University at Albany, State University of New York

Scholars Archive

Emergency Preparedness, Homeland Security, and Cybersecurity Faculty Scholarship

Emergency Preparedness, Homeland Security, and Cybersecurity

3-2023

Communicating about Extreme Heat: Results from Card Sorting and Think Aloud Interviews with Experts from Differing Domains

Jeannette Sutton

Nicholas Waugh

Savannah Olivas

The University at Albany community has made this article openly available.
Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/ehc_fac_scholar

Part of the Atmospheric Sciences Commons, Climate Commons, and the Emergency and Disaster Management Commons

Rights Statement

IN COPYRIGHT

License

Accepted for publication in Weather, Climate, and Society, March 2023.
https://doi.org/10.1175/WCAS-D-22-0108.1

Communicating about Extreme Heat: Results from Card Sorting and Think Aloud Interviews with Experts from Differing Domains

Jeannette Sutton,[a] Nicholas Waugh,[a] Savannah Olivas,[a]

[a] *College of Emergency Preparedness, Homeland Security, and Cybersecurity, University at Albany, State University of New York, Albany, New York*

Corresponding author: Jeannette Sutton, jsutton@albany.edu

Accepted for publication in Weather, Climate, and Society, March 2023.
https://doi.org/10.1175/WCAS-D-22-0108.1

ABSTRACT

Climate trends indicate that extreme heat events are becoming more common and more severe over time, requiring improved strategies to communicate heat risk and protective actions. However, there exists a disconnect in heat-related communication from experts, who commonly include heat related jargon (i.e., technical language), to decision makers and the general public. The use of jargon has been shown to reduce meaningful engagement with and understanding of messages written by experts. Translating technical language into comprehensible messages that encourage decision makers to take action has been identified as a priority to enable impact-based decision support. Knowing what concepts and terms are perceived as jargon, and why, is a first step to increasing communication effectiveness. With this in mind, we focus on the mental models about extreme heat among two groups of domain experts –those trained in atmospheric science and those trained in emergency management to identify how each group understands terms and concepts about extreme heat. We use a hybrid data collection method of open card sorting and think-aloud interviews to identify how participants conceptualize and categorize terms and concepts related to extreme heat. While we find few differences within the sorted categories, we learn that the processes leading to decisions about the importance of including, or not including, technical information differs by group. The results lead to recommendations and priorities for communicating about extreme heat.

SIGNIFICANCE STATEMENT

Effective communication between domain experts is a priority for informed decision making under extreme weather conditions. As severe heat events increase in frequency and severity, the ability to communicate about heat and its impacts in a clear manner will become vital to life safety. The use of jargon-filled technical, scientific language, can serve as a barrier to understanding and engagement, delaying decision making and action. By identifying how extreme heat terms and concepts are understood among domain experts, risk communicators can determine where to focus on the development of plain language messaging which improves decision support and decision making.

Accepted for publication in Weather, Climate, and Society, March 2023.
https://doi.org/10.1175/WCAS-D-22-0108.1

1. Introduction

Extreme heat continues to be a significant problem that is increasing over time (Climate.gov Staff 2012). Because heat waves are slow and silent, they are relatively invisible and receive less attention than sudden emergent and environmentally destructive events. Their victims are primarily the elderly, the poor, the isolated, and other vulnerable populations who generally lack access to resources to prepare for and withstand hot days and nights (EPA.gov 2021). Even disaster scholars fail to study them as frequently as other more highly visible disasters. However, in the U.S. more people die in heat waves than in all other high impact weather events (National Weather Service 2022).

Advances in measurement and prediction of extreme heat, the use of geospatial methods, and analytical tools allow researchers and practitioners to identify the characteristics of heat related vulnerabilities within urban environments (Wilhelmi and Hayden 2016). Researchers have found that early and accurate decision making about heat health can be facilitated by linking heat-related morbidity and mortality data with socio-demographic characteristics in urban and rural environments (Wilhelmi and Hayden 2016). This requires integration of sensing, measurement, and modeling to generate an actionable forecast. It also requires a commitment to clear and actionable communication.

Thus, attention to the ways that heat measurement and heat-health impacts are communicated is also increasing (Grundstein and Williams 2018). However, decision makers, such as emergency managers, need to understand the heat related information they are provided by forecasters because it is used to make life saving decisions on behalf of the public. This is, in fact, a priority of the National Weather Service (NWS) as they shift to a collaborative model that supports Impact-Based Decision Support (IDSS), and make hazard communications more understandable based on social science research (Uccellini and Ten Hoeve 2019). Indeed, there remains a dearth of research investigating the divide between experts in atmospheric and health sciences and experts in other domains, such as emergency management, when it comes to understanding relevant heat-related terms and concepts, and their scientific jargon (i.e., technical language). Therefore, we need an understanding of how different expert groups approach or conceptualize different terms about heat, heat measurement, and heat-health impacts.

Specifically, prior research on science communication has identified knowledge gaps between scientists (i.e., atmospheric science experts) and their audiences (Sivle and Aamodt

https://doi.org/10.1175/WCAS-D-22-0108.1

2019), whereby scientists routinely focus on the science of a hazard by identifying and defining dimensions, variables, and measurement that are domain-specific (Sharon and Baram-Tsabari 2014) which may not be meaningful to experts in other domains. Hazard communication from scientists in varying domains is often highly technical; that is, using professional language, or jargon, which has precise definitions that carry specific meanings for some domain experts and different meanings for others (Bullock et al. 2019). Although non-experts may be motivated to learn about concepts from other domains, in many cases, the terms are "empty words to be filled with meaning" (Sivle and Aamodt 2019, p. 437). Furthermore, the use of jargon makes message processing more difficult for individuals who are unfamiliar with the technical language (Shulman and Bullock 2020; Bullock et al. 2019). Processing becomes particularly important when considering the audiences, domain experts as well as members of the public, making use of scientific content under high impact weather conditions to make decisions about how to mitigate the impacts of extreme heat. Therefore, a better understanding of how atmospheric scientists and emergency managers conceptualize, understand, and organize terms used to communicate about extreme heat may help improve risk communicated between domain experts to support better decision making (Uccellini and Ten Hoeve 2019).

In this paper, we focus on the mental models of communicating extreme heat among two groups of people with different domain expertise – those trained in atmospheric science (ATM) and those trained in emergency management (EM). We use card sorting and think aloud interviews with 37 participants to learn how they define, organize, and prioritize words and concepts that are related to and used to communicate about extreme heat and its impacts. We find that while there is much agreement between the two participant groups about which heat-related words and terms conceptually belong together, differences arise around what terms should be prioritized for decision making and why, including concepts that were missing from the initial list of heat terms. By examining the ways that ATM and EM domain experts interpret and prioritize extreme heat concepts, risk communicators can be prepared to better communicate about extreme heat, its impacts, and the actions that individuals can take to protect themselves. The results from this research suggest that simply making heat information publicly available without also paying attention to the technical language used to communicate it, is not sufficient for effectively communicating about heat risk.

2. Background Literature

a. Impact-based decision making

Following the April 2011 tornado outbreak and the subsequent May 2011 Joplin, Missouri tornado outbreak, the NWS determined that the "meteorological community was not fully connecting forecasts and warnings to decisions made by the wide range of decision-makers, especially public safety officials" (Uccellini and ten Hoeve 2019, p.1927). Recognizing this gap, the NWS determined that the way people receive, understand, and act on information, will stem from the effective delivery of decision support services (Lazo et al. 2020). This resulted in a formalized approach to impact based decision support services, designed to improve communication and to connect information with decision-making (Lazo et al. 2020). Importantly, the NWS recognized that forecasts and warnings must be understood by organizational partners, but these products have been based almost entirely on physical science principles and grounded in technical language resulting in barriers to effective communication and decision making (Uccellini and ten Hoeve 2019). The NOAA Weather-Ready Nation Strategic Plan (2019-2022), emphasizes the importance of clear communication between partners and decision makers, suggesting that the identification of communication barriers will help experts to deliver actionable information (National Weather Service 2019).

b. Communicating technical information and the use of jargon

Understanding technical language is an obstacle for many people (Wellington and Osborne 2001). Scientific concepts, such as meteorological terms that include measurements via indices (i.e. heat index or wet-bulb globe temperature) that are used to describe a forecast may be unfamiliar to many people. These terms represent jargon, which has specific meaning in an expert context that allows for precise communication among those with similar domain knowledge but becomes a barrier to communication for those who lack that knowledge (Silverman et al. 2016). Studies investigating the effects of jargon on metacognition have found that when information looks complicated, individuals are prone to ignore that information rather than meaningfully engaging with the material (Shulman and Bullock 2020). Even the presence of jargon, they argue, is "a cue that signals that the presented information will be effortful to process" leading to resistance and disengagement (Shulman and Bullock 2020, p. 2).

The association between jargon and effective communication has been identified in multiple domains including weather (Sivle and Aamodt 2019), law (Benson 1985), and medicine (Williams and Ogden 2004), demonstrating a communication divide between experts in differing domains as well as their publics. One recent scoping review examined interventions used during extreme heat conditions and identified several assumptions among domain experts about publics (Mayrhuber 2018). They found that risk communicators assume that informing people about heat and heat dangers will lead to behavioral change; at risk individuals will recognize their own vulnerability which will lead to concern and awareness; and the benefits of heat advice are commonly understood and taken seriously (Mayrhuber 2018). Each of these assumptions suggest that there is an expectation that those communicated *to* will understand the meaning and importance of the concepts communicated *by* domain experts.

Within the context of extreme heat for example, multiple indices have been developed to serve as thresholds for issuing heat products (i.e., heat watch, extreme heat advisory, and extreme heat warning). For example, ATMs use the Heat Index (which includes temperature and humidity), and the HeatRisk forecast product (which considers the severity and duration of unusual heat) as criteria for measuring potential impacts related to heat. This technical information is also frequently communicated in risk messages from NWS Weather Forecast Offices, using terms such as “heat index temperature” or “heat index values” with little, or no, explanation to the public about the effect of humidity on increasing dangerous health impacts (Sutton et al, 2021). The use of technical jargon that may be well-understood by those with scientific training or domain expertise but may lack the same meaning for those who have limited knowledge.

As the frequency and severity of extreme heat events increase, the need for impact-based decision support will likely increase among partners in the weather enterprise and emergency managers. However, there is limited research on how information about extreme heat and its effects are understood and communicated by these two groups. This leads to our primary research question: *how do experts from different domains define, describe, and categorize heat terms and concepts for extreme heat events?* To address this question, we utilize card sorting methods that can help to identify the ways such terms are understood.

3. Materials and Methods

Card sorting allows qualitative researchers to investigate how participants understand and organize concepts (Conrad and Tucker 2019). This technique involves participants sorting labeled index cards into categories while they describe how they relate to each other (a running commentary known as “thinking out loud”). Card sorting methods are frequently used within information and design sciences to inform, develop, and test website structures to ensure that they are aligned with end user needs (Rosenfeld et al. 2015). Card sorting methods have been used to identify the mental models of experts and non-experts (Chen et al. 2020; Asgharpour et al. 2007), optimize information organization for medical processes (Reese et al. 2018), and aid in the design of webpages (Faiks and Hyland 2000; Wentzel et al. 2016).

Card sorting methods include closed, open, and hybrid approaches. Closed card sorting requires participants to organize cards into pre-defined categories that are already labeled. In an open card sorting approach, participants create and label their own groups of cards (Spencer and Garrett 2009). In both cases, sorted cards are used to generate structured data that can be organized in a similarity matrix, visually illustrating the relationships between concepts (Fincher and Tenenberg 2005). A hybrid variation of card sorting includes the use of participant-generated cards in addition to those already labeled. This allows researcher insight into the participants' language and domain (Conrad and Tucker 2019). Adding a ranking order task to card sorting invites participants to prioritize concepts and offers insight into their mental models (Conrad and Tucker 2019).

Qualitative think aloud interviews conducted during card sorting tasks provide insights into *what* the participant thinks about as they identify the terms and concepts on each card, *how* they are related to other cards, and *why* they make category decisions (Conrad and Tucker 2019). The running commentary provided by participants offer insight into the mental model they hold about the relationships between terms and their decision-making process about hierarchical ordering (Saunders 2015).

In this study, we used a hybrid variation of card sorting, where participants organized cards that were already labeled, eliminated cards that did not appear to fit, and added new cards that were considered missing. Participants also rank ordered the cards within each sorted pile to prioritize concepts. Each of these tasks, was accompanied by verbal descriptions of their thoughts as they sorted, organized, and structured the concepts.

By conducting card sorting and think aloud interviews in this manner, concepts about extreme heat, including meteorological factors, environmental features, populations at risk, heat impacts, and safety tips, become tangible. Participants visually display and audibly communicate their mental models, illuminating the way that they often group, sort and label tasks and content within their own heads (Rosenfeld et al. 2015) and allow researchers to observe commonalities and differences among participants in each domain.

a. Participants

Altogether, 37 participants were recruited in two waves via email announcements and social media advertisements on Facebook inviting them to contribute to a study on communicating extreme heat. The number of participants for card sorting research can range from 10-15 participants (open card sorting) to as many as 48 participants (fixed card sorting) (Conrad and Tucker 2019). However, studies have demonstrated that statistical correlations stabilize at 20-30 participants (Tullis and Wood 2004). Initial piloting testing and training was conducted with two professional ATMs. This was followed by the first wave of 15 participants representing two knowledge domains, who participated in face-to-face interviews. This wave consisted of 15 graduate and advanced undergraduate students majoring in atmospheric science (ATM) and emergency management (EM) at a large university in the Northeast. The second wave of participants participated via Zoom using kardSort (Balachandran n.d.), an online software for card sorting, and included 20 professional EMs who were recruited via advertisements posted to Facebook groups dedicated to emergency management. While professional experience differed among participants, the extent of formal education in their respective fields was similar. All participants were compensated for their participation in the activity via $20 Amazon gift card. The procedures were approved by and conducted in accordance with the standards of the university Institutional Review Board.

Table 1. Participant Characteristics

Characteristic	Atmospheric Scientists (n = 12)		Emergency Managers (n = 25)	
	n	%	*n*	%
Employment				
Students	10	83.3	5	20
Professionals	2	16.7	20	80
Sex				
Male	8	66.7	17	68
Female	4	33.3	7	28
Other/Decline	0	0	1	4
Race/ethnicity				
African American	1	8.3	0	0
American Indian/Alaskan Native	0	0	1	4
Hispanic, Latino	1	8.3	1	4
Middle Eastern	0	0	1	4
Southeast Asian	1	8.3	0	0
White	9	75.0	21	84
Age				

18-29 years	9	75.0	7	28
30-49 years	2	16.7	13	52
50-69 years	1	8.3	3	12
Number of years in profession				
0 years	5	41.7	3	12
1-5 years	4	33.3	9	36
6-10 years	2	16.7	5	20
11-15 years	0	0	1	4
15+ years	1	8.3	8	32

4. Card Sort Task

Terms and concepts used in sorting were obtained from webpages about extreme heat maintained by NWS Forecast Offices across the U.S. and included words or phrases used to describe extreme heat, its measurement, its effects on human and animal populations, vulnerable populations, and protective actions (NWS Green Bay n.d.; NWS Albuquerque n.d.). This task was followed by a review of webpages and glossaries to compile terms associated with measurement of "wet-bulb globe temperature" (American Meteorological Society 2019). From these reviews, we identified 44 terms (see Table 2).

Table 2. A list of terms, presented alphabetically, included in the card sorting exercise

Terms Used in Exercise			
Absorption	Active	Air conditioning	Air density
Air temperature	Animals/pets	Athletes	Children
Cloud cover	Dew point	Drink water	Dry heat
Elderly	Evaporation	Excessive heat warning	Excessive heat watch
Green spaces	Heat advisory	Heat cramps	Heat disorders
Heat exhaustion	Heat index	Heat stroke	Heat warnings
Human heat index	Humidity	Hyperthermia	Moisture content
Pre-existing conditions	Pressure	Relative humidity	Safety tips
Solar radiation	Sun angle	Sunburn	Sunstroke
Take breaks	THSW index	THW index	Urban heat island
Vulnerable populations	Wet-bulb-globe-temperature	Wind speed	Working outdoors

In each face-to-face interview session, participants were handed a stack of 44 index cards, each displaying one term or concept seen in Table 2. Additional blank cards were also provided for participants to add terms they considered to be missing or to create labels for the groupings they created. Participants were asked to sort cards into piles based on perceived similarities among the terms. They were told that their piles should include 2-8 cards each, and that they could create as many piles as necessary. Participants were assured there were no right or wrong answers. Participants were also instructed to put any terms they did not recognize, know the meaning of, had difficulty sorting, or chose not to include in any categories into a discard pile.

Throughout the process participants were also asked to "think aloud," or verbalize their thoughts about their decision making, how terms they were viewing fit together, and any other thoughts that came to mind as they completed the exercise. Throughout each interview, the facilitator prompted participants to talk out loud and took handwritten notes describing the process each participant used to create piles as well as to record relevant comments made by participants while sorting. The facilitator answered participants' questions about the card sorting process but declined to answer questions relating to definitions or terms included on the cards.

When all the cards were sorted, each participant was asked to create a label for each of their created piles. Participants were then asked to arrange the cards in each pile into a hierarchical order that they saw as best fitting. At the end of each interview, participants were asked a series of questions including what terms were confusing, should not be included, were missing, and could be present in multiple groups, as well as which groups were easiest and most difficult to create.

Virtual card sorting interviews were conducted in a similar manner as those that were done face-to-face. After logging into Zoom, the facilitator invited the individual participant to silently read and digitally signed a participant agreement, acknowledging their consent to be recorded. The facilitator verbally outlined how to use the kardSort program (software that allowed the participant to organize labeled digital cards by dragging and dropping into ordered piles, or to add cards by providing new labels to blank digital cards) and explained the activity in the same way as in face-to-face interviews. The facilitator interacted with the participants throughout the interview by encouraging them to think out loud and asked the same series of questions at the conclusion of card sorting as those who participated face-to-face. All interviews lasted between 30 and 45 minutes, were audiotaped, transcribed, and checked for accuracy.

a. Analysis – Card Sorting

All physical and virtual card sort data was entered into kardSort (Balachandran n.d.) for data processing. Participants' card sort data were exported to SynCaps software version 3, a free online cluster analysis software, for data analysis (Hudson 2018). Prior to analysis, data was processed and cleaned. In this process, new cards that were added by the participants (n = 99) were removed and saved for thematic analysis and named categories that were functionally identical were grouped together. No significant outlying or unusual responses emerged. Two participant groups were then created based on participant background: one for ATMs (*n* =12) and one for EMs (*n* = 25). All further analysis was performed using these two groups.

b. Pairs Maps and Similarity Matrices

Using SynCaps v3, two item-by-item pairs maps were generated, one for each participant group, and visualized as similarity matrices (Figures 1 and 2). Pairs maps show the frequency by which pairs of terms are grouped together. A similarity matrix displays the percentage of times (between 0 and 1.00) each possible pair of terms appeared together in the same pile across the sample of participants (Ocampo-Agudelo and Maya 2021). This analysis is visualized within the matrix using a variety of colored shapes, each representing the level of similarity. Pairings have strong similarity, represented by green circles within the matrix, with ranges between 75% - 100%; moderate similarity, represented by yellow triangles, with ranges between 50% - 74%; weak similarity, represented by red diamonds, with ranges from 25% - 49%; and no similarity, represented by no symbol, with less than 25% (Savage et al. 2019; Righi et al. 2013).

Cluster analysis was also performed using SynCaps v3 on the card sort data for each group of participants. Clusters, numbered from 1-8, are represented in the similarity matrices (Figures 1 and 2). Average maximum similarity, a measure identifying how similar overall responses were within each group, was also generated automatically within SynCaps v3. Clusters were visualized in dendrograms (Figure 3), offering another view of the level of association between each term determined by the participants from each background.

c. Analysis - Interviews

Interview transcripts and recordings were independently coded by a member of the research team and verified by the interview facilitator through a series of discussions (Creswell and Miller 2000). Initially, coding focused on terms or words that participants said they found particularly confusing or difficult to categorize and terms that had low levels of agreement in the card sorting analysis. These included the words "green spaces," "dry heat," and "wet-bulb globe temperature." As insights emerged about the definitions participants applied to the terms, attention was then turned to the processes that each participant used to organize the words, that is, what they said out loud about their card sorting and why specific decisions were being made.

5. Card Sorting Results

Study participants included twelve ATMs and twenty-five EMs (Table 1). ATM participants were primarily white (n = 9; 75%) and identified as male (n = 8; 66.6%); 10 (83.3%) were students enrolled in related coursework and two were full-time professionals. ATM participants were an average age of 26 and had 0-20 years of experience, ranging from internships to professional meteorology, with an average of 4.5 years. EMs were primarily white (n = 21; 84%) and identified as male (n = 17; 68%); five (20%) were students enrolled in related coursework and twenty (80%) were full-time professionals. EM participants were an average age of 33 and had 0-40 years of experience with an average of 10.5 years.

a. ATM Card Sort

The ATMs formed 8.08 ±3 piles with each pile containing an average of 5.28 cards. Based on the overall similarity matrix, the ATMs card sort data had an average maximum similarity 0.74 ±.079, indicating a moderate to strong similarity between all responses in this group. Cluster analysis yielded 8 clusters of terms (Figure 1), numbered 1-8, corresponding to the following categories: warning language, indices, air and sun, humidity, other, populations, safety tips, and health. Figure 1 displays the similarity matrix from the ATMs card sort data and indicates the degree of similarity (0 to 1.00, expressed as percentages) between each pair of 44 terms, organized by cluster. For example, 100% of ATMs agreed that children and athletes should be categorized together (strong similarity), 67% agreed that dew point and relative humidity should be categorized together (moderate similarity), 25% agreed that evaporation and dry heat should be categorized together (weak similarity), and 4% agreed that sun angle and heat index should be categorized together (no similarity).

Accepted for publication in Weather, Climate, and Society, March 2023.
https://doi.org/10.1175/WCAS-D-22-0108.1

b. EM Card Sort

The EMs formed 7.4 ±3 piles with each pile containing an average of 5.73 cards. Based on the overall similarity matrix, the EMs card sort data had an average maximum similarity that was calculated to be 0.77 ±.065, indicating a strong similarity between all responses in this group. Cluster analysis yielded 8 clusters of terms (Figure 2), numbered 1-8, corresponding to the following categories: safety tips, built environment, indices, humidity, air and sun, warnings, vulnerable populations, and health. Figure 2 displays the similarity matrix from the EMs card sort data and indicates the degree of similarity (0 to 1.00, expressed as percentages) between each pair of 44 terms, organized by cluster. For example, 100% of EMs agreed that children and athletes should be categorized together (strong similarity), 64% agreed that dew point and relative humidity should be categorized together (moderate similarity), 40% agreed that evaporation and dry heat should be categorized together (weak similarity), and 0% agreed that sun angle and heat index should be categorized together (no similarity).

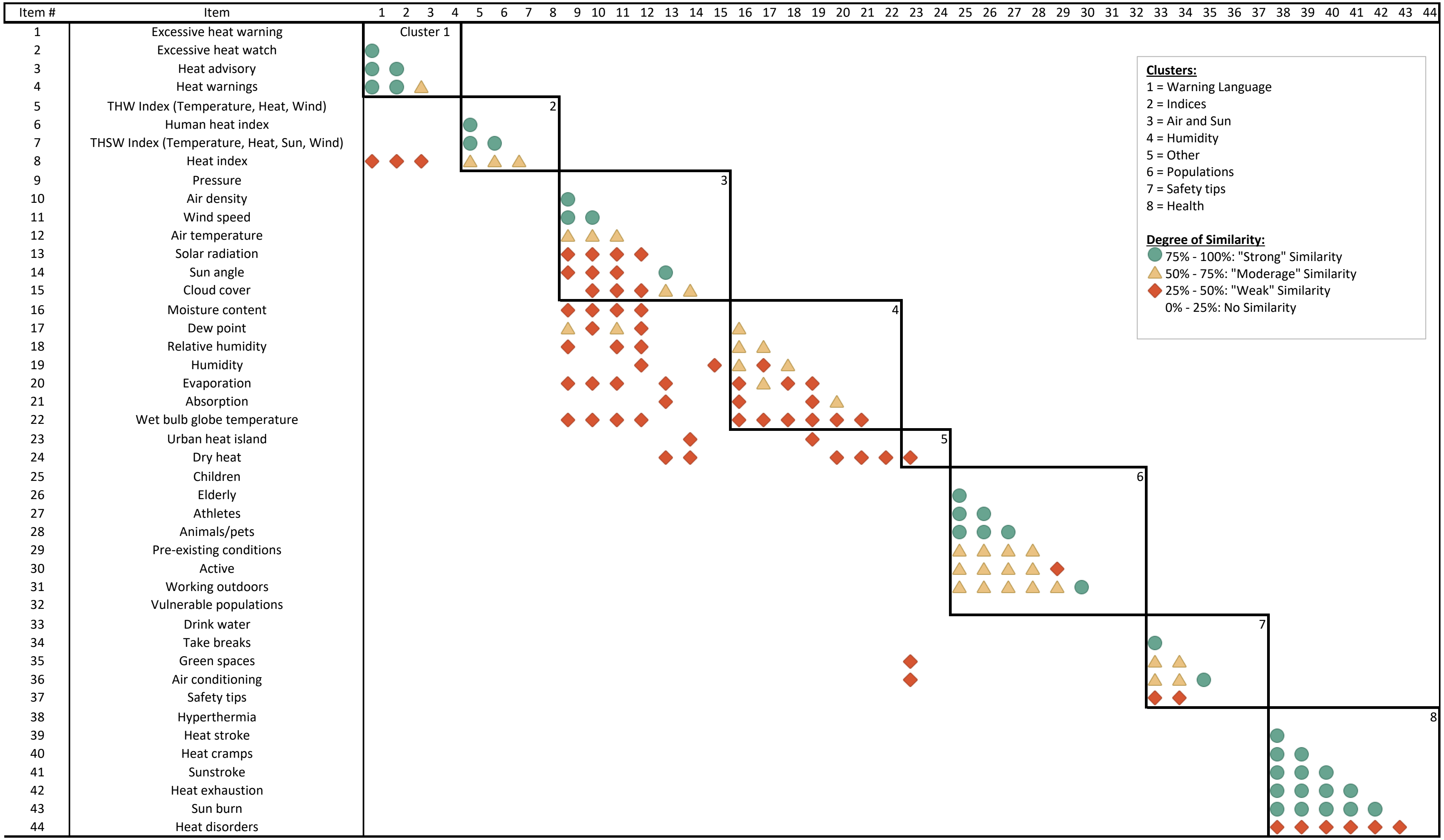

Figure 1. Similarity matrix of ATMs card sort of 44 items into piles.

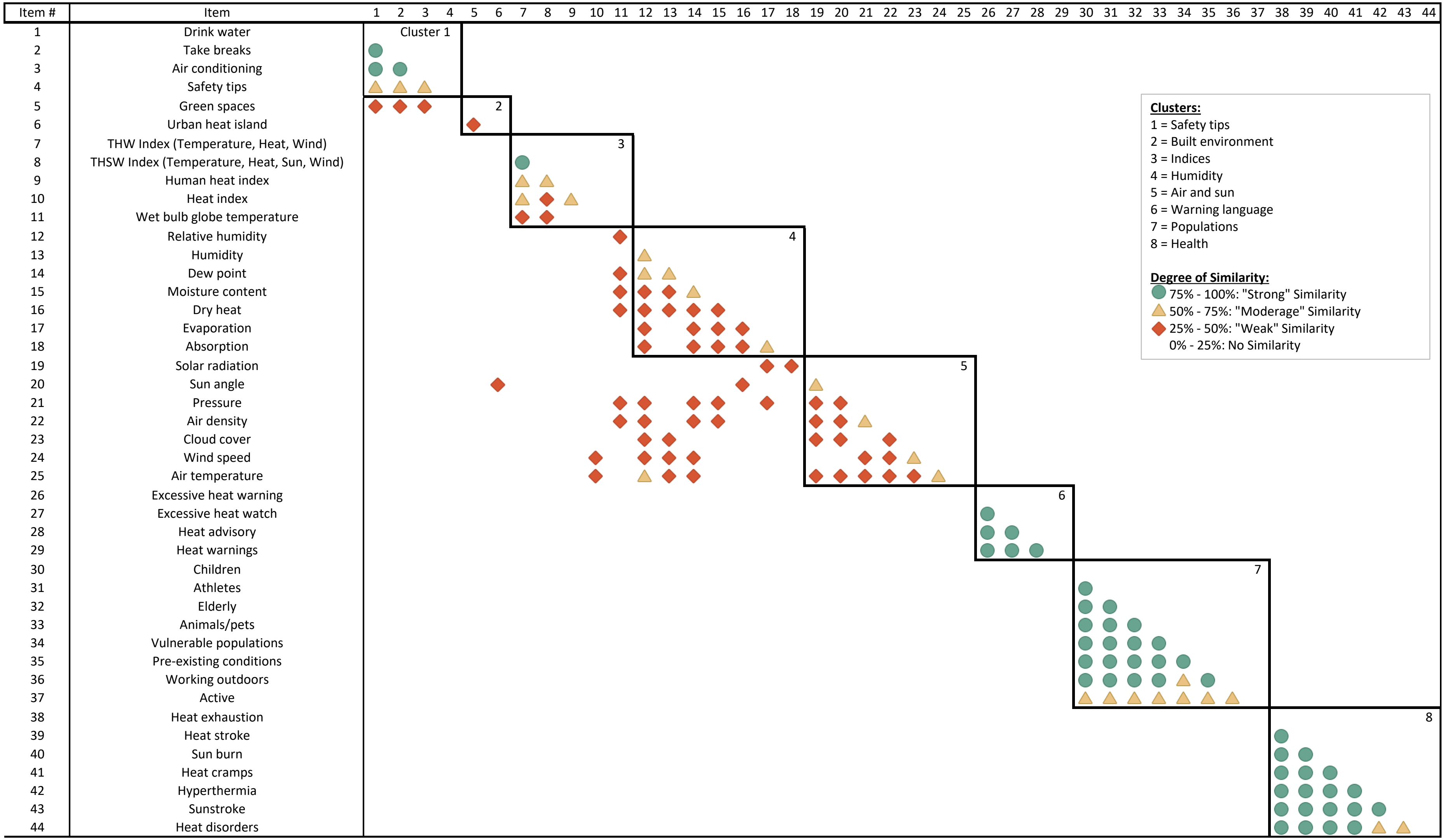

Figure 2. Similarity matrix of EMs card sort of 44 items into piles

c. Comparing Groups

When visually inspecting the results of both participant groups side by side via the dendrogram (Figure 3), we see agreement qualitatively. Cluster analysis revealed that both groups created 8 different clusters, with most of the clusters being almost identical. For example, clusters 5, 6, 7 and 8 created by EMs match clusters 3, 1, 6, and 8 created by ATMs, respectively, in terms of content. Overall, the placement of 41 of 44 items (93.18%) was identical among both groups.

Only 3 of the 44 original terms were placed within different clusters by the two groups. While the EMs grouped "green spaces" and "urban heat island" within cluster 2 (built environment), ATMs placed the term "green spaces" within cluster 7, among safety terms such as "drink water" and "take breaks." The placement of "wet-bulb globe temperature" differed as well, with EMs placing this term among the indices within cluster 3 while ATMs placed it with along with other weather variables such as "dew point," "moisture content," and "humidity" within cluster 4. Finally, EMs placed "dry heat" in cluster 4 with other weather terms such as "dew point and absorption," while ATMs placed "dry heat" in cluster 5 along with "urban heat island."

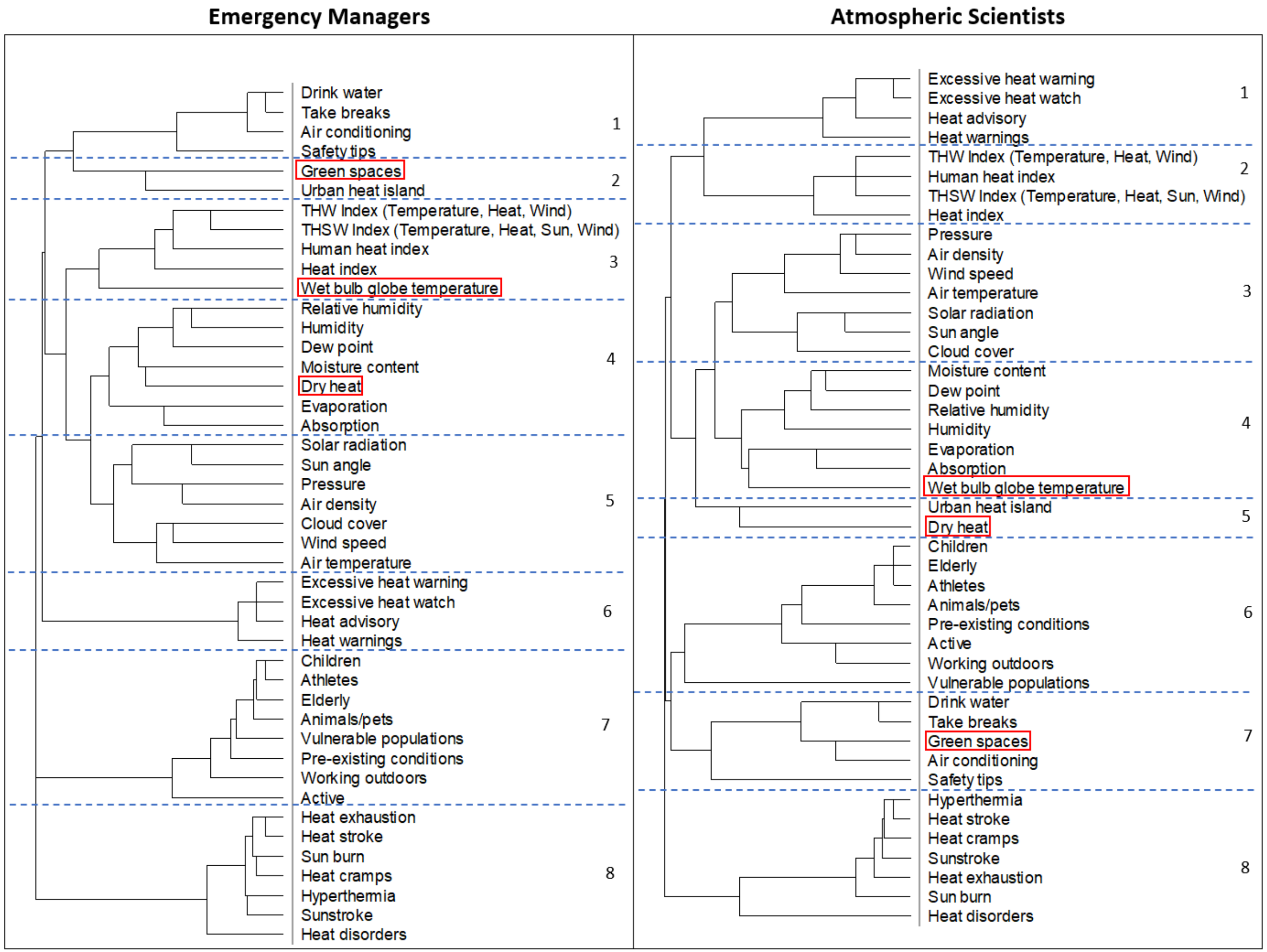

Figure 3. Side-by-side dendrograms of EM and ATM card sort with highlighted differences

Accepted for publication in Weather, Climate, and Society, March 2023.
https://doi.org/10.1175/WCAS-D-22-0108.1

d. Added Cards

During the card sort, participants were also invited to add terms and concepts that they thought were missing from the original set of cards. In total, participants provided 97 new terms; 31 were added during the face-to-face interviews; 66 were added during the virtual interviews (see Table 3). The majority of new terms (74%; *n* = 72) were added by EMs. Four health-related impact terms were added including excessive sweating, sun poisoning, dehydration, and death; 38 safety tips and/or actions were added including shade, fans, and cooling centers. In addition, 15 variations of vulnerable populations were added including unhoused populations, persons living without air conditioning, and persons with pre-existing health conditions. Participants also added 4 weather-related terms; "real feel" and "feels like" temperature, UV index, and wind direction.

In contrast, ATMs provided 22 new terms and concepts (22%). These terms included 7 safety tips and/or actions, such as stay indoors and cooling centers; and 3 variations of vulnerable populations including persons living without air conditioning, outdoor workers, and those who speak English as a second language. They also proposed 2 weather-related terms: the misery index and the summer glory index. A sub-set of words representing the terms added by the two groups are presented in Table 3.

Table 3. Terms added during card sorting exercise (* terms were added by both groups).

Terms Added During Exercise			
Emergency Managers		**Atmospheric Scientists**	
Excessing sweating	Real feel temperature	Find Shade*	Sports drink (electrolytes)
Sun poisoning	Feels like temperature	Emergency Information	Seeking relief
Dehydration*	UV index	Cool Compress	Stay indoors
Death	Wind direction	Check for local heat warnings	Misery index
Shade*	Removal of clothing	General humidity	Summer glory index
Fans	Sharing (Public Outreach)	911 emergency information	Dehydration*
Cooling Centers*	Sunny day floods	Signs and Symptoms	English as a second language communities
Unhoused populations	Sunscreen	Those without air conditioning*	Cooling centers*

Persons without AC*	Pop-up misting stations	Reducing time outdoors	Splash pads*
Persons with pre-existing health conditions	Splash pads/pools*	Cooling off at beach/pool	Sea breeze

There was largely agreement on the placement of terms between the two participant groups and across the clusters. However, the words and terms participants said were confusing or difficult to categorize and the processes used to create groupings differed. To learn more about this, we turn to the results of the think-aloud interviews.

6. Think-aloud Interviews

Interviews revealed insights into the definitions that individuals within each group applied to the terms/concepts. Definitions illuminate *what* each term means to the participant. Interviews also offer insights into the processes individuals used as they organized the terms/concepts. The process shows *how* cards were sorted and illuminate *why* specific decisions were made to group concepts together.

a. Terms/Concepts

There was general agreement on the definition and categorization of most terms. Across the 44 cards, we find that word groupings differed for only three concepts: green spaces, wet-bulb globe temperature, and dry heat. As seen below, even within the two participant groups, there were differences in interpretation.

1) GREEN SPACES AND URBAN HEAT ISLAND

The similarity matrix and dendrogram above illustrate how ATMs grouped "green spaces" with the cluster on safety tips, while EMs placed it within the built environment cluster with urban heat islands. With few exceptions, both participant groups generally defined "green spaces" as a *location* that mitigates heat. 75% of ATMs (n = 9) and 56% of EMs (n = 14) *verbally* described an association between green spaces and cooling effects or heat deterrence. In some cases, "green spaces" was defined as a location that affects local temperature conditions, in others, it was described in relation to places for people to go to, serving as a call to action. For example, one ATM participant explained grouping green spaces with "Safe spaces, because it's, like, *where's there's shade*... green spaces there's trees and stuff. More shade." In contrast, another ATM participant described grouping green spaces with safety tips, saying "If it's warm out, people can *go to greener areas* where there's probably more shade and cooler temperatures." The multiple paths to interpret green spaces was summed up by one interviewee who said:

> *I wasn't 100% sure what was meant by green spaces. Whether it was a suggestion to go into green spaces, as a safety tip to escape the heat and the sun, or if it was a suggestion to create more green spaces. I interpreted it as a suggestion to go into green spaces as protect yourself from the elements, but I could see how someone wouldn't necessarily interpret it the same way and maybe be confused of that.*

Several EM participants linked green spaces with "urban heat island," identifying the relationship between the two concepts; where one generates heat and intensifies its effects, the other absorbs or reduces heat, explaining that "green spaces can affect your urban heat islands." For example, one EM said:

> *All right, so green spaces... put that as heat reduction. And the reason I say that is because when we have green spaces, it's not concrete, it's elements of nature that absorbs heat, cools, refreshes, reduces the impact of concrete and buildings and that kind of stuff.*

Another EM explained that infrastructure and urban design affects local safety and resilience, saying:

> *Green spaces is kind of, you know, that's a more modern way for looking at safety tips. We see in urban planning that they're moving to the notion that having parks will reduce the urban heat island effect. So, I can kind of interchange that one with the safety tip or the weather message adjectives, because having more green spaces is gonna lower the overall temperature in that area.*

2) WET-BULB GLOBE TEMPERATURE

The similarity matrix and dendrogram above illustrate how ATMs grouped "wet-bulb globe temperature" with humidity terms, while EMs placed it with indices measuring variables related to extreme heat. In the interviews, we find that several participants pointed to the inclusion of the word "wet," as an indicator of moisture, while many made guesses or simply stated "I have no idea what that means." Importantly, few participants could define wet-bulb globe temperature, and several argued that other ATM or EM professionals, as well as members of the public, would be unfamiliar with the term. For example, one ATM participant said

It's not that I don't think these terms should be included [in a website about extreme heat]. I just wonder if some of the variables like air density and wet-bulb globe temperature [should be]. I just don't know if that would make sense to a lot of people unless there were definitions and explanations attached to these terms.

An EM professional echoed this sentiment saying:

Wet-bulb globe temperature, I've never heard of that term, or I probably have but it's not something I use in my day-to-day and I don't think... definitely it would be over the heads of people trying to understand a heat advisory or extreme heat language.

Some EMs referenced wet-bulb globe temperature in relation to wildland firefighting, fire weather, and heat stress related injuries, suggesting that it is a measurement applicable, and available, only in certain conditions. For example, one EM said:

And although I know what wet-bulb globe temperature is, and certainly we use that a lot in wildland firefighting, you know, to the general public or to general emergency management, most don't have the capabilities being able to do that [measure or understand] easily.

3) DRY HEAT

The similarity matrix and dendrogram show that ATMs grouped dry heat with urban heat island (a cluster labeled "other"); while most EMs (n = 18) grouped it with terms associated with humidity and moisture. Both groups verbally defined dry heat as heat with little or no moisture associated with it. However, several ATMs struggled to link dry heat to meteorological components because it lacked a specific measurement. For example, one ATM said:

Dry heat. I guess it's kind of variable but it feels like something that's more socially descriptive and, like, people can be impacted differently. It feels like a combination of, like, relative humidity and temperature but it feels more like a social description.

This idea of dry heat as a "social description" was echoed by one EM who drew from their personal experience of heat in a location where the humidity is very low, Colorado, or in Florida, where humidity is, by comparison, very high, explaining:

I have been out to Denver and then experienced the dry heat and did not realize that the temperature was as hot as it was because I'm used to Florida and Florida humidity and

> *sweating, and I wasn't sweating at all [in Denver]. And you can become dehydrated very quickly there.*

Another EM located in the South said:

> *Dry heat. I'm in the South. I'm totally not familiar with that at all. We always have humidity here. So, I really don't know, like, how much of a contribution that really is to the difference between a heat related illness or not.*

The term “dry heat,’ while intuitive to some, appears to be more relevant to persons who have prior experience with the phenomena. This suggests that interpretation of social descriptors will be dependent upon knowledge of local and regional conditions.

b. Processes used to sort and organize cards

The card sorting process illustrates how participants made sense of the terms, placing them in groups according to relationships within systems, such as cause and effect, or types, such as types of people, impacts, or actions. In some cases, participants made decisions that some terms did not belong at all – making judgments about whether words were perceived to be accessible, relevant, or having the most impact. We turn to a description of these processes next.

1) CONCEPT ACCESSIBILITY

Accessibility was described by both groups as words that are readable (at a third-grade level) or knowable (by the general public). For example, dry heat was explained as being knowable by one EM who said it is “pretty accessible to most people. Anyone can know what dry heat is, unlike human heat index and moisture content,” suggesting that knowable concepts can also be explained by drawing from experience in contrast with learned knowledge. Readability also made concepts accessible. For example, one ATM explained that they ordered groups of terms by readability where “the terms at the top are what would be most accessible to general populations.” One EM similarly summed up their concern about accessibility stating:

> *And this is something which is important, you know, the literacy we always are told to write for a third-grade reading level in emergency management when it comes to educating, so I wouldn't use that word I would just stick to heat exhaustion and heat stroke.*

2) CONCEPT RELEVANCE

Relevance was also identified as a primary motivator for including terms and concepts or sorting them into groups. Many participants organized clusters based upon a term's perceived relevance to their local context rather than the broader community. For example, one EM explained why they placed urban heat island at the bottom of a pile, saying "it sounds a bit narrow to people that actually live in urban areas… I question how much bearing it would have since so much of [my community] is mostly suburbs."

This was echoed by one EM who explained that they selected terms and concepts based upon what they thought a member of the public would want to see, asking "what would actually be most relevant to their life for their own planning?" in contrast with "objective measurements of the atmosphere."

Relevance was also highlighted by several ATMs, particularly regarding the inclusion of wet-bulb globe temperature. A lack of familiarity with its measurement and the perceived difficulty of explaining its usefulness led several to suggest that it had limited relevance to the public, and to themselves, in their understanding about extreme heat or decision making.

3) DETERMINING HAZARD IMPACT

Finally, in some categories, participants identified an objective hierarchy of impact and organized terms so that they were ordered from highest to lowest severity. This ordering occurred primarily in the warning cluster and the impacts cluster. For example, one EM explained:

> *As far as injuries and illness, that went from most impactful to least impactful on the individual. Moving down in warnings is kind of the same way: most impactful, to least impactful [ordered as] excessive heat warning, excessive heat watch, and heat advisory.*

One ATM offered a similar explanation, but ordered terms based upon the severity of the language. They said:

> *I'm going to put Excessive Heat Warning first, because that is something that I think would get my attention—the excessive really sells it, followed by excessive heat watch for the same reasons. I think that it communicates to people that … it is a dangerously hot day.*

4) CHOOSING TO DISCARD

While some participants had clear explanations about their organizing decisions, many remarked about their lack of expertise to define some words, to explain their groupings, or to provide categorical labels. This was primarily expressed in response to the summative question asked by the interview facilitator "what was the most difficult group for you to put together?" While some ATMs struggled to define specific terms like wet-bulb globe temperature or green spaces, a lack of weather expertise was clearly articulated by several EM participants as creating barriers. For example, one person stated: "As someone who's not a scientist, I've got an inkling of what many of these entail, but some of them might seem very esoteric to me." Another said, "I don't know a lot about meteorology and atmospheric science, so it's kind of hard to sort these out." And a third expressed "I'm not a weather person. So, pressure, air density, maybe these things are more important than I realized," raising questions about whether the absence of knowledge affected their ability to correctly determine the relevance of some concepts.

During the exercise, some participants chose to discard terms or concepts that they did not understand or placed cards in groups that were "difficult to label" (such as "related information") if the participant thought that the term held significance in a way they did not understand. For example, one EM said:

Next, I have green spaces, urban heat island, and wet-bulb globe temperature. So I don't really know what any of these are. I'm just gonna put related information as the category because I don't know if it's directly involved with these types of categories. I'm going to put it as like information that would be given but isn't necessarily as important.

Another EM said similarly:

Okay, so we have temperature, heat, sun, wind index, temperature, heat, wind. These mean nothing to me. I have no idea what wet-bulb globe temperature is. Air density. I don't know how that I can I know what those words mean. But I don't know what it means. I know what sun angle means. But I don't know how it's relevant. I don't really understand dewpoint.

Importantly, most of the discarded terms were confined to EM participants; ATMs were reluctant to remove any cards saying instead that "Overall, they have this importance. And I guess depending on how you interpret it, they can go anywhere, so they all should be included, no changes."

One ATM said:

There are some things that I don't think a large portion of the population would care to see. But at the same time, I appreciate the value of national services, providing all the available data for someone who wants to go see it.

In some cases, participants were familiar with a term, but chose to remove it because it was perceived to be inaccessible or irrelevant for public use. For example, one EM said:

I'm going to say this is my 'ditch-it' terminology. I don't think the public needs to know what wet-bulb globe temperature is. Unless you can describe it better…for messaging to normal people…not that NOAA folks aren't normal, but regular folk, I will say green spaces is a little academic.

In other cases, cards were perceived as being irrelevant due to the context – decision making under high stress conditions requires useful information; educational information, such as defining and explaining terms, can occur under conditions of preparedness. For example, one ATM said, "for the general public, these terms might be a little too in depth, because based on what I've seen, people just want to get information, get it fast and take action." Another explained, "But that's [information about urban heat islands] really educational for the public. That's not... I don't expect, again, you know, Miss Ethel that lives down the street, I don't expect her to want to read about urban heat islands [during high heat]."

7. Discussion

The primary aim of this study was to identify the mental models that two groups of domain experts, ATMs and EMs, have about extreme heat by examining the ways that they define, organize, and categorize heat terms and concepts. We used a qualitative hybrid card sorting approach, using pre-identified terms and blank cards that were sorted into unlabeled, or open, groups. Participants were asked to "think aloud" as they sorted cards and to organize them into hierarchies within each grouping.

Three terms stood out as being unfamiliar and difficult to categorize for both participant groups: green spaces, urban heat island, and wet-bulb globe temperature. Green spaces and urban heat island were recognized by both groups as being physical places that function to mitigate or exacerbate heat; however, their domain knowledge affected their mental models of how each place functions. ATMs described green spaces as a place where people can go to protect themselves during heat events and described urban heat island as an effect due to trapped heat. EMs discussed urban heat islands in terms of its occurrence resulting from planning decisions within the built environment. Many EMs also described who has access to green spaces and who is most affected by urban planning, linking extreme heat to the broader context of vulnerability, equity, and access to resources, which are topics that also undergird their emergency decision making and activation of heat plans. As explained by Klinenberg (2015), meteorologists tend to focus on the ”principal cause” of an extreme heat event, such as by measuring the slow moving, hot, humid mass of air, the upper-level ridge of high pressure, or the moist ground conditions; however the impacts of extreme heat are dependent upon the material and social structures of society (Klinenberg 2015). These are things that cannot easily be measured with a thermometer and require additional domain knowledge.

Wet-bulb globe temperature was also a challenging term for both participant groups to define and categorize. Notably, many EMs chose not to try to define the term by opting to group it with other indices. In contrast, ATMs frequently placed it with other weather terms but also suggested that it had limited use as a tool for measurement. There have been recent calls within the ATM community for the increased use of wet-bulb globe temperature by those who recognize the value it adds in comparison with the more commonly referenced heat index (CBS News 2022). Heat index takes into consideration temperature and humidity and is calculated for shady areas; wet-bulb globe temperature is a measure of heat stress in direct sunlight, which considers temperature, humidity, wind speed, sun angle, and cloud cover (solar radiation) (NWS Tulsa n.d.). For those who work or exercise in direct sunlight, such as athletes and outdoor workers, monitoring wet-bulb globe temperature can help to guide and manage activities safely (Hosokawa et al. 2019). The lack of familiarity with wet-bulb globe temperature among both participant groups was met with some resistance, suggesting that jargon is sometimes challenging for domain experts as well.

Indeed, researchers have identified key problems that are likely to occur when jargon is used. Specifically, jargon is alienating, undermines comprehension, and reduces content engagement among those who lack familiarity or expertise with technical language (Shulman and Bullock 2020). In this study, while we find that ATMs and EMs frequently sorted cards into similar groupings, their sorting approaches differed. ATMs with high levels of skill and knowledge approached the activity as a bit of a challenge, testing their knowledge and demonstrating their ability to correctly organize terms they use in their daily tasks. EMs also made use of their knowledge and approached card sorting in a manner meaningful to their profession of serving the public. When faced with scientific jargon, we observed some EMs struggle to define concepts and others to question their relevance to communicating about heat risk. These finding echoes that of Sivle and Aamodt (2019) who wrote: "weather language containing technical terms and jargon is precise and highly effective for those familiar with the language. Non-expert end-users might feel excluded by such language, and for this audience everyday words might improve the readability and understandability" (p. 438).

Specifically, for ATMs, there was a stated or implied familiarity with most weather or atmospheric heat-related words. In contrast, many EMs expressed a lack of confidence in what many terms meant and were unsure about their usefulness in explaining, measuring, or deciding about actions to recommend in response to heat. Some EMs chose to guess at terms they were unfamiliar with, others elected to discard them. Prior research on jargon suggests that this disengagement is not an unusual response (Bullock et al. 2019). Because people dislike effortful processing, when information looks complicated, they are more likely to stop processing this information (Shulman and Bullock 2020). Participants in both groups described concerns about the use of terms and concepts that were too technical, suggesting instead the need for a plain language approach that is accessible, timely, and relevant.

Overall, as extreme heat events increase in frequency and intensity, there will also be a need for better communication about the risks it poses, identifying the populations that are most vulnerable, and the protective actions that should be taken to reduce the loss of life. Scientific expertise will be necessary for measuring and modeling the future impacts using jargon for clarity's sake among other scientists; however, transforming those ideas and communicating them clearly will require a conscious and deliberate effort, recognizing the needs of the audience.

8. Recommendations and Conclusions

As excessive heat events become endemic, it will become even more important to understand what people understand about heat, its effects, and the terminology used to describe it. For example, heat index is widely used by atmospheric scientists in their social media communication, but this jargon frequently lacks an explanation about its measurement and its relationship to temperature or heat impacts (Sutton et al. 2022). While ATM domain experts may clearly understand heat index and what it represents, communicating it effectively may prove difficult if users have only a vague idea of what the information actually means.

Risk communication scholars on warnings for imminent threat have found that understanding the message, including the hazard severity and impacts, and the actions to reduce harm, is the first step necessary prior to taking protective action (Sutton & Kuligowski 2019). If technical concepts are used to communicate threat, adding plain language (i.e. heat index can also be referred to "feels like" or "apparent temperature") will make this information more meaningful to end users. This is likely to be even more important for terms such as wet-bulb globe temperature, which was largely unfamiliar to both groups of participants.

The challenge of communicating about heat is exacerbated by its slow onset and persistence as a "silent killer." Not only have researchers found that people tend to tune out when jargon is used, the lack of dramatic images to portray the devastation of heat coupled with the tendency for people to equate hot weather with normal summer conditions suggests that communicating risk under these conditions will become increasingly difficult. ATMs and EMs alike will need to make watches, warnings, and advisories meaningful by emphasizing not just temperatures and index values, but explaining the potential impacts associated with extended exposure and protective actions, especially for those groups that are more vulnerable.

Importantly, both EMs and ATMs identified additional terms and concepts during the card sorting exercise that should be considered for inclusion in future risk communication efforts. They specifically identified additional members of vulnerable populations (i.e., homeless or unhoused, non-native English speakers, outdoor workers, and persons without access to indoor air conditioning), additional impacts associated with exposure to sun and heat (i.e., dehydration, sweating, and sun poisoning) and plain language that can replace or supplement technical jargon. These added terms capture aspects of heat that may be traditionally overlooked but can add depth and raise awareness when communicating about extreme heat.

a. Limitations

This is a mixed methods study utilizing a novel method drawn from usability studies to investigate the mental models of expert and non-expert groups through card sorting. While we do not present statistical tests for significant differences between the two groups, this study offers insight into the thoughts, processes, and prioritization of information about communicating extreme heat, that can lead to clearer communication in the future. The two participant groups differ by their area of focus, but they also differ by levels of experience – we interviewed mostly ATM students, who appeared to concentrate on how to correctly organize cards, and EM professionals, who appeared to focus on how to serve the public. The inclusion of ATM practitioners may have revealed different foci in comparison with the students who are at the beginning stages of their meteorology careers.

b. Next Steps

Future research should include investigations of the historical use of scientific and technical language used in public-facing risk communication. Research focused on identifying the mental models, knowledge, and understanding about heat, heat impacts, and heat health risks can lead to plain language explanations that are better understood by all populations. Future attention should also be placed on tailoring risk communication, with a focus on websites. Determining for whom content is designed (domain experts, decision makers, or the public) should guide the decisions about what content that is included or prioritized. Understanding publics motivations for information seeking should affect content organization and identifying how information is consumed should determine how it is presented.

Acknowledgments.

The authors thank the professionals from emergency management and atmospheric science who shared their perceptions of extreme heat for this research. This research was supported by funding from NOAA (NA210AR4590360). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funder.

Data Availability Statement.

Due to privacy and ethical concerns, the data are not publicly available. Anonymized and de-identified interview data will be made available upon request to the corresponding author.

Accepted for publication in Weather, Climate, and Society, March 2023.
https://doi.org/10.1175/WCAS-D-22-0108.1

REFERENCES

American Meteorological Society, 2019: wet-bulb globe temperature. *American Meteorological Society*.

Asgharpour, F., D. Liu, and L. J. Camp, 2007: Mental Models of Security Risks. *International conference on financial cryptography and data security*, 367–377, https://doi.org/10.1007/978-3-540-77366-5_34.

Balachandran, K., kardSort.

Benson, R. W., 1985: THE END OF LEGALESE: THE GAME IS OVER. *N.Y.U Review of Law & Social Change*, **13**, 520–573.

Bullock, O. M., D. Colón Amill, H. C. Shulman, and G. N. Dixon, 2019: Jargon as a barrier to effective science communication: Evidence from metacognition. *Public Understanding of Science*, **28**, 845–853, https://doi.org/10.1177/0963662519865687.

CBS News, 2022: What is wet-bulb globe temperature? The weather index that's becoming more important. *WTOP News,* https://wtop.com/weather-news/2022/08/what-is-wet-bulb-globe-temperature-the-weather-index-thats-becoming-more-important/ (Accessed September 5, 2022).

Chen, D. A., G. D. Hoople, N. Ledwith, E. Burlingame, S. D. Bush, and G. E. Scott, 2020: Exploring faculty and student frameworks for engineering knowledge using an online card sorting platform. *International Journal of Engineering Pedagogy*, **10**, 62–81, https://doi.org/10.3991/ijep.v10i1.11336.

Climate.gov Staff, 2012: With rising greenhouse gases, U.S. heat waves to become more common & longer-lasting. *Climate.gov,* https://www.climate.gov/news-features/understanding-climate/rising-greenhouse-gases-us-heat-waves-become-more-common-longer (Accessed August 30, 2022).

Conrad, L. Y., and V. M. Tucker, 2019: Making it tangible: hybrid card sorting within qualitative interviews. *Journal of Documentation*, **75**, 397–416, https://doi.org/10.1108/JD-06-2018-0091.

Creswell, J. W., and D. L. Miller, 2000: Determining Validity in Qualitative Inquiry. *Theory Pract*, **39**, 124–130, https://doi.org/10.1207/s15430421tip3903_2.

EPA.gov, 2021: Climate Change Indicators: Heat-Related Deaths. *EPA.gov*, https://www.epa.gov/climate-indicators/climate-change-indicators-heat-related-deaths#:~:text=Some%20statistical%20approaches%20estimate%20that,set%20shown%20in%20Figure%201 (Accessed August 30, 2022).

Faiks, A., and N. Hyland, 2000: Gaining User Insight: A Case Study Illustrating the Card Sort Technique. *Coll Res Libr*, **61**, 349–357, https://doi.org/10.5860/crl.61.4.349.

Fincher, S., and J. Tenenberg, 2005: Making sense of card sorting data. *Expert Syst*, **22**, 89–93.

Grundstein, A. J., and C. A. Williams, 2018: Heat Exposure and the General Public: Health Impacts, Risk Communication, and Mitigation Measures. *Human Health and Physical Activity During Heat Exposure*, 29–43.

Hosokawa, Y., and Coauthors, 2019: Activity modification in heat: critical assessment of guidelines across athletic, occupational, and military settings in the USA. *Int J Biometeorol*, **63**, 405–427, https://doi.org/10.1007/s00484-019-01673-6.

Hudson, W., 2018: SynCaps.

Klinenberg, E., 2015: *Heat wave: a social autopsy of disaster in Chicago*. Second edition. University of Chicago Press.

Lazo, J. K., H. R. Hosterman, J. M. Sprague-Hilderbrand, and J. E. Adkins, 2020: The Value of Impact-Based Decision Support Services: Case Studies with Winter Storms. *Bull Am Meteorol Soc*, **101**, 975–980, https://doi.org/10.1175/BAMS-D-18-0153.1.

National Weather Service, 2019: *2019-2022 Strategic Plan*.

——, 2022: Weather Related Fatality and Injury Statistics National Program Weather Related Fatality and Injury Statistics. *Weather.gov*, https://www.weather.gov/hazstat/ (Accessed August 31, 2022).

National Weather Service Albuquerque, NWS ABQ Monsoon Awareness - Heat Stress. *National Weather Service*.

National Weather Service Green Bay, Heat Safety. *National Weather Service*.

National Weather Service Tulsa, Wet-Bulb Globe Temperature. *National Weather Service*.

Ocampo-Agudelo, J., and J. Maya, 2021: Core Concepts Linking Exhibit Design and the Visitor Experience in Science Centers: An Early Framework. *Culture and Computing. Design Thinking and Cultural Computing*, 113–130.

Reese, T., N. Segall, P. Nesbitt, G. del Fiol, R. Waller, B. C. MacPherson, J. E. Tonna, and M. C. Wright, 2018: Patient information organization in the intensive care setting: Expert knowledge elicitation with card sorting methods. *Journal of the American Medical Informatics Association*, **25**, 1026–1035, https://doi.org/10.1093/jamia/ocy045.

Righi, C., J. James, M. Beasley, D. L. Day, J. E. Fox, J. Gieber, and C. Howe, 2013: *Card Sort Analysis Best Practices*. 69–89 pp. https://uxpajournal.org/card-sort-analysis-best-practices-2/ (Accessed August 30, 2022).

Rosenfeld, L., P. Morville, and J. Arango, 2015: *Information Architecture: For The Web and Beyond*. 4th ed. O'Reilly Media, Inc.

Saunders, M. N. K., 2015: Using mixed methods – combining card sorts and in-depth interviews. *Handbook of Research Methods on Trust*, Edward Elgar Publishing, 134–144.

Savage, J. S., H. A. Harris, J. A. Bleser, and B. Y. Rollins, 2019: Parents' and children's categorization of candy are similar: A card sort protocol. *Nutrients*, **11**, https://doi.org/10.3390/nu11102472.

Sharon, A. J., and A. Baram-Tsabari, 2014: Measuring mumbo jumbo: A preliminary quantification of the use of jargon in science communication. *Public Understanding of Science*, **23**, 528–546, https://doi.org/10.1177/0963662512469916.

Shulman, H. C., and O. M. Bullock, 2020: Don't dumb it down: The effects of jargon in COVID-19 crisis communication. *PLoS One*, **15**, https://doi.org/10.1371/journal.pone.0239524.

Silverman, J., S. Kurtz, and J. Draper, 2016: *Skills for Communicating with Patients*. 3rd ed. CRC Press, 328 pp.

Sivle, A. D., and T. Aamodt, 2019: A dialogue-based weather forecast: adapting language to end-users to improve communication. *Weather*, **74**, 436–441, https://doi.org/10.1002/wea.3439.

Spencer, D., and J. J. Garrett, 2009: *Card sorting: designing usable categories*. 162 pp.

Sutton, J. N., M. K. Olson, and N. A. Waugh, 2022: Communicating extreme heat risk to at risk publics: Content analytic study of tweets from 2021. *6th Conference on Weather Warnings and Communications*, Milwaukee.

Tullis, T., and L. Wood, 2004: *How Many Users Are Enough for a Card-Sorting Study?*

Uccellini, L. W., and J. E. ten Hoeve, 2019: Evolving the National Weather Service to Build a Weather-Ready Nation: Connecting Observations, Forecasts, and Warnings to Decision-Makers through Impact-Based Decision Support Services. *Bull Am Meteorol Soc*, **100**, 1923–1942, https://doi.org/10.1175/BAMS-D-18-0159.1.

Wellington, J., and J. Osborne, 2001: Language and Literacy in Science Education. *McGraw-Hill Education (UK)*.

Wentzel, J., F. Müller, N. Beerlage-de Jong, and J. van Gemert-Pijnen, 2016: Card sorting to evaluate the robustness of the information architecture of a protocol website. *Int J Med Inform*, **86**, 71–81, https://doi.org/10.1016/j.ijmedinf.2015.12.003.

Wilhelmi, O., and M. Hayden, 2016: Reducing Vulnerability to Extreme Heat Through Interdisciplinary Research and Stakeholder Engagement. *Extreme Weather, Health, and Communities: Interdisciplinary Engagement Strategies*, S.L. Steinberg and W.A. Sprigg, Eds., Springer International Publishing, 165–186.

Williams, N., and J. Ogden, 2004: The impact of matching the patient's vocabulary: A randomized control trial. *Fam Pract*, **21**, 630–635, https://doi.org/10.1093/fampra/cmh610.