 9. Color and risk word rankings when asked to rank from least to most threatening

First Word	Second Word	Third Word	Fourth Word	Fifth Word	*n*	Percent of participants (%)
Marginal	**Slight**	**Enhanced**	**Moderate**	**High**	452	42.5%
Marginal	Slight	Moderate	Enhanced	High	94	8.8%
Slight	Marginal	Moderate	Enhanced	High	80	7.5%
Slight	Marginal	Enhanced	Moderate	High	62	5.8%
Marginal	Slight	Enhanced	High	Moderate	47	4.4%
Note: The bolded words denote the official ranking used by the SPC.						

Summary

In summary, the results we presented in this section confirm many of the findings that emerged in Study 1; however, there were a few differences, that when considered in the context of a larger sample, become noteworthy and began to offer some clarity on the usability and intuitiveness of the SPC's risk category system: (1) Most participants deviated from the SPC's official ranking of the risk category words. This was observed in Study 1, and similarly, the larger sample also frequently interchanged both Marginal and Slight and Enhanced and Moderate. These results suggest that both Marginal and Enhanced are likely the most challenging words to interpret, and may lead to wide variations in severe weather risk interpretation by the general public. (2) A majority of individuals deviated from the SPC's official color ranking; however, contrary to Study 1, a larger number of individuals ranked colors more accurately. Those that deviated tended to either rank Red as the most threatening color and/or struggled to rank Magenta against the other colors--resulting in Magenta appearing both high and low on participant's rankings. These patterns were similarly observed in Study 1. Finally, (3) most participants improved their ranking of the risk category words when color was added as a secondary cue. This finding provides a bit more clarity to the results found in Study 1, and highlights the value of color (and a secondary cue) when orienting the general public to the SPC's risk category system.

STUDY 3

Study 1 and Study 2 similarly highlighted that members of the public experienced usability challenges when attempting to navigate the SPC's risk category system using risk words alone. A new research article, by Ernst et al. (2021), came to the same conclusion. Across all three studies, members of the public frequently interchanged the positions of the words Marginal and Slight, as well as Moderate and Enhanced when using the SPC's risk category system. With these convergent findings, the research team decided that the next area worth exploring was whether there were any changes that could be made to improve the usability of the SPC's risk category system. Therefore, drawing on the methodologies of the 2019 Severe Weather and Society Survey (Silva et al. 2020) and a study by MacLeod and Pietravalle (2017), Study 3 asked members of the public to quantitatively rank commonly used risk words to determine if there are any suitable alternatives that may help alleviate some of the severe weather interpretation challenges that were found in previous studies.

Methodology

Survey with Members of the Public

This study was reviewed and approved prior to its beginning in October 2019 by the University of Georgia Institutional Review Board (IRB; PROJECT00001227). The IRB was modified to change the scope of the project and expand the targeted population to a representative sample of the general public. These changes were approved by the University of Georgia IRB on July 31, 2020.

Recruiting Participants

Members of the public were solicited by Qualtrics, a survey sample company, to complete a survey on severe weather graphics and messaging. In particular, the research team purchased a demographically represented survey panel from Qualtrics. Therefore, Qualtrics contacted individuals via email who have expressed interest in completing surveys in the past and were then dynamically sampled to generate a representative sample based on U.S. Census data (U.S Census Bureau 2019). Qualtrics respondents received an incentive based on the length of the survey, their specific panelist profile, and target acquisition difficulty.

Survey Instrument and Procedure

Although the original intent of the survey was to further evaluate the perceived consistency between two severe weather graphics (these results will be discussed in the second report), an additional goal of the survey was to explore commonly used risk words to determine if there are any suitable risk category alternatives that may help alleviate some of the severe weather risk

interpretation challenges that were found in both Study 1 and Study 2. Before being asked to evaluate the consistency of two Convective Outlook graphics, participants were asked to evaluate common words or phrases that are currently in use or have been used in the past to communicate risk.

To determine the words or phrases that participants would evaluate, the research team explored various research articles, real-world risk messaging and risk category systems across a variety of hazards, as well as other health, environmental, and science communication contexts. This revealed a total of 19 risk words or phrases. After further discussion with the research team, we determined that asking members of the public to evaluate all 19 risk words or phrases would be cognitively demanding. As a result, we decided to narrow down the list of risk words or phrases in order to provide a shortened list for evaluation. Through discussions among the research team, it was determined that many of the risk words or phrases were on the higher end of the scale. Therefore, when shortening the list, several of the higher end words were considered for removal to balance the risk words or phrases that remained. Finally, when considering the finalized list of words and/or phrases, we consulted with a bilingual weather researcher who provided direct translations for all of the risk words and phrases we had collected (Joseph Trujillo-Falcon, personal communication, March 2, 2020). Given the ongoing diversity, equity, and inclusions (DEI) efforts to make weather risk information more accessible to non-English speaking audiences, the research team decided to remove certain risk words and/or phrases that did not translate well into Spanish--as the Spanish language is the second most common language spoken in the U.S. (Census 2019). Given the ongoing Spanish language initiatives happening within NOAA at this time and the number of people in the U.S. speaking Spanish, this language was prioritized; however, other language translations should be considered for future social science research. This elimination process left 13 risk words or phrases for members of the public to evaluate. These words and/or phrases can be found in Table 10 below.

Table 10. Thirteen risk words participants were asked to evaluate and their equivalent Spanish translation.

Risk Word	Spanish Translation
Very Low	Muy Bajo
Minimal	Riesgo Minimo
Low	Riesgo Bajo
Marginal	Riesgo Marginal (NWS), Riesgo Mínimo (SPC)
Slight	Riesgo Leve (NWS and SPC), Riesgo Bajo (Linguistic Experts)
Enhanced	Riesgo Aumentado/Realzado (NWS), Riesgo Moderado (Linguistic Experts), Riesgo Elevado (SPC)
Medium	Riesgo Mediano
Moderate	Riesgo Moderado
High	Riesgo Alto (NWS, SPC) Riesgo Extremo (Linguistic Experts)
Very High	Muy Alto Riesgo
Extreme	Riesgo Extremo
Considerable	Considerable
Catastrophic	Catastrofico
Note: For more information on the Spanish translations recommended by linguistic experts vs. the NWS, please refer to Trujillo-Falcon et al. (2021)	

Using both the question wording from the 2019 Severe Weather Society Survey (Silva et al. 2020) and a study by MacLeod and Pietravalle (2017), two survey questions were adapted to elicit the perceived likelihood and the perceived severity that was associated with the 13 risk words outlined above. Perceived risk is often conceptualized as perceived susceptibility and perceived severity (Brewer, Weinstein, Cuite, and Herrington 2004; Witte 1992); therefore, these two dimensions were explored to determine if there were any differences in each of these dimensions across the 13 words. At the beginning of the survey, participants were randomly assigned to complete either perceived likelihood or perceived severity. Participants were then asked to provide their perceived likelihood or severity for the 13 risk words. The risk words were randomized on the page to avoid any order effects. For additional information on the survey items used in this study, please refer to Table 11.

Qualtrics launched the survey on 20 August 2020 by sending an email to their list of recipients that align with our demographically representative sample of the U.S. population. We completed data collection on 26 August 2020. After reviewing the data, however, the research team discovered some inconsistencies and poor data quality associated with this survey item. Therefore, we went back to Qualtrics and reopened the data collection process to receive additional responses that provided a higher quality of data. Therefore, we officially closed the survey on 3 September 2020. At the close of the survey, we had received 1041 usable (complete) responses.

Table 11. Survey items used in Study 3.

Survey Item	Adapted Source
When forecasting the weather, meteorologists use different phrases to describe the ***LIKELIHOOD*** that severe weather will hit within 25 miles of your area. When you see the following phrases, what percent chance comes to mind? Please use the sliding scale below to indicate the likelihood of each phrase as a percent that ranges from 0 to 100, where 0 means not at all likely and 100 means very likely.	Silva et al. (2020) - WX19 MacLeod and Pietravalle (2017)
When forecasting the weather, meteorologists use different phrases to describe the ***SEVERITY*** that severe weather will hit/impact your area. When you see the following phrases, what level of severity comes to mind? Please use the sliding scale below to indicate the severity of each phrase as a percent that ranges from 0 to 100, where 0 means not at all severe and 100 means very severe.	Silva et al. (2020) - WX19 MacLeod and Pietravalle (2017)

Description of Sample

A total of 1044 members of the general public were surveyed. Participants' ages ranged from 18 to 88, with age being represented fairly equally across all age groupings. However, the sample had lower representation from younger adults (18-24, *n* = 62, 6%) and adults 75 and older (*n* = 58, 5.6%). The sample contained more females (*n* = 636, 61.1%), and a majority of survey respondents identified as White or Caucasian (*n* = 791, 76%). There were 100 participants (9.6%) who identified as being of a Hispanic or Latinx origin. Participants were asked to provide their highest educational degree earned, with a bachelor's degree (*n* = 346, 33.2%), some college credit (*n* = 227, 21.8%), and a master's degree (*n* = 203, 19.5%) being the most represented categories in the sample. As a result, the sample was more educated than the average public (Census 2019). In terms of household income, the sample was skewed more toward a higher income ($80,000 to $150,000+; *n* = 417, 40%) audience. For more information on the demographics associated with Study 3, please see Table 12 below.

We also asked participants about their weather information habits when severe thunderstorms are in the forecast. Survey respondents were asked for specific information about their preferred

medium for searching for weather information, the number of sources they use, and the amount of time they spend searching for weather information when severe thunderstorms are in the forecast. When asked which medium individuals preferred, a large majority indicated that they preferred watching weather on the television to get the latest severe thunderstorm forecast information (*n* = 542, 52.1%). The other two most frequently used mediums included getting weather information from a smartphone or mobile application (*n* = 218, 20.9%) and getting weather information from the Internet (*n* = 175, 16.8%). Survey respondents were next asked the number of sources they usually consult when searching for weather information when severe thunderstorms are in the forecast. Most individuals indicated that they use at least two sources when searching for severe thunderstorm information (*n* = 509, 48.9%). Only 20 individuals indicated that they stick to only one source when searching for the latest weather information. In terms of time spent searching for weather information when severe thunderstorms are in the forecast, the participants provided a lot of variation in their responses. The most frequently chosen amount of time was 5-10 minutes (*n* = 305, 29.3%); however, 3-5 minutes (*n* = 286, 27.5%) and 1-3 minutes (*n* = 186, 17.9%) were not far behind. Surprisingly, over 10% of respondents (*n* = 110) indicated that they spend anywhere from 10-20 minutes searching for weather information when severe thunderstorms are in the forecast. For more specific information on the sample's weather information habits, please see Table 13 below.

Table 12. Basic demographic information for participants in Study 3.

Variable	*n*	%		Variable	*n*	%
Age:				**Hispanic or Latinx?**		
18-24	62	6%		Yes	100	9.6%
25-34	133	12.8%		No	941	90.4%
35-44	194	18.6%				
45-54	164	15.8%		**Ethnic Identification:**		
55-64	212	20.4%		White	791	76%
65-74	218	20.9%		Black or African American	129	12.4%
75+	58	5.6%		American Indian/Alaska Native	12	1.2%
				Asian	68	6.5%
Gender:				Native Hawaiian or Pacific Islander	3	0.3%
Male	403	38.7%		Other	38	2.1%
Female	636	61.1%				
Education:				**Income:**		
No HS Diploma	10	1%		Low (<$39,000)	295	28.3%
HS Graduate	144	13.8%		Middle ($40,000 - $79,999)	329	31.6%
Some College	227	21.8%		High ($80,000 - $150,000)	417	40%
College Graduate	346	33.2%				
Master's Degree	203	19.5%				
Doctorate or Professional	27	2.6%				

Table 13. Weather information habits for participants in Study 3.

Variable	*n*	%
Svr Weather Source Preference:		
Smartphone Applications	218	20.9%
Social Media	27	2.6%
Internet Websites	175	16.8%
Friends and Family	6	0.6%
Watching Television	542	52.1%
Radio	73	7%
Time Searching for Wx Info:		
0-1 Minute	48	4.6%
1-3 Minutes	186	17.9%
3-5 Minutes	286	27.5%
5-10 Minutes	305	29.5%
10 - 20 Minutes	110	10.6%
20 - 30 Minutes	50	4.8%
30 minutes to 1 hour	36	3.5%
1 - 2 hours	16	1.5%
Other	4	0.4%

Results and Findings

Perceived Likelihood of Risk Words

Table 14 shows the words and phrases used to describe perceived likelihood on a scale from 0 to 100 in ascending rank order of mean score with the score range, interquartile range, and number of responses included as well for comparison purposes. There are several interesting patterns that similarly arose in Study 1 and Study 2. First, the interchange between Marginal and Slight, and Moderate and Enhanced is seen yet again in this study. According to participants, the word 'Slight' was

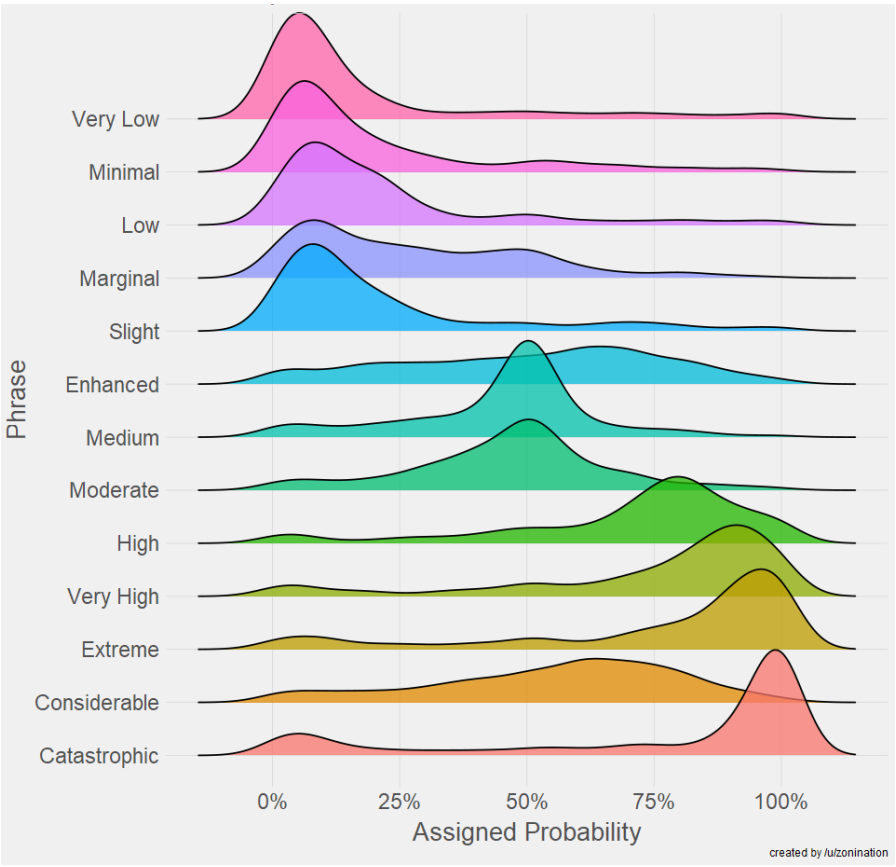


Figure 8: Spread of the data for perceptions of extreme weather likelihood.

associated with a lower perceived likelihood than ‘Marginal’ and, similarly, the word ‘Moderate’ was associated with a lower perceived likelihood than ‘Enhanced.” The interquartile range (IQR) was also calculated to better understand the spread of the data (Figure 8). For example, we see that Marginal, Enhanced, Considerable, and Catastrophic have a relatively higher IQR than many of the other words. This indicates that there was more variability in the data, and that there was a lack of agreement about the perceived likelihood of these words. Therefore, in addition to participants interchanging the perceived likelihood of both ‘Marginal’ and ‘Enhanced’ from their official counterparts in the SPC risk category system, these data also suggest that ‘Marginal’ and ‘Enhanced’ may have different meanings to different people when communicating severe weather likelihood to the public. Although there is less research on the use of the words ‘Considerable’ and ‘Catastrophic’ in the weather enterprise, given their large IQR, it is likely these risk words are just as ambiguous when trying to communicate severe weather likelihood.

Perceived Severity of Risk Words

Table 14 shows the words and phrases used to describe perceived severity on a scale from 0 to 100 in ascending rank order of mean score with the score range, interquartile range, and number of responses included as well for comparison purposes. When compared against perceived likelihood, for the most part, the means and IQRs of all the risk words remain relatively similar. In terms of the overall means, the only two sets of words that switched places between perceived likelihood and perceived severity is Medium vs. Moderate and Extreme vs. Catastrophic (Figure 9). When examined more closely in terms of perceived likelihood, both Medium and Moderate share a very similar mean. Therefore, the interchange between the two when examining perceived severity is not as meaningful. However, a closer look at the IQRs that exist between Extreme and Catastrophic across perceived likelihood and severity provides a more interesting finding.

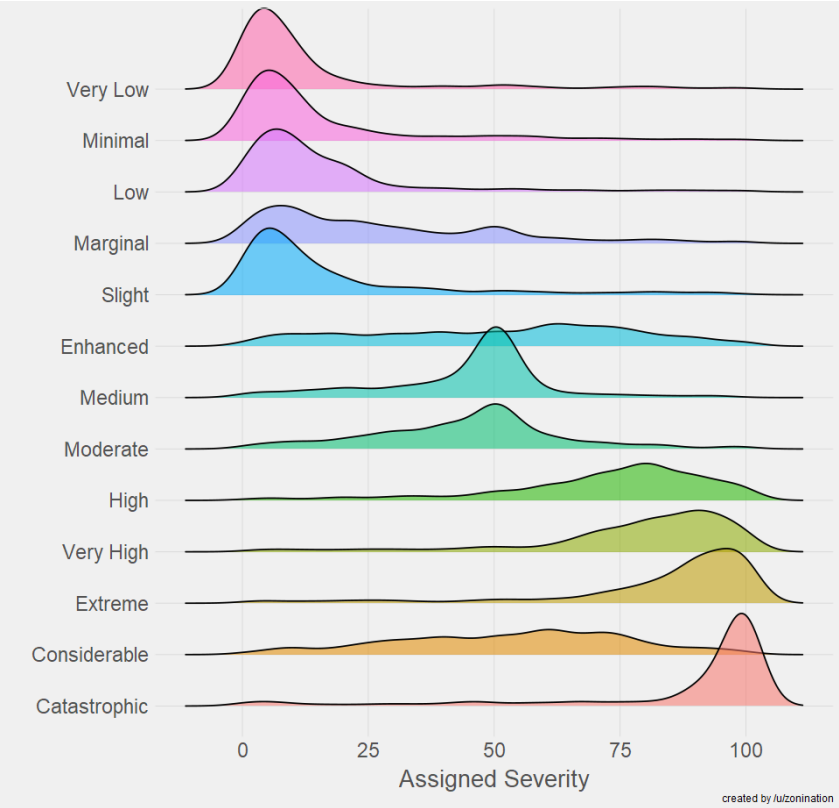


Figure 9: Spread of the data for perceptions of extreme weather severity.

In terms of perceived likelihood, the term ‘Catastrophic’ provided a high IQR indicating that there was less consistency when participants rated the perceived likelihood of this phrase. When looking at the IQR for perceived severity, however, ‘Catastrophic’ now has a very low IQR. This likely suggests that participants had difficulty interpreting the term ‘Catastrophic’ in terms of perceived likelihood, but when asked to consider it in terms of perceived severity, the term ‘Catastrophic’ made more conceptual sense (i.e., there was more agreement around the mean). Since we did not see any other dramatic decreases or increases in the IQRs of other risk words, it is likely that those risk words work when communicating both perceived likelihood and perceived severity. However, it is important to remember that when risk words have a high IQR, they are not as intuitive as other risk words. As a result, there are still some terms that participants do not have as much agreement around when rating perceived severity (i.e., Marginal, Enhanced, and Considerable)

Table 14. Perceived likelihood and perceived severity for each of the thirteen risk words on a scale from 0 to 100.

Perceived Likelihood				*Perceived Severity*			
Risk Word	**Mean**	**SD**	**IQR**	**Risk Word**	**Mean**	**SD**	**IQR**
Very Low	18.27	24.59	16	Very Low	13.20	19.56	8
Minimal	20.31	22.75	20.75	Minimal	15.83	19.91	14
Low	21.67	22.50	17.75	Low	16.30	18.77	15
Slight	21.80	23.54	19	Slight	17.30	21.21	15
Marginal	28.01	22.14	34	Marginal	27.39	23.24	34
Medium	44.90	18.44	14.75	Moderate	44.35	19.00	21
Moderate	46.41	19.12	18.75	Medium	44.45	16.80	14
Enhanced	50.23	25.70	40*	Enhanced	50.68	26.30	41
Considerable	55.12	23.88	32*	Considerable	53.10	22.82	34
High	70.51	23.83	23	High	72.01	20.79	24
Very High	74.35	26.97	27	Very High	77.79	20.83	20
Catastrophic	74.82	35.05	45	Extreme	82.78	21.87	18
Extreme	75.70	29.57	29	Catastrophic	86.72	24.90	10

Hierarchical Cluster Analysis

Although the raw scores and statistics offer meaningful findings, we also wanted to conduct some higher order statistical analyses to better understand how these risk category words are grouped or connected. We believe this will provide some insight to the Storm Prediction Center on how to best optimize their risk category system. In particular, we used a hierarchical clustering analysis to group similar risk words into groups called clusters (Everitt 1993, MacLead and Pietravalle 2017). Using the mean perceived likelihood scores, risk words are first treated as individual clusters and are then slowly merged together based on similarity. The similarity between two clusters of words is defined as the smallest distance between any two points from the two clusters (i.e., single linkage clustering). This process is continued until all of the clusters are merged together. This results in a variety of dendrograms (Figure 10), which provide a visual representation of the hierarchical relationship that exists between the different clusters.

We have plotted the results of a number of different clusters (*Cluster size = 2, 3, 4, 5)* to compare the results (Figure 10). When examining the clustering dendrogram that separated the risk words into two distinct clusters, we see that the low words are separated from the high words. However, if you look closely towards the top of the diagram, it appears as though the middle-ranged risk words (i.e., Considerable, Enhanced, Medium, and Moderate) are close to forming their own branch. This likely indicates that we may need to look at the next cluster to see if this grouping emerges. When examining the clustering dendrogram that separated the risk words into three distinct clusters, we see that the middle-range group does emerge as their own cluster. When looking across the three distinct groupings, there are not any other obvious groupings that still need to emerge. However, it is important to still look at higher order groupings to determine whether any additional groups may form. A closer look at the clustering dendrograms for both the four clusters and five clusters, does not reveal any additional groups. It does, however, continue to highlight that participants' perceived likelihood responses for "Enhanced" and "Considerable" were very spread out. This can be seen as both "Enhanced" and "Considerable" form their own groupings in the four cluster and five cluster dendrograms. Not only that, but the dendrograms similarly show that participants' perceived likelihood responses for "Marginal" were very spread out as well. This is indicated by the long lines in each dendrogram that connect "Marginal" to the other low risk words.

In sum, it appears that the three cluster dendrogram best describes these data. This means that the risk words that participants were asked to examine are only different enough to form three distinct categories. This has implications for the SPC's risk category system, as it currently requires five distinct risk categories to convey severe weather risk to partners and the public. However, these data reveal only that only three levels of risk exist for the risk words that were used in this study. As such, if the Storm Prediction Center aims to continue using risk category words to communicate severe weather risk, there are likely no such risk words that can be used to uniquely communicate five levels of risk. The risk words that are commonly used in practice today (i.e., those evaluated in this study) only differentiate into three levels--low, medium, and

high. This may be adding to the interpretation challenges surrounding both ‘Marginal’ and ‘Enhanced.’

Due to the similarity between both perceived likelihood and perceived severity, only one clustering analysis was performed. However, recall that “Catastrophic” had a smaller IQR and, as a result, a reduced spread. Therefore, it is likely that in the perceived severity clustering analysis, that the word “Catastrophic” would be considered more similar to the higher risk words than in the perceived likelihood clustering analysis. Otherwise, the clustering analysis should be relatively the same.

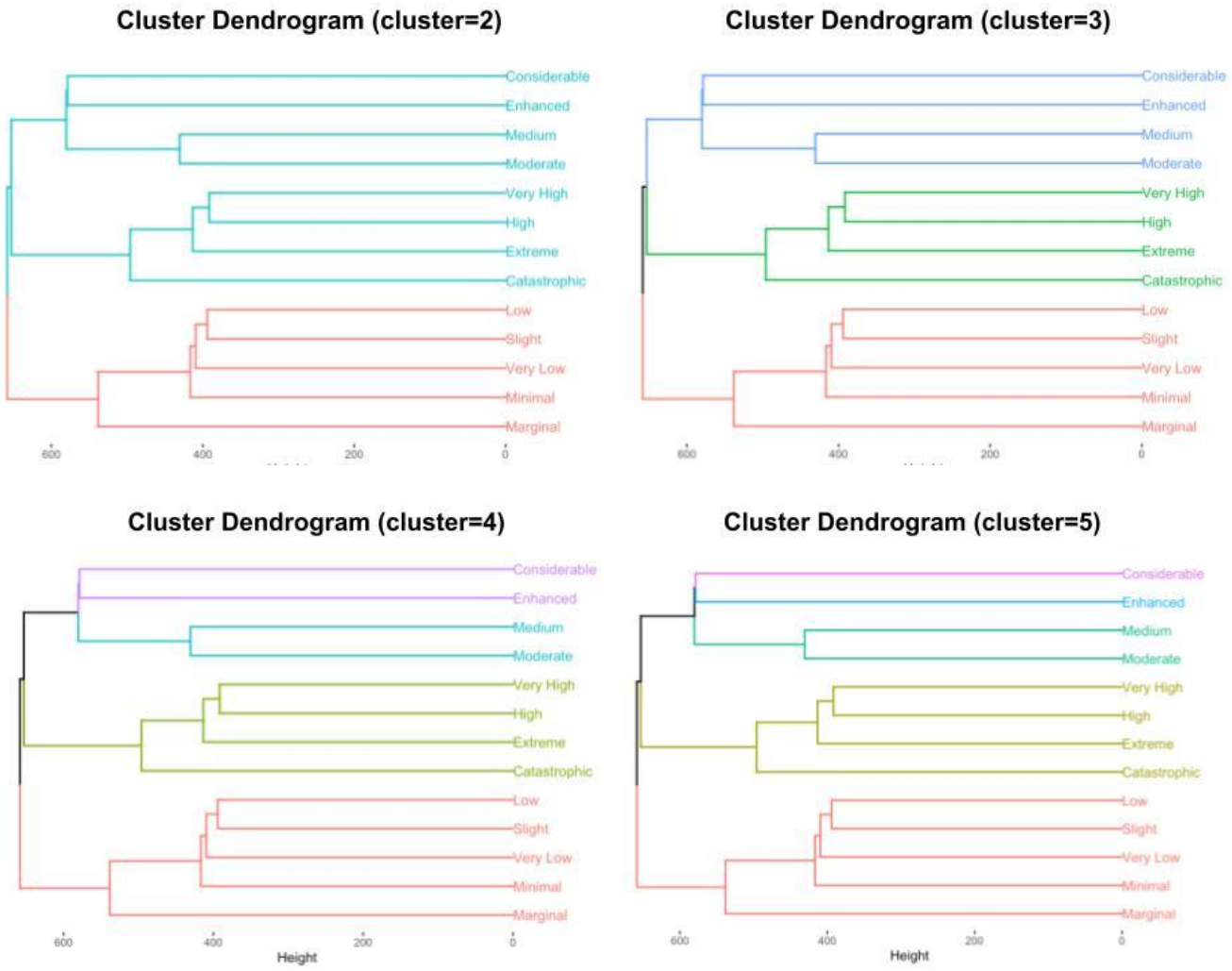


Figure 10: Hierarchical cluster analysis dendrograms for clusters 2, 3, 4, and 5.

Summary

In summary, the results we presented in this section confirm many of the findings that emerged in Study 1 and Study 2; however, findings from this study explored a variety of different risk words to investigate whether there are any suitable alternatives to overcome the observed challenges with both ‘Marginal’ and ‘Enhanced.’ Therefore, the following findings are offered: (1) As observed in Study 1 and Study 2, participants similarly interchanged ‘Marginal’ and ‘Slight’ and ‘Enhanced’ and ‘Moderate.’ According to participants, the word ‘Slight’ was associated with both a lower perceived likelihood and perceived severity than ‘Marginal,’ and similarly, the word ‘Moderate’ was associated with both a lower perceived likelihood and perceived severity than ‘Enhanced.’ (2) The analysis also revealed that ‘Marginal’ and ‘Enhanced’ had higher interquartile ranges (IQR). A high IQR indicates more variability in the data, and a lack of agreement around the mean. Therefore, in addition to participants interchanging the perceived likelihood and perceived severity of both ‘Marginal’ and ‘Enhanced,’ these data also suggest that these risk words likely have different meanings to people when communicating severe weather risk information. (3) When making comparisons using the raw statistics across both perceived likelihood and perceived severity, there are not many differences. This suggests that these words are similarly interpreted for both dimensions of risk. However, the term ‘Catastrophic’ provided a high IQR for perceived likelihood and a low IQR for perceived severity. This suggests that participants had difficulty interpreting the term ‘Catastrophic’ in terms of perceived likelihood, but when asked to consider it in terms of perceived severity, the term ‘Catastrophic’ made more conceptual sense (i.e., there was more agreement around the mean). This difference was not observed with any of the other words. (4) Hierarchical cluster analyses revealed that the three cluster dendrogram best describes these data. This means that the risk words that participants were asked to examine are only different enough to form three distinct categories. This has implications for the SPC’s risk category system, as it currently requires five distinct risk categories to convey severe weather risk to partners and the public. As such, if the SPC aims to continue using risk category words to communicate severe weather risk, there are likely no such risk words that can be used to uniquely communicate five levels of risk. The risk words that are commonly used in practice today (i.e., those evaluated in this study) only differentiate into three levels--low, medium, and high.

Triangulated Findings & Research-Guided Recommendations

After looking across all three studies, the following section will provide some triangulated findings and research-guided recommendations on how to improve severe weather risk communication among members of the public. The conclusions and research-guided recommendations that are provided below are listed in order based on how often they appeared across the three studies.

Conclusion #1: There is strong evidence that the words used in the SPC's risk category system are not easily understood by members of the public. The results of Study 1, Study 2, and the research by Ernst et al. (2021) all show that members of the public frequently interchanged both Marginal and Slight, as well as Enhanced and Moderate. Similarly, Study 3 provides additional insight suggesting that both Marginal and Enhanced are likely the most challenging words to interpret, and may lead to variability in severe weather risk interpretation based on their large interquartile ranges (IQR). Therefore, in addition to participants interchanging both Marginal and Enhanced, Study 3 also suggests that these risk words likely have different meanings to people when communicating severe weather risk information.

- **Recommendation: Additional social science research is needed to make the current risk category system more intuitive for *all* end users.** Although Study 3 provided some insight on possible risk category alternatives for Enhanced and Marginal, additional social science research is needed to experimentally evaluate new risk category system prototypes. Not only that, but specific metrics must be identified ahead of the prototype evaluation to determine how best to conceptualize "improvement" between the current risk category system and the prototype risk category system(s). Identifying and documenting these metrics will provide support for any operational changes or decisions that are made. These efforts should also include *all* end users. Given the ongoing diversity, equity, inclusion, and accessibility efforts, the risk category prototypes must consider language translation and ensure that any new system considerations are thoroughly tested among a diverse set of end users.

- **Recommendation: Consult emergency managers, broadcast meteorologists, and other core partners on any proposed operational changes to the risk category words.** Although most of the research to-date on the SPC's risk category system has involved emergency managers, broadcast meteorologists, and other core partners, it will be incredibly important to include them when evaluating any changes to the SPC's risk category system. Not only will it be important to similarly evaluate the experimental prototypes with these users, but it will also be necessary to qualitatively explore this topic with core partners to obtain feedback on how any changes will affect their products, processes, and policies.

- **Recommendation: Risk levels represent a promising alternative to using risk category words.** Although the current studies did not explore the use of risk levels (i.e., Level 1) as a replacement for risk category words, findings from Study 3 suggest that there are likely no risk words that can be used to uniquely communicate five levels of risk. Based on the hierarchical cluster analysis, a variety of risk words that are used in practice today only differentiate into three levels—low, medium, and high. Therefore, if the SPC aims to continue using risk category words to communicate severe weather risk, they will struggle to find five different words that convey those five levels of risk. This is why using numerical risk levels might be a promising alternative. However, more social science research is needed before this can be implemented operationally.

- **Recommendation: Reevaluate whether the 'Thunder' category should be included in the Convective Outlook graphic and, if so, consider changing its name.** In Study 1, many participants expressed concerns with the 'Thunder' category. In particular, individuals were confused by the purpose and/or meaning of this category. Therefore, it is recommended that the SPC reconsider the value of placing this category on the Convective Outlook graphic. If it provides valuable information, then a name change is recommended. Because individuals questioned whether this would be the only zone that could expect thunder, it is recommended that the category be renamed to "Thunderstorms" or "Non-Severe." This would provide additional clarity on the threats associated with this category, but also what makes it unique in comparison to the other risk categories.

Conclusion #2: Some participants had difficulties interpreting magenta in the context of the other risk category colors. Although less prominent in Study 2, both the results from Study 1 and Study 2 showcase the challenges that participants faced when trying to rank magenta against the other risk category colors. In the qualitative study of the SPC risk category system, a majority of participants (n = 19, 63%) thought that red was a more threatening color than magenta and that "pink is not a threatening color." A similar pattern emerged in Study 2 with a larger sample of participants, however this time, it was a smaller portion of participants that perceived red to be more threatening than magenta (n = 436, 40%). Even though it was not a pattern favored by a majority of participants, it is still a large enough proportion to be noteworthy. Not only that, but a closer look at the color rankings in Study 2 (Table 7, pg. 30) reveal that participants simply interchanged and moved Magenta to various spots within the five color ranking. This adds to the evidence that some participants had difficulty interpreting magenta in the context of other colors. This is concerning considering this misinterpretation is occurring at the high end of the risk category scale.

- **Recommendation: Additional social science research is needed to determine if there is a color alternative that would better communicate high end weather threats.** There is not enough social science evidence at this time to recommend that the SPC should alter their color scheme. However, it is suggested that additional social science research explore color alternatives to determine if they can better communicate high end weather threats. In particular, this social science research might also inform

NOAA/NWS' ongoing consistent depiction of risk efforts. Therefore, prior to finalizing the policies surrounding those efforts, additional social science research is warranted that provides theoretical advancements on how to effectively use color for scales, indices, categories, and/or risk risk levels.

Conclusion #3: There is emerging evidence that combining colors and risk category words does improve the intuitiveness of the SPC's risk category system. Recall that when the new SPC risk categories became operational, SPC's leadership made a conscious decision to use colors, numbers, and risk category words to provide multiple cues to help communicate severe weather risk. Therefore, the research team evaluated whether combining colors *and* risk categories would improve the intuitiveness of the SPC's risk category system among members of the public. The small qualitative sample in Study 1 revealed mixed results when presenting participants with two distinct cues, however in Study 2, it became obvious that combining both colors and risk category words does improve the intuitiveness of the SPC's risk category system. This was seen in Study 2 as a large majority of participants (67%) improved their risk category ranking when color was added as a secondary cue. Not only that, but almost 40% of people improved their risk category ranking *and* matched the official risk category word ranking used by the Storm Prediction Center when color was included as a secondary cue. That means there was a 35% increase in individuals correctly matching the SPC's official ranking when color was also used. However, this does not mean that the previously documented challenges with the SPC's risk category words simply disappear. A closer look reveals that the word misinterpretations still exist among a smaller portion of the sample.

- **Recommendation: Continue using colors, numbers, and risk category labels in all SPC risk communication with the general public.** Although the SPC currently uses multiple cues in their severe weather risk messaging and graphical products, this research showcases the value of using color in addition to risk category words when communicating severe weather risk. This also means that the use of color and numbers should continue, even when researchers explore risk category alternatives and experimental prototypes.

Conclusion #4: The qualitative study revealed that participants experienced several interpretation challenges while attempting to use and interpret the Day 1 Convective Outlook graphic. Although a majority of participants were able to navigate the Day 1 Convective Outlook and obtain information that they felt was helpful for making decisions, there were several interpretation challenges that resulted in participants incorrectly interpreting information that is not necessarily being communicated by the graphical product. For example, when asked to interpret their severe weather risk, some individuals interpreted the graphic in terms of likelihood that their location might experience severe weather whereas others described it in terms of severity. A closer look revealed that some participants were using color as a proxy for severity information. Another common misinterpretation was the assumption that a storm's movement or motion would affect the risk category of their location. Therefore, a lot of participants looked downstream from their location to determine their severe weather potential. After examining the results a bit closer, the research team questioned whether participants

might be connecting or comparing the Day 1 Convective Outlook with the graphical output of a weather radar. This misinterpretation may also explain why some participants associated color with the intensity of the severe weather potential.

- **Recommendation: When communicating with the public and sharing information on social media, the SPC should use the Public Severe Weather Outlook graphic more frequently.** With more participants preferring the PWO in Study 1, and several commenting that it was "a lot cleaner, crisper, and easier to follow" it is recommended that the design of the National SPC Categorical Outlook graphic be used less frequently for public communication. Not only that, but participants struggled to understand and interpret the abbreviations in the legend of the National SPC Categorical Outlook graphic (e.g., SLGT = Slight). These graphics can still be used for more expert audiences; however, when publishing graphics to social media and other channels frequently used by the public, it is recommended that the SPC use the PWO design (see Figure 2). *Note:* This does not mean that the SPC should stop sharing national graphics. Instead, the SPC should promote national graphics that use the PWO's streamlined design.

- **Recommendation: Emphasize the use of "today," time updated, and valid time until on Public Severe Weather Outlook graphics.** In an attempt to overcome the perception that storm movement plays a large role in an individual's risk perception, it is recommended that the SPC consider using "today" in the title of the graphic or the date, the time the graphic was updated, and the time that the graphic is valid until on the PWO graphic. Currently, if a PWO graphic contains these pieces of information, they are often tiny and underemphasized. Perhaps increasing their visibility could better convey that this is a static graphic. However, as an enterprise, meteorologists have sought to convey that the atmosphere is always in motion. As such, many weather products can be put into motion. Based on the findings from this study, it is anticipated that the pervasiveness of the weather radar and weather products may be negatively affecting an individual's ability to consider this graphic in a static state. Beyond these recommendations, other ways to communicate that an individual's severe weather risk lasts throughout the day should be considered. For example, the SPC might consider adding timing information. This would provide additional context clues and help the user understand when they are most at risk for severe weather.

- **Recommendation: Emphasize the threat or hazards being depicted by the graphic, and/or change the title of the PWO graphic to "Severe Storm Outlook."** Many participants in Study 1 struggled to identify and/or understand the hazard that was threatening their area. Therefore, it is recommended that PWO graphics include information about the specific hazards that are threatening the risk area. For example, some recent SPC graphics have provided this information on the side of the graphic (Figure 11). This should provide more emphasis on the threats directly impacting individuals in these areas. Another possibility is to change the name or title of the PWO graphic. Throughout the interviews, individuals were unsure what encompassed "severe weather" and usually attributed various types of weather as "severe." Participants

described the Day 1 Convective Outlook as depicting excessive rain, snow, hurricanes, thunderstorms, and even extreme heat. Therefore, the SPC might consider changing the title of the graphic to “Severe Storm Outlook.” This would emphasize the threat for “storms,” and eliminate the possibility of other weather hazards. A combination of a name change and the addition of the threat information would provide more clarity to the end users.

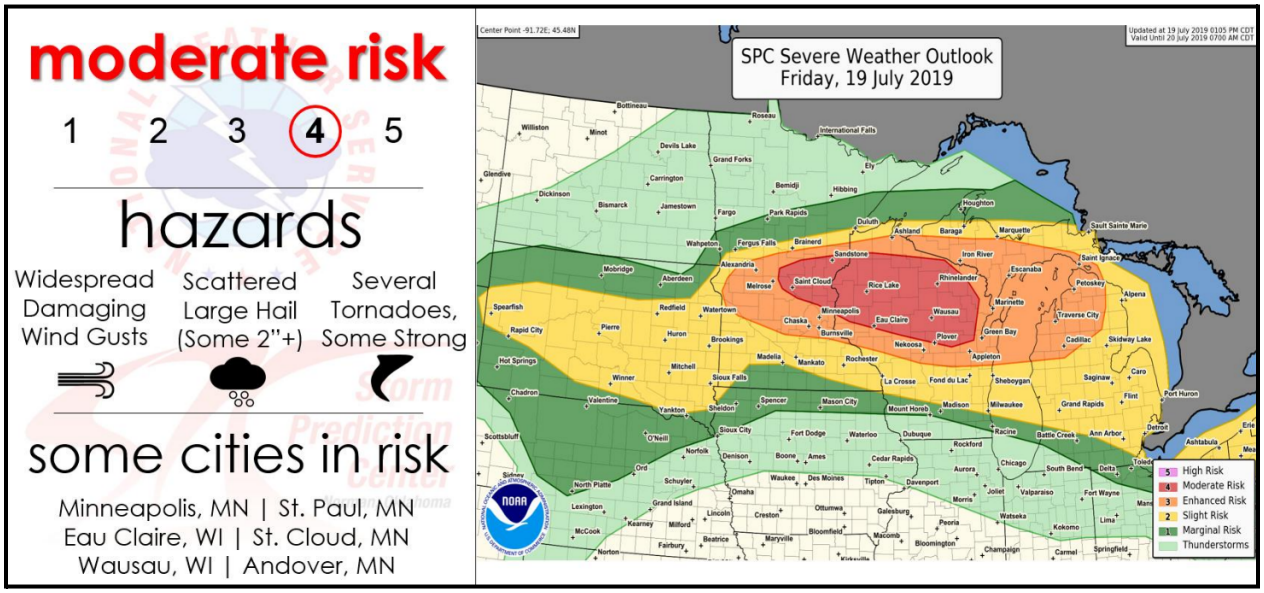
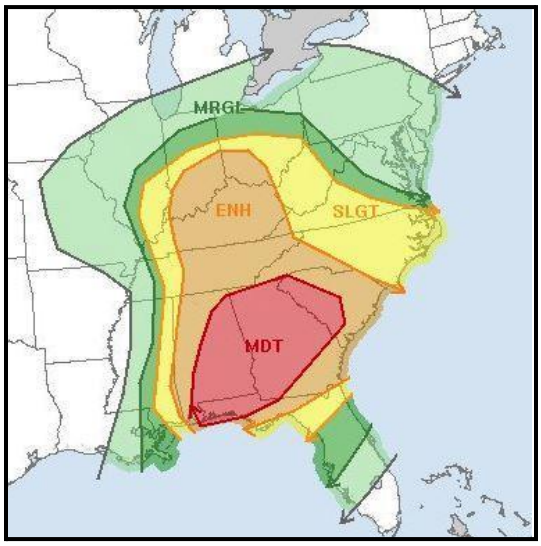


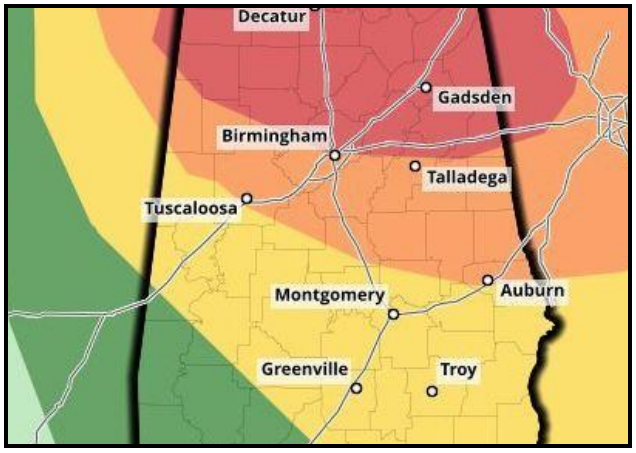
Figure 11: New SPC Day 1 Convective Outlooks that emphasize hazard-related information.

- **Recommendation: Remove the arrows on the SPC National Categorical Outlook graphic.** Although it was recommended that the SPC National Categorical Outlook graphic be used less frequently when communicating with members of the public, it is important to note that individuals overly emphasized the arrows in the SPC National graphic and frequently used them to infer storm direction. Therefore, to reduce confusion, it is recommended that these arrows be removed from the risk boundaries (see Figure 3).



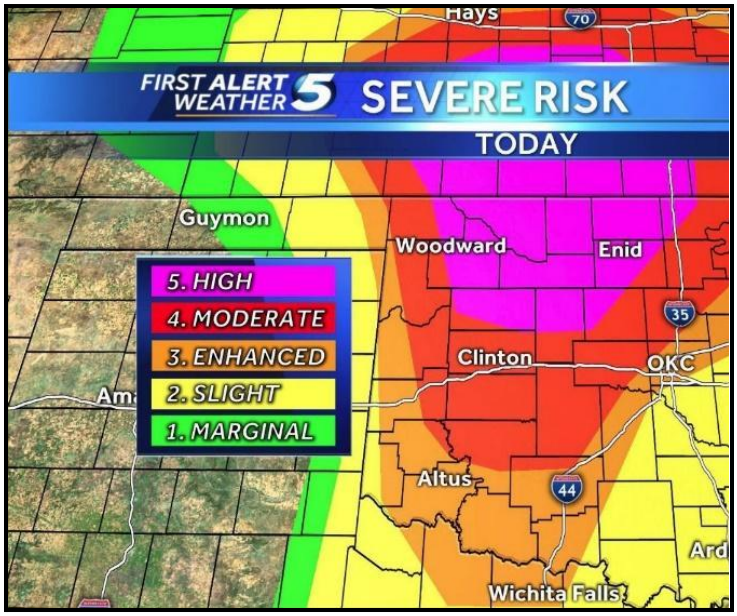
See Figure 3. Arrows on the National Convective Outlook Graphic.

- **Recommendation: The risk boundaries should be completely removed, or made less prominent on the PWO graphic.** With the risk boundaries creating many different interpretations of risk and uncertainty in Study 1, it is recommended that they be completely removed or made less prominent on the PWO graphic. By emphasizing the boundaries, the participants focused on them and attempted to understand their meaning. Therefore, making them less prominent on the PWO graphic, should lead to more individuals focusing on the categorical information and make transitioning between categories more fluid. To bolster the generalizability of these findings, further social science research is needed to explore the implications of removing the risk boundaries from the PWO graphic (see Figure 4)..



See Figure 4: Tuscaloosa is on a risk boundary between a 'Slight' and 'Enhanced' risk.

- **Recommendation: The legend on the PWO graphic should be more strategically placed and graphically modified to improve its usability.** In viewing a variety of Convective Outlook graphics, participants in Study 1 were able to examine various legends. This provided the opportunity for individuals to offer graphical preferences and recommendations for the PWO graphic. In particular, individuals thought the legend should be strategically placed so that it was closer to the risk areas. This would draw the eye more to the risk areas and the legend at the same time. Sometimes strategically placing a legend can interfere with the usability of the graphic, especially if it covers up a city, town, or region. Therefore, legends should be strategically placed when possible. Other recommendations for the legend include (1) making it vertical, (2) putting the entire risk category word in a long-colored box (see Figure 5), and (3) adding the word "Risk" to the end of each risk category word. This offers more context clues in describing the risk categories.



See Figure 5. Broadcaster graphic preferred by many because they were drawn to the vertical legend.

Note: Newer SPC Day 1 Convective Outlooks have started making the legend vertical and adding the word “Risk” to the end of each risk category (see Figure 9). This practice should be repeated when possible.

Conclusion #5: Social and physical scientists working on similar NWS products, services, policies, or processes should be connected when possible. In addition to the results of these three studies offering relevant conclusions and recommendations to the NWS and Storm Prediction Center, it is also important to highlight the collaborations that were fostered by NWS and OAR’s Weather Program Office that informed our research process. During 2020, the NWS, SPC, and Weather Program Office helped connected researchers that had been funded by NOAA to improve, alter, and/or change the SPC’s Convective Outlook graphic. This resulted in collaborations between researchers at different institutions, the sharing of knowledge across research projects, and also knowledge sharing efforts for the broader weather enterprise. For example, our collaboration with researchers at the University of Oklahoma resulted in an entire American Meteorological Society session on social science findings associated with the SPC’s Convective Outlook graphic. Not only that, but our fruitful collaboration also resulted in a review of all social science research that had been conducted on the Convective Outlook graphic (see Krocak et al. 2021). Therefore, by connecting with researchers that had similar goals and research questions, we were able to brainstorm together, build complementary research projects, and work together to improve NWS products, services, policies, and processes.

References Cited

American Meteorological Society, 2018: (In)consistency in a social media world: Communication reflections of the 2017 hurricane season. *Matt Parker Communication Workshop*, Austin, TX, Amer. Meteor. Soc. Retrieved from: https://annual.ametsoc.org/2018/index.cfm/programs/short-courses-workshops/workshop-the-matt-parker-communication-workshop-on-consistent-weather-messaging/

Backhaus, A. L., 2004: An examination of the relationship between college students' conceptual level development and their perceptions of advertisements aimed at alcohol intervention (unpublished thesis). Emporia State University, Emporia, Kansas, United States.

Brewer, N. T., N.D., Weinstein, C. L. Cuite, and J. E. Herrington, 2004: Risk perception and their relation to risk behavior. *Annals of Behavioral Medicine,* **27**(2), 125-130.

Demuth, J. L., J. K. Lazo, and R. E. Morss, 2012: Assessing and improving the NWS point-and-click webpage forecast information. *NCAR Technical Note.* Retrieved from https://opensky.ucar.edu/islandora/object/technotes%3A504

Elstad, E., D. M. Carpenter, R. F. Devellis, and S. J. Blalock, 2012: Patient decision making in the face of conflicting medication information. *Int J Qualitative Stud Health Well-being,* **7**(1). doi: https://doi.org/10.3402/qhw.v7i0.18523

Ernst, S., J. Ripberger, M. Krocak, H. Jenkins-Smith, and C. Silva, 2021: Colorful language: Investigating public interpretation of the Storm Prediction Center's Convective Outlook. *Weather and Forecasting*, **36**(5), 1785-1797.

Everitt, B.S., 1993: *Clustering analysis.* John Wiley and Sons. New York.

Grams, J., B., Bunting, & S., Weiss, 2014: SPC Convective Outlooks. Retrieved from: www.spc.noaa.gov/misc/SPC_probotlk_info.html

Hsieh, H. F., and S. E. Shannon, 2005: Three approaches to qualitative content analysis. *Qual Health Res,* **15**(9), pg. 1277-1288.

Kain, D. J., and C. F. Smith, 2010. Risk Perceptions and Emergency Communication Effectiveness in Coastal Zones: Preliminary Findings on Interpretations of Weather Related Messages and Maps. Report provided to the Director of the US National Hurricane Center.

Klockow, K. E., and S. A. Jasko, 2016: Building a consistent message: Establishing ground rules for the weather enterprise. 11th Symposium on Societal Applications, New Orleans, LA, Amer. Meteor. Soc., PD2.1, Retrieved from: https://ams.confex.com/ams/96Annual/webprogram/Session39835.html

Krocak, M., S. Ernst, J. Ripberger, C. Williams, J. Trujillo-Falcon, B. Gallo, and P. Marsh, 2021: The National Weather Service Storm Prediction Center's Convective Outlook: Conclusions from Past Research and Recommendations for Future Development. Retrieved from: https://www.spc.noaa.gov/publications/krocak/otlk-res.pdf

MacLeod, A., and S. Pietravalle, 2017: Communicating risk: Variability of interpreting qualitative terms. *EPPO Bulletin*, **47**(1), 57-68.

Mayhorn, C. B., M. S. Wogalter, J. L. Bell, and E. F. Shaven, 2004: What does code red mean?. *Ergonomics in Design,* **12**(4), pg. 12-14.

National Oceanic and Atmospheric Administration, 2016: Interpretation and use of the SPC Day 1 Outlook and recommendations for Increasing temporal and spatial resolution. Report provided to Storm Prediction Center.

National Weather Association, 2017a: Should weather messaging wear uniforms? The role of consistency versus uniformity of visuals and messaging within the weather community. *Wednesday Webinar Series,* Retrieved from: https://www.youtube.com/watch?v=IcnXfL92orE

National Weather Association, 2017b: Inconsistency and weather communication – The good, the bad, and the ugly. *Broadcasters Workshop,* Retrieved from: https://nwas.org/annual-meeting-events/past-meetings/2017-annual-meeting-agenda/

Psytoolkit, 2019: Wisconsin card sorting task (WCST). Retrieved from: https://www.psytoolkit.org/experiment-library/wcst.html

Sherman-Morris, K., 2005: Tornadoes, television, and trust—A closer look at the influence of the local weathercaster during severe weather. *Global Environmental Change Part B: Environmental Hazards,* **6**(4), 201-210.

Silva, C., J. Ripberger, H., Jenkins-Smith, and M. Krocak, 2020: Continuing the baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Data and reference report retrieved from Harvard Dataverse: https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/MLCJEW/ZSXOQU&version=1.1; doi: https://doi.org/10.7910/DVN/MLCJEW/ZSXOQU.

U.S. Census Bureau, 2019: *Resident population estimates by age, sex, race, and hispanic origin. Retrieved from: https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates/2020-evaluation-estimates/2010s-county-detail.html*

Williams, C. A., D. M. Carpenter, P. K. J. Han, and S. Joslyn, 2019: Identifying and assessing the effects of forecast and warning communication (in)consistency on different recipients. *14th Symposium on Societal Applications: Policy, Research, and Practice,* Retrieved from: https://ams.confex.com/ams/2019Annual/meetingapp.cgi/Paper/350959?fbclid=IwAR0MJMOChAQxhR5ML3zqhtszSW-qIKQOoN_sfJvbR4FMVv-MVis3UPvCQmg

Williams, C. A., and G. M. Eosco, 2021: Is a consistent message achievable? Defining "message consistency" for weather enterprise researchers and practitioners. *Bulletin of the American Meteorological Society,* **102**(2), E279-E295

Witte, K., 1992: Putting the fear back into fear appeals: The extended parallel process model. *Communications Monographs,* **59**(4), 329-349.