



Using Subjective Measures for Numeracy and Spatial Cognition to Predict Accuracy of Interpreting a Tropical Forecast Graphic

ZOEY ROSEN

Institute for Public Policy Research & Analysis, Norman, OK

MARILEE LONG

Colorado State University, Fort Collins, CO

(Manuscript received 22 January 2025; review completed 28 August 2025)

ABSTRACT

The weather domain has given a lot of attention towards numeracy, meaning a person's ability to understand numerical information, for predicting how accurately different groups interpret forecast information. However, these forecasts are often presented graphically with a map, yet not many studies include a geographic equivalent to numeracy in their measures. Additionally, despite the cognitive load on participants, much of the literature measures numeracy and spatial cognition with mathematical tests. This study provides justification from interdisciplinary literature for using subjective scales to measure spatial cognition, or the ability to understand maps, along with numeracy to predict accuracy. Using a new wind speed exceedance threshold (WSET) tropical cyclone graphic, we tested the relationships between numeracy, spatial cognition, and accuracy in a public survey ($n = 624$) and with nineteen experts (emergency managers and meteorologists) from Louisiana and Florida. Numeracy independently predicted accuracy of interpretation, with spatial cognition having a small, statistically nonsignificant effect. This study provides empirical support for measuring numeracy and spatial cognition using subjective scales, as well as an accuracy measure for a new tropical forecast graphic. Future studies should continue to measure both numeracy and spatial cognition when assessing how individual characteristics impact forecast graphic interpretation.

1. Introduction and background

Forecast graphics are critical products during extreme weather, presenting complex probabilistic information for a wide range of users to understand. Forecasters use several methods for communicating uncertainty in a single map graphic. For instance, a typical weather graphic contains (a) a map with geographic scaling to show the spatial distribution of the forecasted hazard's potential effects, and (b) a probabilistic numerical range in the title or legend to show the percent chance of a hazard's impacts. Studies show that when presented clearly, numerical

probability information generally improves decision quality, which is important for weather threats (Morss et al. 2010; Ripberger et al. 2022a,b). In fact, many studies across social science disciplines have been done to understand how numeracy, or a person's numeric literacy, influences how accurately they interpret probabilistic information (Adams et al. 2017; Peters et al. 2019; Ripberger et al. 2022b). This study contributes to our understanding of whether there is a similar relationship between accuracy and spatial cognition, which describes a person's literacy in reading maps, by experimentally testing numeracy, spatial cognition, and accuracy measures with a new forecast map graphic.

Numeracy and spatial cognition are not both typically measured when researchers study the accuracy of map interpretation, yet together they would more fully describe the abilities people use to understand a forecast graphic. Furthermore, it is important to understand how different audiences interpret meteorological data products, because how they understand this information influences how they may make decisions.

1.1. Numeracy

Numeracy is defined as a person's ability to understand or work with numbers, such as probabilities (Drake 2012; Peerbolte and Collins 2013; Adams et al. 2017; Peters et al. 2019; Ripberger et al. 2022a,b). On average, the public and emergency managers (EMs) are less numerate than meteorologists, which impacts their comprehension of forecast graphics that contain probability information (Wernstedt et al. 2019; Ripberger et al. 2022a). Designing graphics so that people with lower numeracy understand them has become an increasing practice by forecasters to have the greatest likelihood of user groups accurately comprehending the forecast (Ripberger et al. 2022a,b). For instance, studies indicate that graphics with probability information should use both text and numbers to assist those with lower numeracy (Lenhardt et al. 2020; Rosen et al. 2021). Especially for graphics about tropical cyclones that have large impacts on communities and are regularly communicated with numbers, it is important to know how the public and experts interpret the numerical information (Drake 2012; Losee et al. 2017; Dormady et al. 2022).

Numeracy is an important skill in adults, as the ability to understand and accurately interpret numbers is necessary for everyday financial and health decision-making (Peters et al. 2019). However, numeracy is more nuanced than just an objective ability to work with numbers; there is also numeric self-efficacy, which describes one's confidence in their abilities to use numbers (Peters et al. 2019). Having higher numeric self-efficacy is related to higher engagement with numerical information and higher completion rates for mathematical tasks (Peters et al. 2019). Self-efficacy is a key part of numeracy, as numeracy depends on a person's ability to use numbers both accurately and confidently (Scott 1999; Peters et al. 2019).

For this study, numeracy is defined as numeric literacy, meaning one's ability to process and understand basic probability and numerical concepts (Peters et

al. 2006, 2019). Numeracy has been measured in other studies primarily through objective measures, like mathematical tests (Fagerlin et al. 2007). Some common mathematical tests that have been used include answering items that convert percentages to proportions and vice versa, calculating fractions, and converting probabilities to percentages and proportions (Schwartz et al. 1997; Lipkus et al. 2001; Peters et al. 2006). Mathematical testing is a direct way to assess numeracy and has been used successfully in studies of risk perception; however, many authors have found problems with using this approach (Lipkus et al. 2001; Fagerlin et al. 2007). One major issue is the additional cognitive load put onto participants when they must take a math test, which has led to lower completion rates for studies and higher attrition rates in longitudinal studies (Fagerlin et al. 2007). Also, there have been concerns with using these tests in online experimental designs because participants can use a calculator or look up the answer in another tab without the researcher knowing, thus inflating their numeracy levels (Fagerlin et al. 2007). Thus, subjective scales were developed to provide a quicker, less difficult-to-obtain measure that still accurately reflects numeracy levels (Zikmund-Fisher et al. 2007).

1.2. Spatial cognition

Spatial cognition is defined as the ability to use, navigate, and interpret information effectively in a spatial representation, like a map (Tomaszewski et al. 2020). Similar terms like spatial awareness and spatial thinking ability are used in the cartography, emergency management, and forecasting disciplines to describe how people familiarize themselves in space and draw conclusions based on where items are located in a map (Alexander 2004; Hegarty 2010; Kane 2014; Houston et al. 2019). In this study, "spatial cognition" means a person's cognitive ability to understand the location of items in a space when those items are shown in maps. Like numeracy, spatial cognition has been measured using both objective and subjective assessments. An example of a widely used objective assessment is the spatial thinking ability test (STAT), a 16-question, standardized assessment that measures people's ability to think spatially, by integrating items to assess geography knowledge and spatial skills (Lee and Bednarz 2012).

Although STAT is widely used, it has been criticized. In a companion publication to Lee and Bednarz (2012),

Bednarz and Lee (2011)¹ could not confirm the reliability or the validity of the components in STAT. For reasons such as that STAT does not measure verbal strategies to solve spatial problems, the authors noted that STAT was not the “optimal assessment instrument for uncovering the components of spatial thinking” (Bednarz and Lee 2011). Objective measures may not be the most appropriate test to use for spatial cognition in this study because while forecast graphics use geography content, the graphics are layered with verbal information as well that contribute to how the data are read and used. Thus, to capture this ability more robustly, this study took a self-report approach to measure spatial cognition, like the one used to measure numeracy.

1.3. Tropical context and hypotheses

A recent collaboration between the Cooperative Institute for Research in the Atmosphere (CIRA) and the National Hurricane Center (NHC) provided the opportunity to test the relationships between numeracy, spatial cognition, and map interpretation accuracy using a new graphic. In 2022, CIRA and NHC began collaborating on redeveloping the windspeed probability model (WSP2.0) to better estimate wind speed over land as well as make other technical improvements to the model. Wind speed exceedance thresholds (WSETs) describe tropical cyclone (TC) sustained (1-minute) surface wind hazards, corresponding to a pre-specified probability of occurrence and can be mapped from WSP2.0². The WSET graphic was developed to communicate wind exceedance for hurricanes to the public and experts, such as meteorologists and EMs. As the graphic had not been made before, we had the opportunity to test the relationships between numeracy, spatial cognition, and accuracy in both audiences with a graphic that has no prior interpretations. Additionally, this study used subjective measures intentionally for numeracy and spatial cognition to extend previous work that aimed to limit cognitive load on participants, while still capturing the relationships of interest in this study (Hegarty et al. 2002; Zikmund-Fisher et al. 2007; Fagerlin et al. 2007; Hegarty et al. 2018). By measuring

numerical and map abilities, we hope to gain insight into these groups’ accuracy with interpreting meteorological data in different formats.

Relationships between spatial cognition and numeracy have been found, though these concepts are not often measured together (Kane 2014). Decision-making often involves numerical assessment as well as a consideration of the local context, so including spatial cognition and numeracy in studies of how people interpret extreme weather forecast graphics could elaborate on how choices are made by individuals for a given situation in a community or a professional domain (Dowling 1997). EMs and meteorologists interpret graphical forecasts as part of their professions, so examining these groups as well as the public affords an opportunity to see where or if there are differences. This study aims to extend our understanding of the relationship between numeracy, spatial cognition, and accuracy by investigating it in the context of hazard graphic interpretation for expert user groups and the public through the following hypotheses:

H1: Experts (EMs and meteorologists) will be more accurate in interpreting the probability information in the WSET graphic than will a public sample.

H2: Numeracy and spatial cognition will be higher in the expert group (EMs and meteorologists) than a public sample.

H3: For both groups, numeracy and spatial cognition will predict the accuracy of interpretation of the graphic.

2. Measurement and methods

This mixed-methods study consisted of semi-structured interviews asking quantitative and qualitative questions with experts that use hurricane graphics (meteorologists and EMs) and a survey of the public. Mixed-methods approaches are useful to compare within and across types of questions, as well as for elaborating on the experiences and perceptions of individuals and groups more broadly (Frels and Onwuegbuzie 2013). Both the expert participants ($n = 19$) and the public participants ($n = 624$) resided in Florida or Louisiana, states that routinely experience hurricanes. During sampling, which is described more in Section 2.4, effort was made to ensure the expert sample represented those that work at the county or state level for the two states. Thus, whereas the expert group is smaller than the public sample, we checked for normality before determining what statistical tests were

¹ Interestingly, the authors released a critique of STAT in 2011 through conference proceedings before the test was published in 2012 in the *Journal of Geography*.

² This study is part of a larger project on the development of WSETs, focusing on how the information is interpreted by its intended audiences. See Rosen et al. (2024) for more information.

appropriate for the data; non-parametric tests were used for the expert sample and parametric tests were used for the public sample.

The following sections detail empirical support for measuring numeracy and spatial cognition with subjective scales, as well as rationale for extending these variables together to predict accuracy. After, the sampling procedure and scale reliabilities for each population of interest are discussed.

2.1. Measuring numeracy

In this study, participants were asked to self-report their level of numeracy. They completed the subjective numeracy scale (SNS), an 8-item Likert scale that assesses subjective numeracy through participants' reported abilities with numeric information (Fagerlin et al. 2007). The SNS (Appendix A) has been well-validated through multiple surveys of representative samples; these studies have shown that the SNS predicts objective numeracy abilities (Zikmund-Fisher et al. 2007). Important for this study, the SNS has been used in the weather domain in a study of how EMs dealt with meteorological uncertainty and probabilistic formats for forecasting in a testbed experiment, finding that they had a high preference and ability with using numbers (Adams et al. 2017). The SNS has also been used in a risk communication study of the public; Hess et al. (2011) found the self-report scale was the most appropriate for risk graphics because it captures both self-efficacy and mathematical ability for numeracy. Thus, using a subjective numeracy scale is appropriate for this study as it has been found to work in both the populations and domains of this study.

2.2. Measuring spatial cognition

Sense of direction has been measured subjectively through self-report items, which have shown to be a more promising approach to predicting spatial cognition (Hegarty et al. 2002, 2018). The Santa Barbara Sense of Direction (SBSOD) scale is widely used, reliable, and has been validated empirically in other studies (Hegarty et al. 2002; Weisberg et al. 2014). The 15-item scale (Appendix B) also correlates with people's ability to locate landmarks, describing not only a general sense of direction, but people's navigation ability and comfort with operating spatially (Hegarty et al. 2002). Even though there are some concerns around how predictive self-report scales are when compared

to tests of spatial cognitive abilities (Thorndyke and Goldin 1981; Taekeuchi 1992), Hegarty et al. (2002) found in their multi-study experimental design that people were accurate in their assessments of their own spatial abilities. This result suggests that the SBSOD is a more suitable way to assess spatial cognition and was therefore used in this study.

2.3. Measuring accuracy

Accuracy can be defined as the correctness someone has in their calculations (Schwartz et al. 1997). This study is concerned with accuracy in interpreting the WSET forecast graphic, meaning how accurate the participant is when they are asked to determine a value from the map graphic. To measure accuracy, participants were asked what the wind exceedance was on prototype graphics of the WSET at two designated locations, Point A and Point B (Fig. 1). Their answers were compared to the correct value. Using the legend on the graphic, the answer was given an accuracy score. Participants were given a zero if they were correct, and their answer was scored one additional point for each category the answer was away from accurate. For

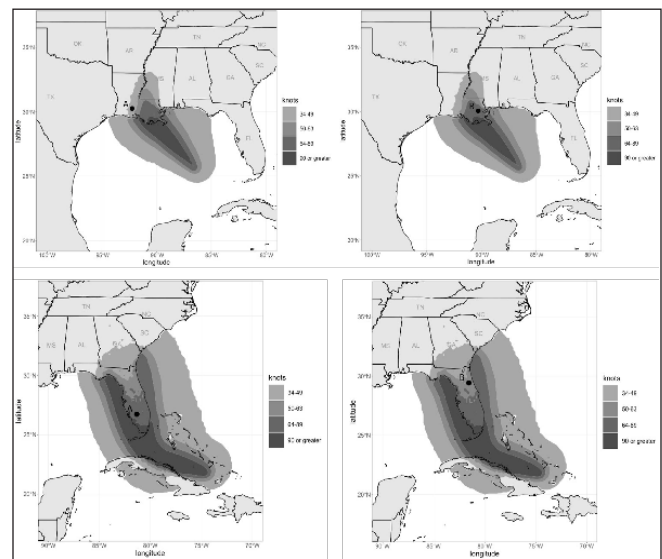


Figure 1. WSET graphics shown to LA (top) and FL participants (bottom) with Point A (left) and Point B (right) to measure accuracy of interpretation. At the recommendation of the research team, the one difference between graphics for participants is the wind speeds were shown in miles per hour for the public sample instead of knots, which was shown to the experts. *Click image for an external version; this applies to all figures and hereafter.*

| | | |
|--|--|------------------|
| What is the windspeed at point A? | | |
| Incorrect | <input type="radio"/> 34-49 knots | Accuracy Score 1 |
| Correct | <input checked="" type="radio"/> 50-63 knots | Accuracy Score 0 |
| Incorrect | <input type="radio"/> 64-89 knots | Accuracy Score 1 |
| Incorrect | <input type="radio"/> 90 knots or greater | Accuracy Score 2 |

Figure 2. Diagram of how accuracy scores were determined from participant responses.

example, if the windspeed was in the $25.8\text{--}32.5\text{ m s}^{-1}$ (50–63 kt) range and the participant said it was $\geq 46.4\text{ m s}^{-1}$ (90 kt), their assessment was scored a 2 as it was 2 categories away from correct. An explanatory figure for this score is shown in Fig. 2.

2.4. Sampling and procedure

2.4.1. EXPERT SAMPLE

Expert participants were asked to answer a quantitative questionnaire ahead of a larger interview, items of which are what is being analyzed for this study. Participants were asked quantitative items for the concepts related to the study such as spatial cognition and numeracy, in addition to demographics like gender or work experience. These questionnaires were collected in October and November 2022, taking the participants no longer than 10 minutes. The final sample size of experts was 19 participants (ten meteorologists and nine emergency managers). The length of their careers was 19.37 years on average (median: 20 years, max: 33 years, min: three years). The sample was predominately male (78.95%; $n = 15$) and of the sample, six experts were from Louisiana (31.58%) and 13 were from Florida (68.42%).

2.4.2. PUBLIC SAMPLE

The public answered an online survey with the same items from the quantitative questionnaire used in the expert sample. Qualtrics was used to recruit a national representative sample. Quotas were set to mirror census data for Florida and Louisiana and are proportional to each state's categories for age, race, and gender (United States Census Bureau 2021a, b). The survey questions and prototypes were asked and shown in the same order as the experts. Participants were

shown prototype graphics for their state of residence. Metadata from Qualtrics showed that the response time was similarly around 10 minutes, $M = 0:13:59$ minutes (median: $0:10:46$ minutes). The survey was launched on 5 December 2022 and closed on 4 January 2023. There were 624 participants in the public sample. Additional demographics for this sample are in Appendix C.

2.5. Reliability of scales and normality of data

2.5.1. EXPERT SAMPLE

Each of the 19 participants answered the survey items, which were analyzed using IBM's SPSS version 26 software. As shown in Table 1, the scales used to measure numeracy and spatial cognition exhibited scale reliability with Cronbach alpha values that were in the acceptable range. Because accuracy was measured using a single item, it did not have an alpha value. To assess the normality of the data, a Shapiro-Wilk test was used because this test is highly recommended for sample sizes under 30 (Ghasemi and Zahediasl 2012). The W statistics for both variables were not statistically significant (Table 1). Therefore, we conclude that the data for numeracy and spatial cognition are normally distributed, so parametric tests are appropriate within the expert sample, and non-parametric tests are appropriate when comparing with the larger public sample.

2.5.2. PUBLIC SAMPLE

For the public sample ($n = 624$), the scales that measured numeracy and spatial cognition were reliable (Table 2). As in the expert sample, accuracy did not have an alpha value as it was measured with a single item. Histograms for each scale variable were inspected to see if the data appeared normal, also known as visually checking the data. The data followed the statistical criteria where normality can be assumed, so parametric tests are also appropriate to use for this sample (Ghasemi and Zahediasl 2012).

To boost the generalizability of the findings, we ran independent samples t-tests to determine whether there were statistically significant differences by state (Florida or Louisiana) in participants' numeracy and spatial cognition scores. As shown in Table 3, numeracy and spatial cognition did not differ by state; thus, we could combine the two groups when testing the hypotheses.

Table 1. Reliability (alpha) and Normality (W) Statistics for numeracy and spatial cognition scales for the expert sample data.

| Variable | Scale | Cronbach's α | Shapiro's W | p-value |
|-------------------|-----------------------------|---------------------|-------------|---------|
| Numeracy | SNS (Fagerlin et al. 2007) | 0.77 | 0.95 | 0.45 |
| Spatial Cognition | SBSOD (Hegarty et al. 2002) | 0.75 | 0.96 | 0.58 |

Table 2. Reliability Statistics for Public Sample Data.

| Variable | Scale | Cronbach's α |
|-------------------|-----------------------------|---------------------|
| Numeracy | SNS (Fagerlin et al. 2007) | 0.81 |
| Spatial Cognition | SBSOD (Hegarty et al. 2002) | 0.86 |

Table 3. Independent samples t-tests for scales in public sample, by state.

| Variable | State | Mean | SD | t | df | p-value |
|-------------------|-----------|------|------|-------|-----|---------|
| Numeracy | Louisiana | 4.01 | 1.04 | -1.44 | 622 | 0.81 |
| | Florida | 4.13 | 0.99 | | | |
| Spatial Cognition | Louisiana | 4.50 | 1.05 | -1.61 | 622 | 0.11 |
| | Florida | 4.63 | 0.95 | | | |

3. Results

H1 predicted that expert users (EMs and meteorologists) would be more accurate in interpreting the probability information in the wind exceedance graphic than would a public sample. On average, the accuracy scores for the expert group were closer to zero (more accurate) than the public sample ($M_{\text{expert}} = 0.08$, $SD_{\text{expert}} = 0.25$, $\text{Range}_{\text{expert}} = 1$; $M_{\text{public}} = 0.66$, $SD_{\text{public}} = 0.60$, $\text{Range}_{\text{public}} = 2.5$). As expected, the range of accuracy scores was also larger in the public sample than the expert group³ (Fig. 3). Not only were the experts more accurate on average, but they also had a tighter distribution of accuracy scores.

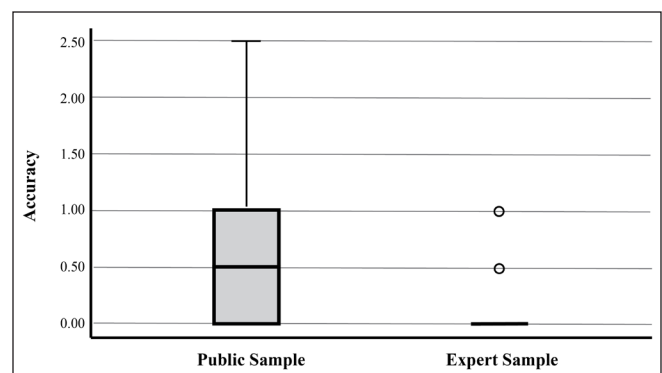
A Mann-Whitney U test was performed to evaluate whether accuracy differed by group (expert or public). The results indicated that the expert group was significantly more accurate in interpreting the wind exceedance graphic than the public group, $z = -4.53$, $p < 0.01$. Therefore, there was support for H1.

H2 predicted that numeracy and spatial cognition will be higher in the expert group (EMs and meteorologists) than in the public sample. The numeracy values in the

public sample were lower than in the expert group (Table 4). Spatial cognition was also lower in the public sample than in the expert group (Table 4).

A Mann-Whitney U test was performed to evaluate whether numeracy differed by group. The results indicated that the expert group was significantly more numerate than the public group, $z = -4.23$, $p < 0.01$. A Mann-Whitney U test was performed to evaluate whether spatial cognition differed by group. The results indicated that the expert group had significantly higher spatial cognition than the public group, $z = -3.89$, $p < 0.01$. As evidenced by both tests, H2 was supported.

H3 predicted that numeracy and spatial cognition will predict the accuracy of graphic interpretation. In the expert user group, the sample size was too small

**Figure 3.** Accuracy scores for the wind exceedance graphic for public and expert samples.

³ Although there were two outliers in the expert group, these experts have many years of experience interpreting these types of graphics and they said in their interviews that their answer may be incorrect owing to the difficulty they had with the grayscale color scheme, not the numerical content itself.

Table 4. Descriptive statistics for numeracy (6-point scale) and spatial cognition (7-point scale) in both samples.

| Variable | Sample | M | SD |
|-------------------|--------|------|------|
| Numeracy | Expert | 5.05 | 0.64 |
| | Public | 4.08 | 1.02 |
| Spatial Cognition | Expert | 5.42 | 0.65 |
| | Public | 4.57 | 1.00 |

to run a regression. Experts' high scores for numeracy, spatial cognition, and accuracy point towards a positive relationship among these variables. However, because of statistical constraints, only the public sample was used for this analysis. In the public sample, a bivariate regression was conducted to test whether numeracy and spatial cognition predicted accuracy of interpreting the wind exceedance graphic. As a reminder for the public sample, the average numeracy score was 4.08 (SD = 1.02) on a 6-point scale, the average spatial cognition score was 4.57 (SD = 1.00) on a 7-point scale, and the average accuracy score was 0.66 (SD = 0.60), where scores closer to zero are more accurate. Results from the bivariate regression showed that numeracy did significantly predict accuracy ($\beta = -0.19$, $t(592) = -4.64$, $p < 0.01$). However, spatial cognition did not predict accuracy ($\beta = -0.02$, $t(592) = 0.64$, $p = 0.53$), and the interaction between numeracy and spatial cognition was also nonsignificant ($\beta = -0.1$, $t(592) = -0.37$, $p = 0.71$). Therefore, H3 is only partially supported.

4. Discussion

Forecast map graphics rely on probabilistic information for dynamic threats that are necessary for a wide range of users not only to understand but quickly use to make decisions during extreme weather. Thus, a major goal is to design graphics that people interpret correctly regardless of their forecast expertise. Previous studies by social scientists in meteorology have focused on how users interpret information, centering around ability to decipher verbal or numerical information (Ripberger et al. 2022a,b). However, numeracy and spatial cognition also influence how accurate someone is with interpreting risk graphics like maps, despite not being measured in studies together (Hegarty et al. 2002; Roth 2009; Hess et al. 2011). This study tested three hypotheses about the relationships between people's ability to interpret risk information accurately and their numeracy and spatial cognition abilities. We tested these abilities to further our understanding of how people process information from hurricane forecast graphics with a new wind exceedance map graphic.

Adding spatial cognition measures into studies on forecast interpretation is relatively new, only being incorporated into research recently (Davenport 2022). This study provides an initial view into how measuring both spatial cognition and numeracy can elaborate into the characteristics of our populations of interest and how they interpret a new tropical cyclone forecast graphic, the WSET. As numeracy explains how people interpret numbers in risk messages like forecasts, spatial cognition can explain more about how people interpret map graphics, specifically, like a geographical equivalent to numeracy (Tomaszewski et al. 2020). Qualitative relationships between levels of spatial cognition and numeracy have been found, so we anticipated this relationship would be upheld within this quantitative study (Kane 2014). The expert group was significantly more numerate than the public group, in line with previous research on similar populations (Ripberger et al. 2022b). The expert group also had significantly higher spatial cognition than the public group (H2).

Bringing together the relationships between numeracy, spatial cognition, and accuracy, we hypothesized that numeracy and spatial cognition would predict the accuracy of interpreting the hurricane graphic (H3). The expert user group was highly accurate, numerate, and had high reported spatial cognition. Results from the public sample partially upheld the prediction; a bivariate regression analysis revealed that numeracy mainly predicted accuracy. Interestingly, this analysis shows that spatial cognition did not significantly predict accuracy. Knowing that one's levels of numeracy and spatial cognition may predict accuracy is valuable for future studies because it adds to our understanding of how information processing abilities could factor into how people interpret map graphics.

4.1. Limitations and future work

Whereas the public sample represented residents of Florida and Louisiana, Hurricane Ian may have influenced these survey respondents. Ian was a major

hurricane that made landfall in Florida about a month before data collection, which could have influenced those that paid more attention than usual to forecast graphics in those weeks in half the sample and were more familiar with interpreting hurricane wind data. To determine the degree of influence this may have had on participants, we measured experience using the hurricane experience score (Ehrlich et al. 2010) and found that there was a statistically significant difference in the scores for hurricane experience between participants living in Louisiana and those living in Florida. When controlling for age of participant, as the average age in Florida is older than Louisiana (United States Census Bureau 2021a,b), or length of residency in state, as newer residents would have less hurricane experience, the effect size of these variables was small and negligible. Therefore, we determined that hurricane experience in general did not impact the variables of interest in this study, and future studies that plan to collect data just after hurricane season should add similar controls and plan farther from hurricane season.

Participants in both studies were recruited from only two states, which could have influenced their accuracy scores because Louisiana and Florida have some of the most active hurricane seasons in the United States, so residents may be more familiar with interpreting forecast graphics. Other Gulf Coast and Atlantic Coast states in the US that experience hurricanes, such as Texas or North Carolina for example, could have provided different findings for both the experts and the public because they may engage with hurricane wind forecast graphics less often. To get a better understanding of how people interpret new hurricane graphics, more coastal states should be sampled in the future.

Wind exceedance has not been depicted in a forecast graphic in this way before, so we do not have previous studies on WSETs that could explain how accurately this phenomenon is interpreted. Although our study of spatial cognition and numeracy did not have an interaction effect with accuracy when modeled in a regression, we think this interaction would be present in a study that used a more well-known graphic. A study on storm surge, for example, may reveal this relationship between accuracy, spatial cognition, and numeracy because there are more previous studies on this hazard with the public and experts as the graphics have been disseminated to the public longer. If we knew more about how understandable the hazard itself is to the audience, we could evaluate how variations in its presentation and communication impact measured

relationships. Additionally, we may see a difference in interaction effects if the experiment was repeated outside of the tropical domain. Future research should also investigate the effects of numeracy, accuracy, and spatial cognition with other hazard forecast graphics.

5. Conclusion

The information an individual receives during disasters like hurricanes often influences their choice in taking protective action (Morss et al. 2016). Because of this, it is critical to have a full understanding of how forecast graphics are understood by their audiences, and ensure they are accessible, accurately interpreted, and prompt decision-making. This study had the opportunity to test individual's characteristics, such as numeracy and spatial cognition, with subjective scales to see how these variables impact accuracy of interpretation in experts and the public for a newly developed tropical forecast graphic.

As predicted, the expert participants had high numeracy and spatial cognition scores and likewise were highly accurate when interpreting the wind exceedance map graphic. The public had lower numeracy levels and lower spatial cognition scores than experts and were less accurate in interpreting the graphic. Despite the differences between samples, relationships between these variables helped to explain how people process and interpret forecast map graphics. Numeracy predicted the accuracy of interpretation for the graphic, though spatial cognition had a small, statistically nonsignificant effect. The relationships between accuracy, spatial cognition, and numeracy should continue to be studied for weather forecast graphics because they provide important insight into how people understand and engage with information, which can ultimately save lives in future storms. Especially because numeracy has been a focus in much of the scholarship around weather and probability communication, having a geographic equivalent like spatial cognition measured should also be a priority in future social science studies testing forecast graphics.

Funding. This work was partially supported by the Cooperative Institute for Research in the Atmosphere.

Institutional review board statement. The study was approved by the Institutional Review Board of Colorado State University protocol code 3592 approved on 21 June 2022, and an amendment to update procedures

was approved on 19 October 2022. Informed consent was obtained from all subjects involved in the study.

Acknowledgments. Thank you to Andrea Schumacher, Mark DeMaria, and Alan Brammer for their help throughout this project.

Data availability statement. Data are available from the corresponding author upon reasonable request.

APPENDIX A

Subjective Numeracy Scale (SNS)

Cognitive abilities (1 = not at all good, 6 = extremely good)

- How good are you at working with fractions?
- How good are you at working with percentages?
- How good are you at calculating a 15% tip?
- How good are you at figuring out how much a shirt will cost if it is 25% off?

Preference for display of numeric information

- When reading the newspaper, how helpful do you find tables and graphs that are parts of a story? (1 = not at all, 6 = extremely)
- When people tell you the chance of something happening, do you prefer that they use words (“it rarely happens”) or numbers (“there’s a 1% chance”)? (1 = always prefer words, 6 = always prefer numbers)
- When you hear a weather forecast, do you prefer predictions using percentages (e.g., “there will be a 20% chance of rain today”) or predictions using only words (e.g., “there is a small chance of rain today”)? (1 = always prefer percentages, 6 = always prefer words; reverse coded)
- How often do you find numerical information to be useful? (1 = never, 6 = very often)

APPENDIX B

Santa Barbara Sense of Direction (SBSOD) Scale

7-point Likert Scale (strongly disagree to strongly agree).

- I am very good at giving directions.
- I have a poor memory for where I left things. (REV)
- I am very good at judging distances.
- My “sense of direction” is very good.
- I tend to think of my environment in terms of cardinal directions (N, S, E, W).
- I very easily get lost in a new city. (REV)
- I enjoy reading maps.
- I have trouble understanding direction. (REV)
- I am very good at reading maps.
- I don’t remember routes very well while riding as a passenger in a car. (REV).
- I don’t enjoy giving directions. (REV).
- It’s not important to me to know where I am. (REV).
- I usually let someone else do the navigational planning for long trips. (REV).
- I can usually remember a new route after I have traveled it only once.
- I don’t have a very good “mental map” of my environment. (REV)

*NOTE: REV means “reverse coded” so the items all go in the same direction when being statistically analyzed.

APPENDIX C

Additional demographic information of the public survey participants. The following tables provide additional demographic information on the 624 participants in the public sample, such as race/ethnicity, highest level of education, employment status, and income.

Table C1. Public sample race/ethnicity demographic information.

| Race/Ethnicity | Frequency | Percent of Responses | Percent of Sample |
|--|-----------|----------------------|-------------------|
| Spanish, Hispanic, or Latino | 117 | 15.14% | 18.75% |
| White/Caucasian | 482 | 62.35% | 77.24% |
| Black/African American | 93 | 12.03% | 14.90% |
| American Indian/Native American or Alaska Native | 13 | 1.68% | 2.08% |
| Asian | 28 | 3.62% | 4.49% |
| Native Hawaiian or Other Pacific Islander | 6 | 0.78% | 0.96% |
| Other | 32 | 4.14% | 5.13% |
| Prefer not to say | 2 | 0.26% | 0.32% |
| Total: *773 | | | |

*Participants could check all that apply; N=624

Table C2. Public sample education demographic information.

| Highest Level of Education | Frequency | Percent |
|---|-----------|---------|
| Some high school or less | 21 | 3.40 |
| High school diploma or GED | 167 | 26.80 |
| Some college, but no degree | 128 | 20.50 |
| Associates or technical degree | 73 | 11.70 |
| Bachelor's degree | 149 | 23.90 |
| Graduate or professional degree (MA, MS, MBA, MD, PhD, JD, DDS, etc.) | 83 | 13.30 |
| Prefer not to say | 3 | 0.50 |
| Total | 624 | 100.00 |

Table C3. Public sample employment demographic information.

| Employment Status | Frequency | Percent |
|----------------------------------|-----------|---------|
| Working full-time | 281 | 45.00 |
| Working part-time | 84 | 13.50 |
| Unemployed and looking for work | 39 | 6.30 |
| Homemaker or stay-at-home parent | 25 | 4.00 |
| Student | 37 | 5.90 |
| Retired | 129 | 20.70 |
| Other | 29 | 4.60 |
| Total | 624 | 100.00 |

Table C4. Public sample employment income information.

| Total household income before taxes during the past 12 months | Frequency | Percent |
|---|-----------|---------|
| Less than \$25 000 | 136 | 21.80 |
| \$25 000 - \$49 000 | 133 | 21.30 |
| \$50 000 - \$74 000 | 131 | 21.00 |
| \$75 000 - \$99 000 | 83 | 13.30 |
| \$100 000 - \$150 000 | 86 | 13.80 |
| \$150 000 or more | 27 | 4.30 |
| Prefer not to say | 28 | 4.50 |
| Total | 624 | 100.00 |

REFERENCES

- Adams, T. M., S. J. Sanders, and C. A. Shivers-Williams, 2017: An exploratory examination of the impact of uncertainty on forecast data preference. *Global Aspect*, **7**, 11–15, [Crossref](#).
- Alexander, D. E., 2004: Cognitive mapping as an emergency management training exercise. *J. Contingencies Crisis Manage.*, **12**, 150–159, [Crossref](#).
- Bednarz, R. S., and J. Lee, 2011: The components of spatial thinking: empirical evidence. *Proc. Soc. Behav. Sci.*, **21**, 103–107, [Crossref](#).
- Davenport, C. E., 2022: *Characterizing the Growth of Spatial Thinking Abilities Across Meteorology Courses*.
- Dormady, N., A. Fasano, A. Roa-Henriquez, D. Flanagan, W. Welch, and D. Wood, 2022: Informational determinants of large-area hurricane evacuations. *Bull. Amer. Meteor. Soc.*, **103**, E642–E656, [Crossref](#).
- Dowling, P., 1997: *The Sociology of Mathematics Education: Mathematical Myths / Pedagogic Texts*. 1st ed. Routledge, 352 pp.
- Drake, L., 2012: Scientific prerequisites to comprehension of the tropical cyclone forecast: Intensity, track, and size. *Wea. Forecasting*, **27**, 462–472, [Crossref](#).
- Ehrlich, M., E. Harville, X. Xiong, P. Buekens, G. Pridjian, and K. Elkind-Hirsch, 2010: Loss of resources and hurricane experience as predictors of postpartum depression among women in southern Louisiana. *J. Women's Health*, **19**, 877–884, [Crossref](#).
- Fagerlin, A., B. J. Zikmund-Fisher, P. A. Ubel, A. Jankovic, H. A. Derry, and D. M. Smith, 2007: Measuring numeracy without a math test: Development of the subjective numeracy scale. *Med. Decis. Making*, 672–680, [Crossref](#).
- Frels, R. K., and A. J. Onwuegbuzie, 2013: Administering quantitative instruments with qualitative interviews: A mixed research approach. *J. Counseling Dev.*, **91**, 184–194, [Crossref](#).
- Ghasemi, A., and S. Zahediasl, 2012: Normality tests for statistical analysis: A guide for non-statisticians. *Int. J. Endocrinol. Metab.*, **10**, 486–489, [Crossref](#).
- Hegarty, M., 2010: Components of Spatial Intelligence. *The Psychology of Learning and Motivation*, Vol. 52, Elsevier, 265–297, [Crossref](#).
- _____, A. E. Richardson, D. R. Montello, K. Lovelace, and I. Subbiah, 2002: Development of a self-report measure of environmental spatial ability. *Intelligence*, **30**, 425–447, [Crossref](#).
- _____, H. Burte, and A. P. Boone, 2018: Individual differences in large-scale spatial abilities and strategies. *Handbook of Behavioral and Cognitive Geography*, D. R. Montello, Ed., Edward Elgar Publishing, 231–246, [Crossref](#).
- Hess, R., V. H. M. Visschers, M. Siegrist, and C. Keller, 2011: How do people perceive graphical risk communication? The role of subjective numeracy. *J. Risk Res.*, **14**, 47–61, [Crossref](#).
- Houston, D., and Coauthors, 2019: The influence of hazard maps and trust of flood controls on coastal flood spatial awareness and risk perception. *Environ. Behav.*, **51**, 347–375, [Crossref](#).
- Kane, P. J., 2014: An investigation into estimation and spatial sense as aspects of workplace numeracy: A case study of recycling and refuse operators within a situated learning model. M. Phil. Thesis, Auckland University of Technology, 208 pp. [Available online at <https://hdl.handle.net/10292/8630>.]
- Lee, J., and R. Bednarz, 2012: Components of spatial thinking: Evidence from a spatial thinking ability test. *J. Geogr.*, **111**, 15–26, [Crossref](#).
- Lenhardt, E. D., R. N. Cross, M. J. Krocak, J. T. Ripberger, S. R. Ernst, C. L. Silva, and H. C. Jenkins-Smith, 2020: How likely is that chance of thunderstorms? A study of how national weather service forecast offices use words of estimative probability and what they mean to the public. *J. Operational Meteor.*, **8** (5), 64–78, [Crossref](#).

- Lipkus, I. M., G. Samsa, and B. K. Rimer, 2001: General performance on a numeracy scale among highly educated samples. *Med. Decis. Making*, **21**, 37–44, [Crossref](#).
- Losee, J. E., K. Z. Naufel, L. Locker, and G. D. Webster, 2017: Weather warning uncertainty: High severity influences judgment bias. *Wea. Climate Soc.*, **9**, 441–454, [Crossref](#).
- Morss, R. E., J. K. Lazo, and J. L. Demuth, 2010: Examining the use of weather forecasts in decision scenarios: Results from a US survey with implications for uncertainty communication. *Meteorol. Appl.*, **17**, 149–162, [Crossref](#).
- _____, J. L. Demuth, J. K. Lazo, K. Dickinson, H. Lazrus, and B. H. Morrow, 2016: Understanding public hurricane evacuation decisions and responses to forecast and warning messages. *Wea. Forecasting*, **31**, 395–417, [Crossref](#).
- Peerbolte, S. L., and M. L. Collins, 2013: Disaster management and the critical thinking skills of local emergency managers: Correlations with age, gender, education, and years in occupation. *Disasters*, **37**, 48–60, [Crossref](#).
- Peters, E., D. Västfjäll, P. Slovic, C. K. Mertz, K. Mazzocco, and S. Dickert, 2006: Numeracy and decision making. *Psychol. Sci.*, **17**, 407–413, [Crossref](#).
- _____, M. K. Tompkins, M. A. Z. Knoll, S. P. Ardoin, B. Shoots-Reinhard, and A. S. Meara, 2019: Despite high objective numeracy, lower numeric confidence relates to worse financial and medical outcomes. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 19 386–19 391, [Crossref](#).
- Ripberger, J., J. Allan, K. Gupta, A. Bitterman, C. Silva, H. Jenkins-Smith, and E. Cokely, 2022a: There's a Chance of What? Assessing Numeracy Skills of Forecasters, Emergency Managers, and Publics to Improve TC Product Uncertainty Communication, IDSS, and Training. Non-series report, The University of Oklahoma Institute for Public Policy Research & Analysis, 24 pp. [Available online at <https://repository.library.noaa.gov/view/noaa/47959>.]
- _____, J. T., A. Bell, A. Fox, A. Forney, W. Livingston, C. Gaddie, C. Silva, and H. Jenkins-Smith, 2022b: Communicating probability information in weather forecasts: Findings and recommendations from a living systematic review of the research literature. *Wea. Climate Soc.*, **14**, 481–498, [Crossref](#).
- Rosen, Z., M. J. Krocak, J. T. Ripberger, R. Cross, E. Lenhardt, C. L. Silva, and H. C. Jenkins-Smith, 2021: Communicating probability information in hurricane forecasts: Assessing statements that forecasters use on social media and implications for public assessments of reliability. *J. Operational Meteor.*, **9** (7), 89–101, [Crossref](#).
- _____, M. Long, A. Schumacher, M. DeMaria, and A. Brammer, 2024: Expert and public preferences and perceptions of different color schemes and map overlays on a tropical cyclone wind exceedance forecast graphic. *Wea. Climate Soc.*, **16**, 497–510, [Crossref](#).
- Roth, R. E., 2009: The impact of user expertise on geographic risk assessment under uncertain conditions. *Cartogr. Geogr. Inf. Sci.*, **36**, 29–43, [Crossref](#).
- Schwartz, L. M., S. Woloshin, W. C. Black, and H. G. Welch, 1997: The role of numeracy in understanding the benefit of screening mammography. *Ann. Intern. Med.*, **127**, 966–972, [Crossref](#).
- Scott, D., 1999: Essential Ingredients for Numeracy. *Australian Primary Mathematics Classroom*, **4**, 4–8, [Crossref](#).
- Taekeuchi, Y., 1992: Sense of direction and its relationship with geographical orientation, personality traits and mental ability. *Jpn. J. Educ. Psychol.*, **40**, 47–53, [Crossref](#).
- Thorndyke, P. W., and S. E. Goldin, 1981: Ability Differences and Cognitive Mapping Skill. Rand Note N-1667-ARMY, 72 pp. [Available online at <https://www.rand.org/content/dam/rand/pubs/notes/2009/N1667.pdf>.]
- Tomaszewski, B., and Coauthors, 2020: Supporting disaster resilience spatial thinking with serious geogames: Project Lily Pad. *ISPRS Int. J. Geo-inf.*, **9**, 405–426, [Crossref](#).
- United States Census Bureau, cited 2021a: QuickFacts: Louisiana. 1–3. [Available online at <https://www.census.gov/quickfacts/LA>.]
- _____, cited 2021b: QuickFacts: Florida. 1–3. [Available online at <https://www.census.gov/quickfacts/FL>.]
- Weisberg, S. M., V. R. Schinazi, N. S. Newcombe, T. F. Shipley, and R. A. Epstein, 2014: Variations in cognitive maps: Understanding individual differences in navigation. *J. Exp. Psychol. Learn. Mem. Cogn.*, **40**, 669–682, [Crossref](#).
- Wernstedt, K., P. S. Roberts, J. Arvai, and K. Redmond, 2019: How emergency managers (mis?)interpret forecasts. *Disasters*, **43**, 88–109, [Crossref](#).
- Zikmund-Fisher, B. J., D. M. Smith, P. A. Ubel, and A. Fagerlin, 2007: Validation of the subjective numeracy scale: Effects of low numeracy on comprehension of risk communications and utility elicitations. *Med. Decis. Making*, **27**, 663–671, [Crossref](#).