Scholar Commons
USF University of South Florida

Numeracy
Advancing Education in Quantitative Literacy

Volume 16 | Issue 1 Article 3

2023

Exploring the Relationship between Quantitative Reasoning Skills and News Habits

Bennett Attaway
Knology, bennetta@knology.org
John Voiklis
Knology, johnv@knology.org
Jena Barchas-Lichtenstein
Knology, jenabl@knology.org
Eric Hochberg
TERC, eric_hochberg@terc.edu
Jim Hammerman
TERC, jim_hammerman@terc.edu

See next page for additional authors

Follow this and additional works at: https://digitalcommons.usf.edu/numeracy

 Part of the Adult and Continuing Education Commons, Journalism Studies Commons, and the Other Social and Behavioral Sciences Commons

Recommended Citation

Attaway, Bennett, John Voiklis, Jena Barchas-Lichtenstein, Eric Hochberg, Jim Hammerman, Uduak Grace Thomas, Nicole LaMarca, Laura Santhanam, and Patti Parson. "Exploring the Relationship between Quantitative Reasoning Skills and News Habits." *Numeracy* 16, Iss. 1 (2023): Article 3. DOI: https://doi.org/10.5038/1936-4660.16.1.1430

Authors retain copyright of their material under a Creative Commons Non-Commercial Attribution 4.0 License.

Exploring the Relationship between Quantitative Reasoning Skills and News Habits

Abstract

Because people are constantly confronted with numbers and mathematical concepts in the news, we have embarked on a project to create journalism that can support news users' number skills. But doing so requires understanding (1) journalists' ability to reason with numbers, (2) other adults' ability to do so, and (3) the attributes and affordances of news. In this paper, we focus on the relationship between adults' news habits and their quantitative reasoning skills. We collected data from a sample of 1,200 US adults, testing their ability to interpret statistical results and asking them to report their news habits. The assessment we developed differentiated the skills of adults in our sample and conformed to the theoretical and statistical assumption that such skills are normally distributed in the population overall. We also found that respondents could be clustered into six distinct groups on the basis of news repertoires (overall patterns of usage, including frequency of news use overall and choice of news outlets). As often assumed in the literature on quantitative reasoning, these news repertoires predicted quantitative reasoning skills better than the amount of quantification in the outlets, but they still predicted only a small fraction of the variance. These results may suggest that news habits may play a smaller or less direct role in quantitative reasoning than has previously been assumed. We speculate that the presence (or absence) of quantification in everyday activities – namely work and hobbies – may be a better predictor of adults' quantitative reasoning, as may additional dimensions of news habits and affective responses to numbers.

Keywords

quantitative reasoning, quantitative literacy, journalism

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 License

Cover Page Footnote

Bennett Attaway is a researcher at Knology who focuses on explaining complex concepts to general audiences. Cognitive and social psychologist John Voiklis leads behaviors research at Knology. Jena Barchas-Lichtenstein is a linguistic anthropologist who leads media research at Knology and is co-PI of Meaningful Math. Eric Hochberg is a senior researcher in the STEM Education Evaluation Center at TERC. Jim Hammerman is a long-time STEM and data science educator and co-directs TERC's STEM Education Evaluation Center. Uduak Grace Thomas was previously communications manager at Knology. She is now an independent science writer and editor. Nicole LaMarca is a project manager and researcher who manages Knology's portfolio of work with *PBS NewsHour*. Laura Santhanam is the health reporter and coordinating producer for polling for the PBS NewsHour. Patti Parson is managing producer of *PBS NewsHour* and PI of Meaningful Math.

Authors

Bennett Attaway, John Voiklis, Jena Barchas-Lichtenstein, Eric Hochberg, Jim Hammerman, Uduak Grace Thomas, Nicole LaMarca, Laura Santhanam, and Patti Parson

This article is available in Numeracy: https://digitalcommons.usf.edu/numeracy/vol16/iss1/art3

Introduction

Whether it's understanding the scope of trillions of dollars in government spending or the significance of a single degree of temperature rise, news users are constantly confronted with numbers and mathematical concepts. Maybe you know someone who pores over sports scores or obsessively refreshes the stock market page. Could the news that people get be related to their ability with numbers? In particular, could it be related to quantitative reasoning (Karaali et al. 2016), the ability to make sense of numbers in context and make decisions on the basis of those numbers? This paper is a methodological foray into questioning these assumptions.

This paper is part of a larger collaborative project co-led by journalists and social scientists, with the goal of helping news users reason with numbers. Creating journalism that can support people's number skills—"better news about math" (Barchas-Lichtenstein et al. 2021)—requires understanding journalists' ability to reason with numbers, other adults' ability to do so, and the attributes and affordances of news. These three requirements are fundamentally interdependent, and interventions must take their interdependence into account.

At least three distinct research traditions have addressed the role of quantitative information, especially statistics, in the news media. Science and technology studies (STS) researchers have proposed approaches that explore how statistics circulate among various types of people, largely focused on describing and identifying reasons for beliefs and behavior (descriptive framing) beyond attributing them to a lack of knowledge on scientific topics (deficit framing). Researchers in quantitative reasoning (QR) and statistical literacy have compared what adults know to what news media expects them to know. Meanwhile, research in journalism studies has focused primarily on journalists' knowledge, practices, and training, and on the quantity and type of statistics reported in news content. All three approaches have shaped our work.

Science and Technology Studies

Science and technology studies (STS) has taken a critical lens to statistics. For example, Porter (1995) traces the history of Western "trust in numbers" to social and political factors, particularly accounting and bureaucracy. In the public health context, Erikson (2012) illustrates how statistics are used to transform contentious policy questions into questions of supposedly apolitical technical expertise. She shows how public health professionals in Sierra Leone can "write an authoritative report and make future recommendations without leaving the hotel or meeting a human being who met her population health criteria" (372), and hospital privatization in Germany has shifted the focus of health statistics from patient wellbeing to customer satisfaction.

Scholars in this vein don't just argue for considering the social life of statistics; they also encourage all of us to begin with empirical description that tests our assumptions. Specifically, Crettaz von Roten and de Roten (2013) noted that claims of collective statistical illiteracy are just-so stories that are not grounded in empirical evidence. This observation informs our first research question:

RQ1: What differences in quantitative reasoning (QR) can be detected among US adults?

Quantitative Reasoning and Statistical Literacy

The terms "numeracy," "quantitative literacy," and "quantitative reasoning" are not defined consistently across frameworks and research groups (Karaali et al. 2016). We follow the example of Vacher (2014), who identified "quantitative reasoning" as primarily concerned with reasoning about and engaging with quantitative information, rather than carrying out numerical operations or working with specific representations of data such as graphs. (See Barchas-Lichtenstein et al. [2021] for further detail on how our team engages with these frameworks.)

Mathematicians, statisticians, and researchers in mathematics and statistics education have analyzed news stories to identify the underlying knowledge such stories assume the public has. For example, John Allen Paulos (1988, 1995) has written several books of critical essays on this topic. Two theorists in this tradition were particularly central to our thinking: both have written extensively about the skills adults need to make sense of statistics and other numbers and take for granted that news is the primary site for this need.

Gal (2002) proposes a framework for breaking statistical literacy into five cognitive elements and two dispositional ones. The cognitive elements include literacy skills, mathematical and statistical knowledge, context knowledge, and critical skills. He further subdivides the required statistical knowledge into five types in increasing order of difficulty. The most basic type is a general sense of the logic of data: why do institutions collect data at all, and through what methods? These skills build up to knowledge of data analysis techniques and procedures as the final type. In contrast, the mathematical knowledge he identifies is surprisingly basic: recognizing that mathematical operations underlie statistical calculations; having at least an informal understanding of percentages, means, and medians; and the conceptual links between summary statistics, graphs, charts, and raw data.

Utts (2003) argues that the general public—including researchers and scientists—misunderstands seven key topics. First, correlation is not causation. Second, a finding can be statistically significant even if it has no real-world importance. Third, the absence of statistical significance may be due to an underpowered study rather than the absence of an effect. Fourth, people need to be aware of potential sources of bias in surveys, such as small changes in question wording and order that can have large effects on survey responses. Fifth, there are

enough people in the world that even extremely unlikely events happen all the time. Sixth, people often confuse conditional probabilities with their inverse—for example, believing that if a test is 90% accurate in detecting a rare condition, someone who tests positive has a 90% chance of having the condition. Seventh, variability is ubiquitous, meaning that averages alone cannot tell you any single observation is typical or "normal."[1]

We built on these frameworks to develop an instrument capable of capturing differences in adults' quantitative reasoning about news across its various manifestations.

Journalism Studies

In general, research on quantification in journalism studies focuses on journalists and news texts (e.g., all papers in Nguyen and Lugo-Ocando [2016]; van Witsen [2018, 2020]). This research relies heavily on several assumptions. First, it assumes a deficit in journalists' own quantitative skills, whether due to a lack of formal requirements in their degree programs (Cusatis and Martin-Kratzer 2009; Martin 2017), a lack of confidence (Maier 2003), or a profession that devalues these skills (Harrison 2016). Second, the research assumes that improving journalists' reasoning will automatically translate to improved public reasoning. For example, Maier (2002) catalogued errors in news stories. While presenting correct information is certainly *necessary* for public reasoning, Maier does not consider whether it is *sufficient*. In fact, this assumption has been disproved repeatedly in research on confirmation bias and debunking misinformation (Nickerson 1998; Seifert 2014; Gorman and Gorman 2021).

Several recent studies have found that different topic areas typically require more or less familiarity with statistics. For example, Cushion et al. (2017) considered 21 different topics covered by the BBC. They found that less than 10% of stories on entertainment, sports, and crime contained statistics, while more than half of stories about the economy (75%), energy (59%), and social policy (54%) did. Sampling from a wide variety of US news sources through Google News, we found that economic and health reporting appeared to require audiences to engage in far more quantitative reasoning than did science and politics reporting (Voiklis et al. 2022).[2]

Given these disparities, we were interested to see how news habits, particularly on the topic level, might correlate with quantitative reasoning, leading to the following research questions:

RQ2: What patterns of news use can people be grouped into?

[1] While Utts arrives at this point last, it is the basis for the other six concepts.

[2] We note that this data was not collected in the lead-up to an election, and that quantification in political news likely varies with the election cycle.

RQ3: To what extent do news habits predict QR, after controlling for individual differences (in education, math anxiety, need for closure)?

Methods

We developed a survey to explore the connection between quantitative reasoning skill, specifically the ability to interpret statistical results, and news habits. The survey was distributed to a sample of 1,200 US adults through the Prolific platform (www.prolific.co; Peer et al. 2017). Distribution through this platform allows polling of a representative sample based on age, gender, and race.[3]

The amount of quantitative content presented in news stories varies between topics and outlets, largely depending on editorial policy. In order to estimate the amount of quantitative content adults had recently encountered in the news, we sampled stories from each outlet we asked about (see *Assessing Quantification in News Outlets* below for details). We hoped that doing so would allow us to assess whether people who rely on more quantitative outlets have stronger reasoning.

We address our analytic methods in each of the three major results sections.

Survey Instrument

The survey contained five content modules. The first module assessed quantitative reasoning in contexts. The second and third modules measured two individual differences related to numbers and statistics: mathematical anxiety and the "need for (cognitive) closure" (Webster and Kruglanski 1994), which is a measure of uncertainty tolerance. We do not assume that these will moderate quantitative reasoning skill, but include them as covariates to account for any effect that may be present. The fourth module collected information about respondents' news habits, and the fifth module included several demographic traits.

News-related quantitative reasoning items. The first module of the survey was a multiple-choice assessment of how people question and interpret statistical content. Questions focused on concepts and inferences rather than calculation. We recognize that existing assessments of quantitative reasoning include calculations based on both graphs and written description, and that many everyday contexts require such calculations (for instance, working with proportions to adjust a recipe, or calculating the percent discount at a store). However, we chose to focus on inferences rather than numerical values, because we anticipate that news users draw conclusions based on what they read but are less likely to perform calculations. The

[3] While online panels are not strictly random, the principles of informed consent assure that no sampling method can ever be fully random.

full text of questions and answer options, as well as percentage of respondents selecting each option, is available in the Appendix.

We adapted several items of varying difficulty from Levels of Conceptual Understanding of Statistics (LOCUS), a well-established and highly reliable instrument designed for students in grades 6–12 (Whitaker et al. 2015). While we mainly chose items focused on interpreting results, we included one involving formulating questions, because thinking critically about quantitative news includes knowing what questions data can answer. LOCUS is based on the Guidelines for Assessment and Instruction in Statistics Education framework, which recognizes that "[e]very high school graduate should be able to use sound statistical reasoning to intelligently cope with the requirements of citizenship, employment, and family" (Franklin et al. 2007, 1). We created five additional questions based on statistician Jessica Utts' (2003) summary of statistical knowledge required in everyday contexts, bringing the total number of assessment questions to ten. The two topics identified by Utts which we did not include were frequency of "unlikely" events and confusion of conditional probabilities with their inverse, because they were uncommon in our review of news content.

Many of the incorrect answers to our quantitative reasoning items were designed to appeal to known heuristics biases. For example, we used language in wrong answers that both deliberately mirrored language in the question prompt and reflected participants' likely personal experiences.

We agree with Oughton (2018) that standardized formal assessments do not reflect everyday numeracy and share her cautions about the use of such assessments. However, we also believe that the questions in this assessment are quite similar to the types of inference that news users are expected to make on a regular basis.

Affective variables. We expected that fear of numbers would make quantitative reasoning more difficult: although none of the assessment questions required calculations, most included specific numbers. To account for any such effect, we measured mathematical anxiety using Numerical Anxiety items from the Brief Version of the long-used and reliable Mathematics Anxiety Rating Scale (Suinn and Winston 2003; first developed by Richardson and Suinn 1972). We reduced the number of items based on the factor loadings in confirmatory factor analysis by Pletzer et. al. (2016) and chose the items most strongly related to Everyday Numerical Anxiety. Cronbach's alpha for this reduced set of items was 0.78, indicating acceptable reliability, and PCA indicated that a single component described 60% of the variance in responses.

Given that, by definition, much quantitative reasoning deals with real-world data, which is uncertain, (dis)comfort with ambiguity may also predict the extent to which audiences fully consider quantitative news content in their judgments and

decision-making. The "need for closure" measures five limiting factors on reasoning and judgment: people's desire for predictability, preference for order and structure, discomfort with ambiguity, decisiveness, and close-mindedness (Roets et al. 2015). We used the well-established 15-item Need for Closure (NFC) scale (Roets and Van Hiel 2011; first developed by Webster and Kruglanski 1994) to measure this construct.

News Practices. The News Practices module of the survey included items on frequency of use of different categories of news (such as Science & Technology, Politics, and Sports) and of news from specific outlets. We were mainly concerned with news users' media repertoires—overall patterns of usage (Hasebrink and Domeyer 2012; Edgerly 2015; Swart et al. 2017)—rather than with their use of individual outlets or topics, but these specific items allowed us to cluster respondents into repertoires based on similar patterns of news use. And while people systematically over-report news habits in surveys (Prior 2009), surveys do give us a picture of norms and ideologies (cf. Gershon 2010). That is, by taking respondents' self-reported news habits as a relative picture, we are still able to gain a sense of news repertoires (compare Edgerly 2015; Dvir-Gvirsman 2020).

In order to investigate potential correlations between quantitative content of news and quantitative reasoning of news users, we needed to be able to sample news stories from the outlets from which respondents were consuming news. To simplify this task, we asked about news use from the 10 most popular news brands (Verto Analytics 2018), as well as the outlet represented by several authors.

Demographics. Finally, we collected demographic data including level of education (which we expected to correlate with both news use choices and QR proficiency), urban/suburban/rural location, gender, and race/ethnicity. We were able to obtain data on participants' ages from the panel provider.

Participant Pool

Participants were roughly evenly split between men and women (583 men, 604 women, 18 non-binary respondents). Because of the way Prolific creates demographic subgroups for race (Prolific 2019), the sample was mostly consistent with US demographics but significantly undersampled people who identified as Hispanic/Latinx (only 69 respondents, or 6% of the total, whereas 18.7% of people in the United States identify as Hispanic/Latinx).

Our respondents were somewhat more educated than the US population as a whole: the median respondent had a bachelor's degree, which is only true of around 32% of US adults 25 years or older (US Census Bureau 2021).

Participants ranged in age from 19 to 91 years old, with a median age of 47.

DOI: https://doi.org/10.5038/1936-4660.16.1.1430

In keeping with national statistics (Parker et al. 2018), around 50% of respondents lived in a suburban area, 30% in an urban area, and 20% in a rural one. Demographic tables are available in Supplement A.

Assessing Quantification in News Outlets

In order to understand the amount of quantitative content present in a "typical" news story from a given outlet, we sampled stories from five dates prior to the end of survey data collection, falling on different days of the week and separated in time for greater variety in topics. The survey was open from March 24–31, 2021; and stories were sampled from Friday, February 26; Monday, March 1; Thursday, March 11; Wednesday, March 17; and Friday, March 26.

We originally planned to look at the 10 most-viewed stories from each outlet from each date, but we only received analytics data from one outlet. Instead, we used the Wayback Machine to access each of the other outlets' homepages as they appeared on the five dates. Many news sites feature a list of "most viewed" stories, and where such a list was available, we included these stories in our sample. For outlets which did not provide this list, we examined the stories which were most prominently placed on the page (for instance, with a larger headline and visible without scrolling). Additional details are available in Supplement B.

One researcher coded the sampled stories based on the types of quantitative reasoning required to fully understand the content presented. The set of five codes used is based on the components of statistical literacy proposed by Gal (2002), namely understanding why data is needed and how it is produced, basic descriptive statistics, figures and tables, probability, and inferential statistics. We build here on the coding system used by Voiklis et al. (2022; also Barchas-Lichtenstein et al. 2022) to classify news stories; we clarified the criteria for assigning some codes and made it explicit that they build off one another. For instance, if a story requires understanding of probability, it is necessary for the reader to also understand basic descriptive statistics such as proportions and averages. The coding scheme is further described in Supplement B.

The single researcher who did the coding was part of the original team who developed the coding scheme and established inter-rater reliability (Voiklis et al. 2022). After establishing inter-rater reliability, it is common practice for a single coder to finish the job.

Authorship

The process by which authorship of collaborative work is determined is infrequently addressed as a methodological issue, but attribution is a critical element of just research practices. How much work is enough to be named as an author? Whose name appears first, and thus receives the lion's share of the credit for a paper typically referenced as (first author) et al.? The Civic Laboratory for

Environmental Action Research (CLEAR) has developed a feminist approach to authorship decisions (Liboiron et al. 2017). Their process has three core features: (1) author order is determined by consensus; (2) they consider care work, administrative work, and other forms of labor that are frequently devalued by scholarly knowledge production; and (3) they account for social position.

Keeping these three tenets in place, we modified this equity protocol as follows: the five Knology authors plus an additional Knology staff member participated in this process, and all other authors agreed with the ordering. Because Knology has historically used alphabetical order as a tiebreaker, and because no group of authors contained more than two people, we used reverse alphabetical order to disrupt this long-standing convention.

Results: Performance on News-related Quantitative Reasoning Tasks

As our quantitative reasoning assessment contained a combination of adapted LOCUS items and original questions based on common statistical misconceptions, it was important to check that the assessment was of appropriate difficulty. We expected the overall scores to be normally distributed and hoped that our original items would be comparable in difficulty to the LOCUS items.

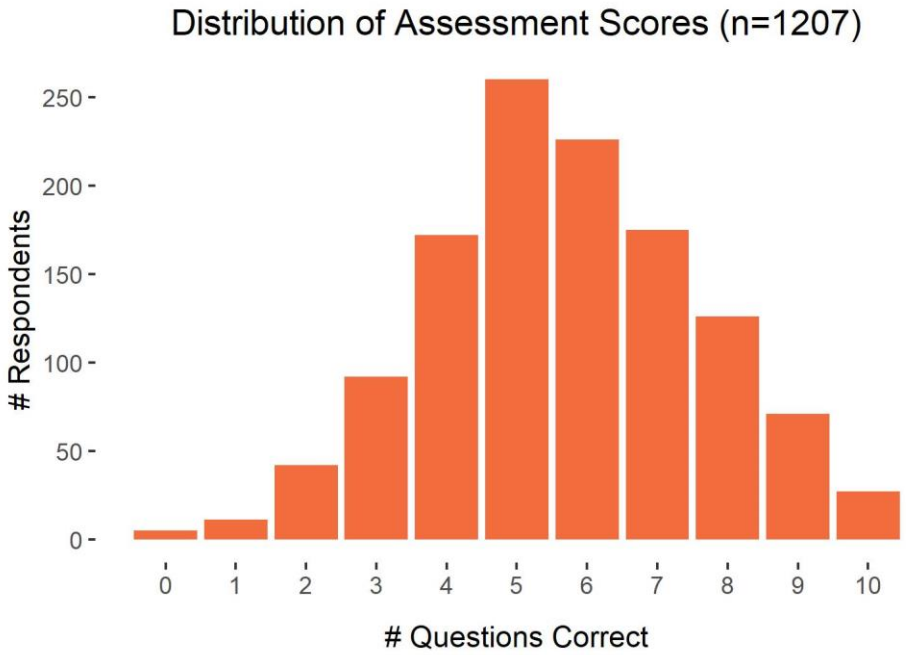

Figure 1. Distribution of assessment scores (n = 1207). This graph represents the number of correct answers on a ten-item multiple-choice assessment (see Appendix).

The average respondent correctly answered 5–6 out of the 10 assessment questions, and visual inspection shows that scores were normally distributed (see Fig. 1). Since the score data is not continuous, to test normality we compared the observed scores to a normal distribution with values rounded to the nearest integer, using a chi-squared test. This test failed to reject the hypothesis that the distributions were the same. The median score was 6 and the mean was 5.61, with a standard deviation of 1.94. We do not observe any floor or ceiling effects.

Assessment questions varied in difficulty. Most (89.7%) respondents answered the easiest question correctly, while only one in four (25.4%) answered the hardest question correctly (see Fig. 2). Items adapted from LOCUS generally matched their assigned LOCUS difficulty (i.e., more people were able to correctly answer the items from lower LOCUS difficulty levels), providing evidence that our adapted items were equivalent to the original items. The difficulty of our original assessment questions generally fell between LOCUS Intermediate and Advanced levels, with the exception of one question with scores comparable to the Basic item. (Full items are available in the Appendix.)

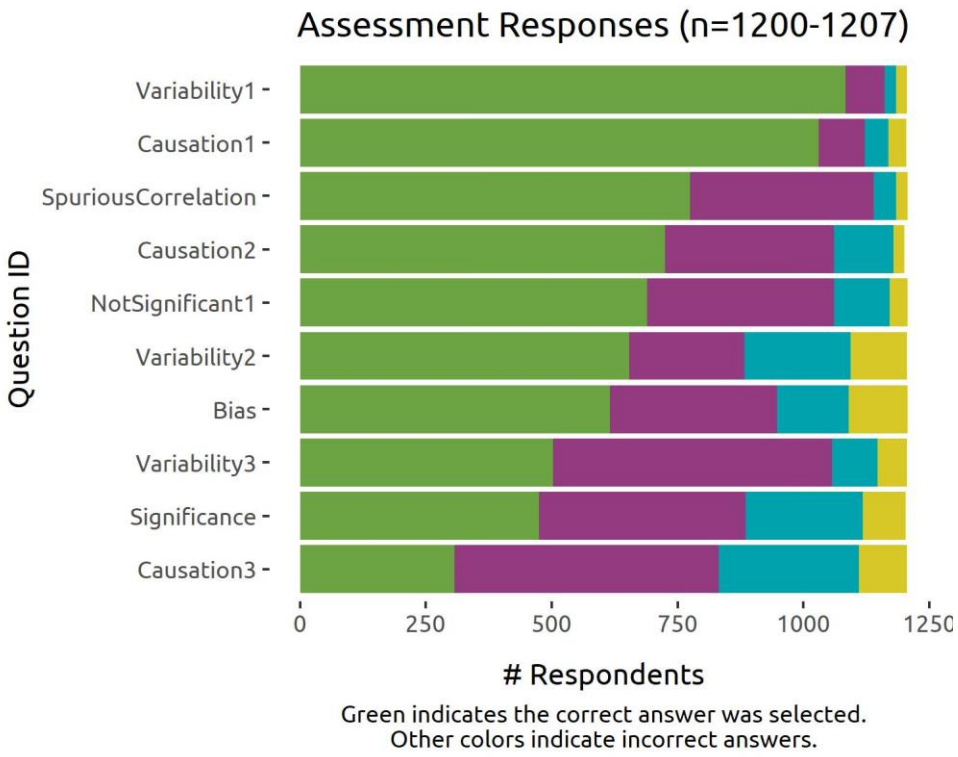

Figure 2. Assessment responses (*n* =1200–1207). Green bars indicate correct responses; purple, blue, and yellow bars indicate "distractor" options in order of how frequently they were selected. (That is, purple always indicates the most popular wrong answer and so on.)

For most questions, one incorrect ("distractor") option accounted for the majority of incorrect answers. Examining these distractors hints at common statistical misconceptions among respondents. For instance, the most commonly chosen distractor for Significance, which required respondents to distinguish between statistical and practical significance of a result with a very small effect size, stated that "[Condition A] is much more relaxing than [Condition B], because the difference was statistically significant." This is precisely the misunderstanding that Utts (2003) identified.

We view QR as a collection of tools for interpreting quantitative content rather than a single skill. For instance, someone who understands that non-random sampling can lead to bias may not also know that an average value alone does not accurately describe a population or understand when it is appropriate to infer causation from data. However, having more of these skills will allow them to reason about a larger proportion of the quantitative content they encounter.

The response data from our assessment shows that, as we expected, correctly answering any specific item is not strongly correlated with responses to others. Principal Components Analysis suggests that the response data could be explained by a model with three underlying factors, but items associated with the same factor in this model do not appear to be conceptually related. We focus our analysis on the overall assessment score and, to account for the relationships between items from different theoretical sources, on the sub-scores for the LOCUS-derived and Utts-derived items rather than attempting to treat QR as a psychometric construct. When examining sub-scores, we exclude one item (Significance) which was not at all correlated with the other Utts items. This item asked respondents to recognize that a very small difference stated to be statistically significant was not "significant" in the colloquial sense of the word. This was the most difficult of the Utts items that we created for the assessment, and less typical of the type of reasoning audiences would typically need to use when reading a news story, as non-specialist news outlets tend not to use the level of detail present in the item to report on scientific research.

Results: News Use Patterns

To solicit respondents' news practices, we asked for the frequency with which they checked news in seven content categories (Sports, Science/Technology, Politics, Health, Entertainment, Education, and Business) and from eleven major outlets. As asking about specific categories for every outlet would have placed high time demands on respondents, we asked about topics and outlets in separate questions. We also expect that many respondents obtain news, and thus encounter quantitative content, from sources not included in the survey.

Most respondents consumed news at least occasionally from each of the categories we asked about (Fig. 3). Politics was the most common category, with around one-third of respondents reporting that they consumed Politics news daily. Sports was the only news category which a large proportion of respondents (35%) said they never checked. Most respondents (71%) reported getting news at least occasionally from one to five of the outlets listed. Breaking down responses by outlet, CNN and the New York Times were the most frequently checked news sources, but each was only used by around half of respondents.

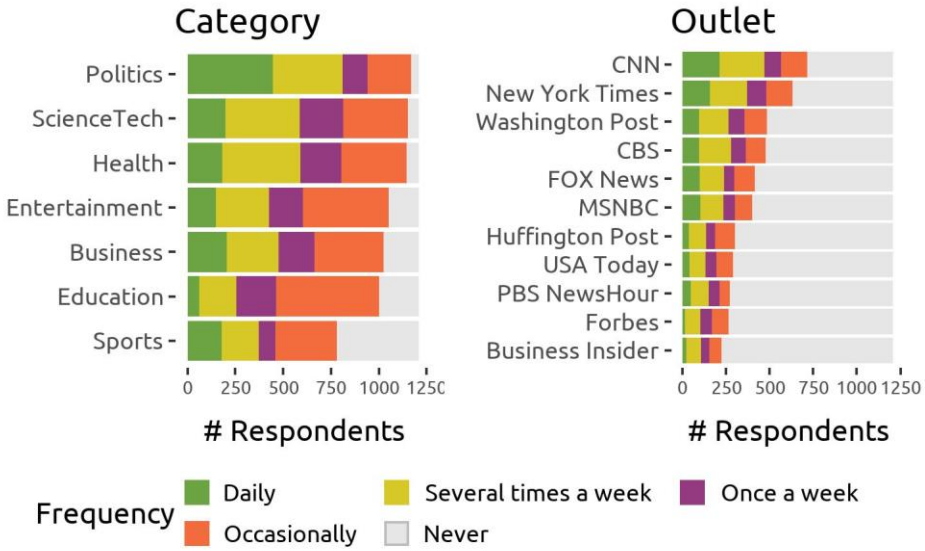

Figure 3. Frequency of news use reported for each topic and outlet (n = 1207). Colored bars indicate the number of respondents selecting each option.

As respondents could identify a separate frequency of use for each of the 18 category and outlet questions, it was necessary to reduce the dimensionality of the response set (that is, to describe the variation in responses using fewer variables) in order to identify patterns. This type of simplification starts with parallel analysis, which involves estimating (through comparison to simulated data) the number of principal components (summary variables) needed to capture variation in the data set, then Principal Components Analysis, which uses these components as a stand-in for the high-dimensional original data. This analysis indicated that patterns of use across news outlets could be described by a single-component model. In other words, people checked news more or less frequently. We tested a higher-dimensional model using three principal components but found that the additional variables loaded only on Sports and Entertainment, respectively, while the first component represented variation across all other topics. Similarly, we performed parallel analysis and PCA on the items about frequency of news consumption from

specific outlets, which indicated these responses could be represented well with a three-component model.

To answer Research Questions 2 and 3, we needed to identify participants with similar news repertoires. To that end, we performed cluster analysis on the responses (see Fig. 4), using a model with three variables to represent news outlets used and one representing use of news across categories (overall news use).[4]

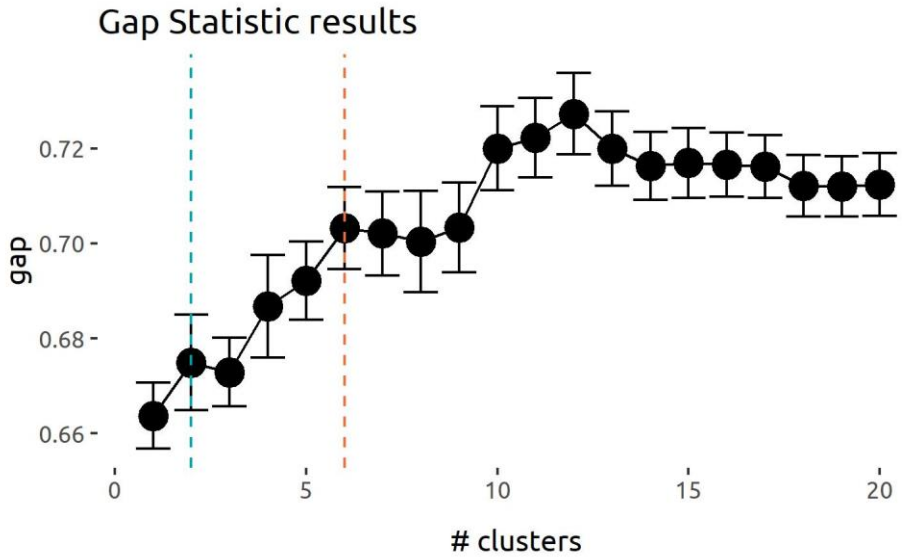

Figure 4. Gap statistic for clustering models with 1 to 20 clusters. Dashed lines indicate values satisfying the gap heuristic for model selection. (n = 12 also satisfies the gap heuristic, but results in a more complex and less interpretable model.)

Cluster analysis involves testing models which separate the data into different numbers of clusters (where distance between items in a cluster and the "center" of that cluster is minimized) and selecting a model which creates clusters that are well-separated and interpretable. We used partitioning around medoids (Kaufman and Rousseeuw 1990), which is more robust than *k*-means clustering, and selected a model based on the gap heuristic (Tibshirani et al. 2001). This technique involves calculating a gap statistic describing the separation between clusters and choosing a value for which increasing the number of clusters does not result in a significantly higher gap. Figure 4 shows the values of the gap statistic for n = 1 to 20 clusters, with dashed lines indicating the first two values satisfying the gap heuristic. As a

[4] Cluster analysis is not a single method but a family of methods. We used Gaussian Mixture Modeling (Fraley and Raftery 2002) as implemented by the *pam* R package. Mixture models assume the presence of sub-populations in the sample and estimate the probability distribution of the data for two or more possible sub-populations (for an accessible description see Harring and Hodis [2016]).

two-cluster model did not create easily interpretable groups of participants while the six-cluster solution did, we continued analysis using six clusters.

Characteristics of Clusters

Because each cluster represents a type of news repertoire, we describe them each here. Where they map fairly clearly to news repertoires described by other studies, we draw explicit parallels. (For example, our "outlet-loyal left" maps fairly closely to Edgerly's (2015) "liberal + online" and our "outlet-loyal right" maps fairly closely to Edgerly's "conservative only" clusters. While we did not ask about political ideology, the news repertoire itself shows considerable similarity.) To illustrate each cluster, we provide data from the "prototypical" respondent, that is, the respondent closest in Mahalanobis distance to the centroid of the cluster, based on the four variables in our model.

Cluster 1: News Grazers (cf. Mitchell et al. 2017). One cluster consisted primarily of respondents who reported never or infrequently getting news from any of the listed outlets but who check news at least weekly across many of the listed topics (see Fig. 5). We acknowledge the possibility that the survey options did not include these respondents' preferred news sources.

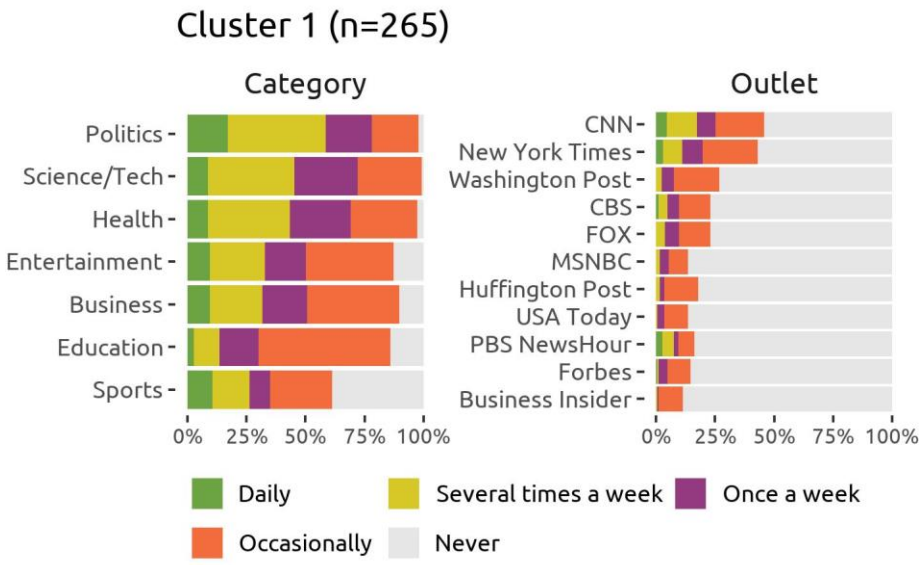

Figure 5. Patterns of news use for News Grazers cluster (*n* = 265).

The respondent closest to the centroid of this cluster reported checking Health news daily, Science/Tech several times a week, and Sports, Politics, and Entertainment once a week. They reported getting news from four outlets (CBS, *New York Times*, *Washington Post*, and *USA Today*), but each only "occasionally."

Cluster 2: the Outlet-loyal Left (cf. Edgerly 2015). The second cluster contained people who tended to get news frequently from CNN, the *New York Times*, and/or the *Washington Post* (as well as, potentially, unlisted news sources). This group was more outlet-loyal than most others; that is, there were several outlets that nearly all of them viewed at least occasionally, while each of the other outlets we mentioned were used by less than half of the members of the cluster. While 96% of members of this cluster checked at least one of CNN, the *New York Times*, and the *Washington Post* once a week or more, 87% reported checking one several times a week or more (see Fig. 6). While these three sources are not the most left-leaning among the options we offered, this group was also the least likely to get news from FOX (only 7% mentioned using it at all, and none more than once a week), even compared to the "news avoiders" cluster (where 23% of respondents used FOX).

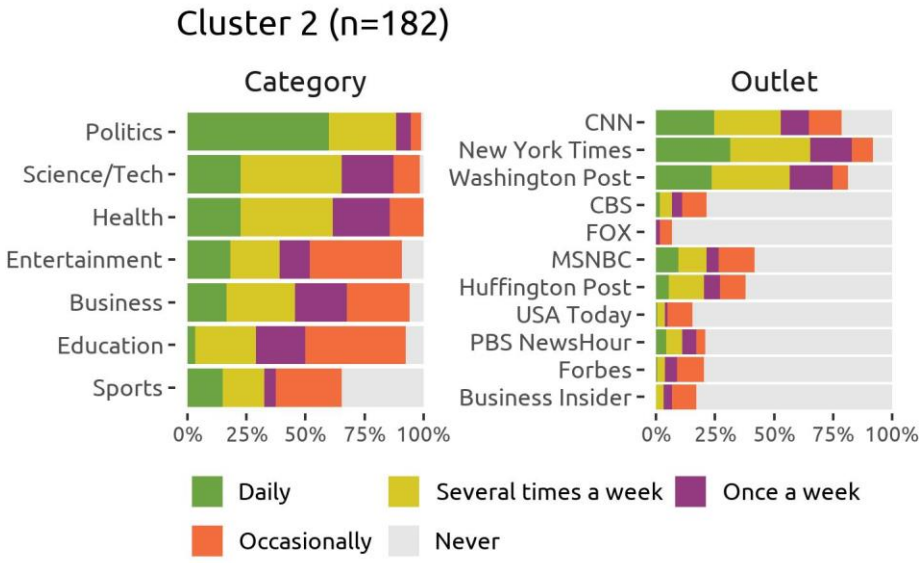

Figure 6. Patterns of news use for Outlet-loyal Left cluster (n = 182).

We label this group the "Outlet-loyal Left" because they roughly parallel Edgerly's (2015) "liberal + online" cluster. Like Edgerly, we found that this cluster was better educated than our sample as a whole. The mean education level for the overall sample fell between associate and bachelor's degrees (3.38, where 3 = associate and 4 = bachelor's). For the group we called the outlet-loyal left, the education level was 3.87, a small deviation from the overall average (Cohen's d = 0.37).

The "prototypical" respondent for this cluster encountered news from all categories at least "occasionally," Health, Sports, and Science/Tech several times a week, and Politics daily. They got news from the *Washington Post* daily, *PBS*

NewsHour several times a week, *Huffington Post* once a week, and CNN and FOX occasionally.

Cluster 3: Omnivores (cf. Edgerly 2015; Dvir-Gvirsman 2020). This group frequently used both video-focused and text-focused news sources, although members tended to use broadcast sources more frequently, particularly CNN, CBS, and MSNBC. 97% of cluster members used one or more of these outlets multiple times per week (see Fig. 7).

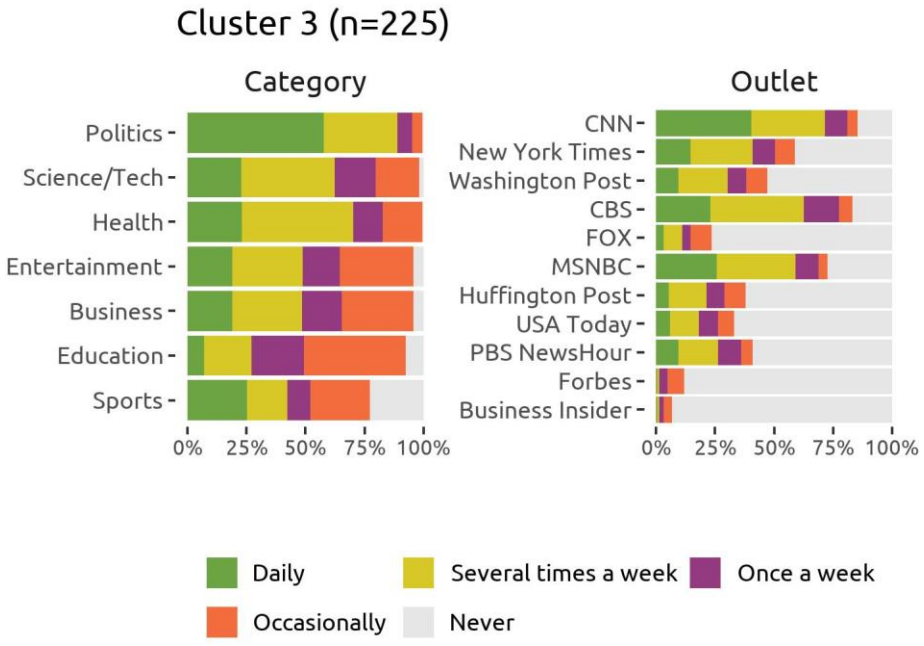

Figure 7. Patterns of news use for Omnivores cluster (n = 225).

The "prototypical" member got Sports, Science/Tech, Politics, and Entertainment news daily, and Health, Education, and Business news occasionally. Out of the provided options, they got news from only CBS, CNN, and the *New York Times*, but accessed news from all three sources daily.

Cluster 4: News Avoiders (cf. Edgerly 2015; Dvir-Gvirsman 2020). Some respondents reported that they never, or only occasionally, consumed news in most of the category areas listed. The "prototypical" member of this cluster consumed Entertainment news daily and Sports, Science/Tech, and Politics occasionally, but out of the listed outlets only got news from CNN and did so only occasionally. Similar to Edgerly (2015), we found that this cluster was less educated than the sample as a whole. The average education level was 2.70 (between "some college"

and an associate degree), a medium deviation from the overall sample average of 3.38 (Cohen's *d* = -0.50), this cluster also contained nearly twice as many women as men, while Edgerly (2015) did not find a significant effect of gender for news avoiders (see Fig. 8).

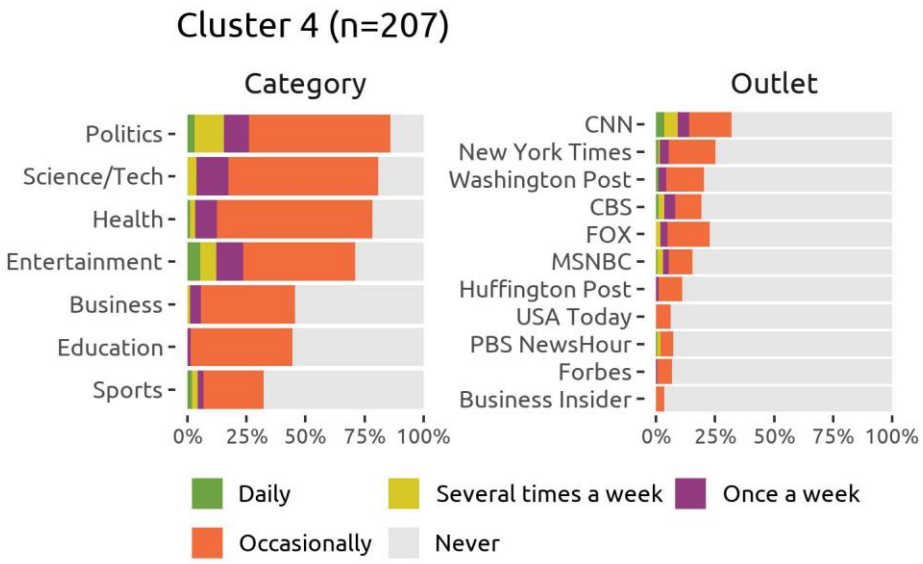

Figure 8. Patterns of news use for News Avoiders cluster (*n* = 207).

Cluster 5: the Outlet-loyal Right (cf. Edgerly 2015). One cluster consisted of respondents who checked FOX News regularly. Most of these respondents (64%) did not report checking any of the print-first outlets listed even occasionally—they likely read news stories on sites or in newspapers not included in our survey. This group was also highly outlet-loyal, with high use of FOX (84% of cluster members used FOX multiple times per week) and relatively low use of most other sources listed.

The "prototypical" respondent for this cluster got news from FOX daily and CNN several times a week. They looked at all news categories except Entertainment at least occasionally, but Politics was the only category they saw daily (see Fig. 9).

Cluster 6: Business-focused news users. The final cluster identified consisted of respondents who checked news frequently across a wide variety of outlets, including *Forbes* and *Business Insider*. This was the only cluster where members checked these two business-oriented outlets regularly. This cluster was also better educated than our sample as a whole. Their average education level was 4.02 (bachelor's degree), a medium deviation from the overall average of 3.38 (Cohen's *d* = 0.54).

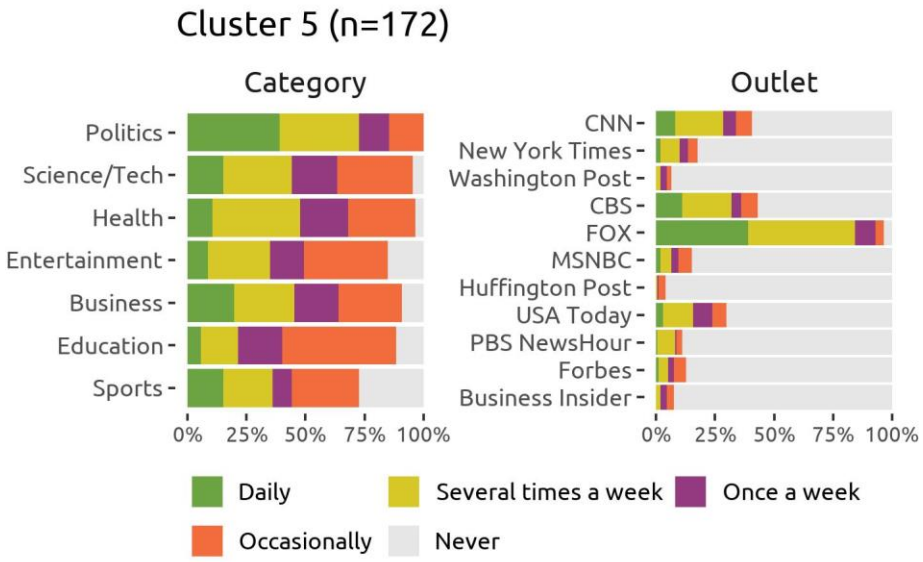

Figure 9. Patterns of news use for the Outlet-loyal Right cluster (*n* = 172).

The prototypical response for this cluster checked news from all categories at least occasionally, and Politics and Business news daily. They got news from CBS, CNN, *Business Insider*, and *Forbes* several times a week, and from *PBS NewsHour*, the *New York Times*, and the *Washington Post* once a week (see Fig. 10).

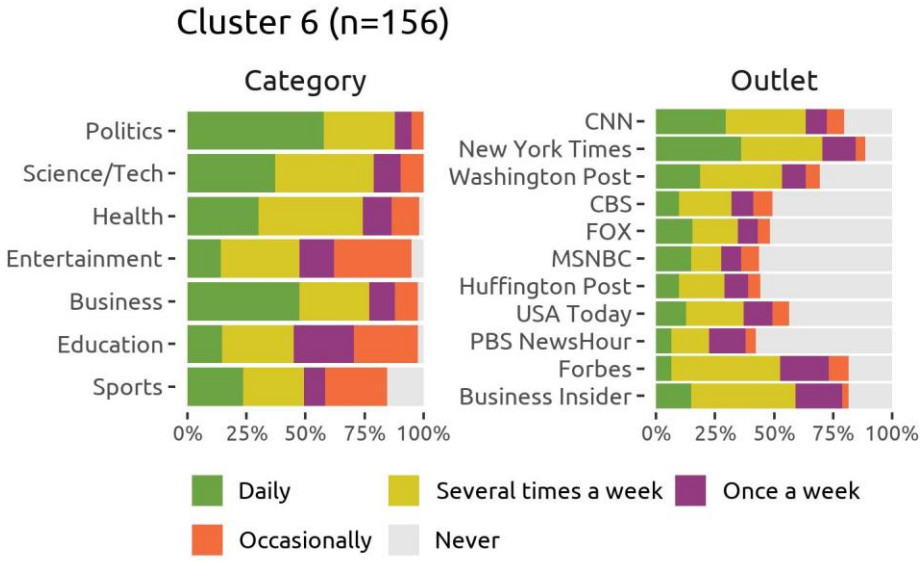

Figure 10. Patterns of news use for the Business-focused cluster (*n* = 156).

Results: Relationship between News Habits and Quantitative Reasoning

People with stronger or weaker quantitative skills might cultivate news habits that match their skill level. It is equally possible that people's news habits might encourage stronger or weaker quantitative skills. In either case, it would be informative to assess the extent to which news habits align with quantitative reasoning skills. The strength of the relationship would justify further exploration to assess causality.

Predictive Value of News Repertoires

We examined a series of linear models predicting scores on the news-related Quantitative Reasoning (QR) assessment based on demographic factors including gender, ethnicity, and education level, as well as normalized scores on the Need for Closure and Math Anxiety scales (Table 1).[5] If ethnicity and gender affect quantitative reasoning, it is likely through a complex set of structural and interpersonal pathways colloquially summarized as "privilege" (see, e.g., Martin et al. 2010). Functionally, we transform these into binary variables—white privilege and male privilege—where the group most likely to benefit from privilege is compared to other groups. The full range of options and number of respondents selecting each option are reported in Supplement A. Given that scores on the LOCUS and Utts-derived sets of questions were moderately correlated, we used a multivariate analysis of co-variance (MANCOVA) model to look at the overall effect of news repertoires on both subscores while accounting for demographic factors and individual differences. As noted in Methods, we have reason to believe these covariates could affect quantitative reasoning, but they are not the primary focus of our research questions. Follow-up ANCOVA models looking at the LOCUS and Utts subscores individually are available in Supplement C.

We found that a model which includes the interaction of news repertoire (i.e., the clusters described above) with education gives the best fit, as measured by residual sum of squares. Repertoire had the largest effect of all covariates (partial $\eta^2 = 0.02$). Partial η^2 is a measure of effect size typically used for (M)AN(C)OVA models and represents the percent of total variance explained by an independent variable. A value of 0.01 is considered a small effect, 0.06 a medium effect, and 0.14 a large effect (Maher et al. 2013).

As expected, higher normed scores on Need for Closure and Math Anxiety were associated with lower scores on the assessment. Also, a higher education level (represented as an ordered factor ranging from "high school or less" to "doctorate,

[5] PCA was used to return scores that approximate a *z*-scale, mean approximately 0, and standard deviation approximately 1.

Table 1
Analysis of Co-variance Model Showing Effect of Variables on QR Scores.

	Partial η^2	df	F	p
Male Privilege	0.002	2	1.398	0.074
White Privilege	0.017	2	10.598	<0.001
Education	0.018	2	11.043	<0.001
Math Anxiety	0.016	2	9.711	<0.001
Need for Closure	0.004	2	2.103	0.045
Repertoire (Cluster)	0.020	10	4.972	<0.001
Education x Repertoire	0.013	10	3.063	0.004
Residuals	0.909	1189		

law, or medical degree") was associated with higher scores. However, all these effects were small. In total, this model explained approximately 9% of variance in QR scores.

As this initial analysis suggested that there were differences between QR scores of certain repertoire groups, we conducted post-hoc testing using the *multcomp* R package (Bretz et al. 2010), which incorporates adjustments for multiple comparisons. We found the following statistically significant differences between clusters, accounting for covariates:

- Looking at overall QR scores:
  - The Outlet-loyal Left cluster scored higher (covariate-adjusted mean of 6.36) than any other cluster.
  - The Business-focused (5.76) and Grazers (5.79) clusters scored higher than the Outlet-loyal Right cluster (5.15).
- On the LOCUS subscale:
  - The Outlet-loyal Left cluster (covariate-adjusted mean of 3.18) outperformed all other clusters except Grazers.
  - The Grazers cluster (2.86) outperformed the Outlet-loyal Right cluster (2.47).
- On the Utts subscale:
  - The Outlet-loyal Left (covariate-adjusted mean of 3.18) and Business-focused (3.05) outperformed the Outlet-loyal Right (2.68) cluster.

While not all pairs of clusters had statistically significant differences, the overall pattern in scores was that the Outlet-loyal Left cluster scored highest, followed by the Grazers and Business-focused clusters, then Omnivores and Avoiders, with the Outlet-loyal Right cluster scoring lowest. A similar pattern was present when looking at LOCUS or Utts subsets of questions. Figure 11 illustrates the covariate-adjusted means for each cluster for overall score across both subsets. We provide this information as a graph rather than a table to make confidence intervals more readily visible.

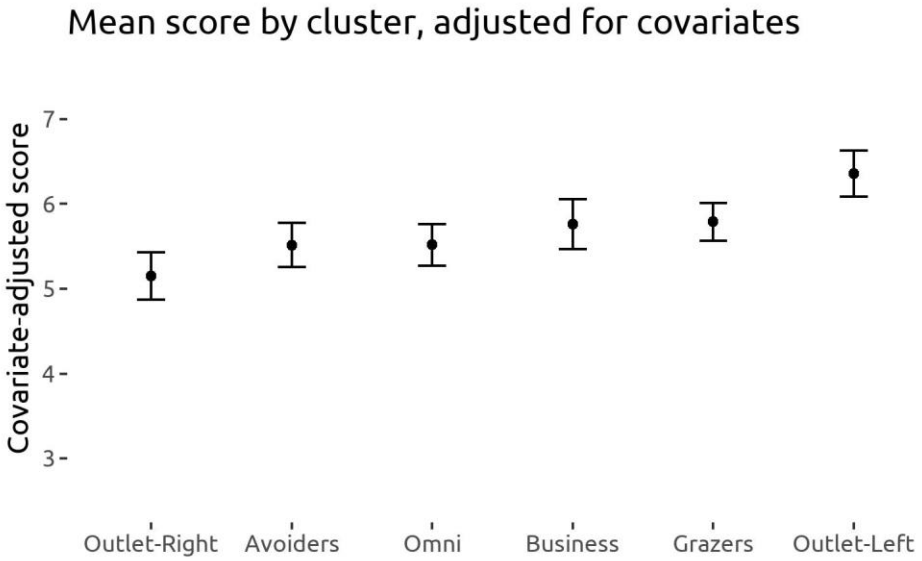

Figure 11. Mean score by cluster, adjusted for covariates. The Outlet-loyal Left cluster had significantly higher scores than all other clusters, and the Grazers and Business-focused clusters scored significantly higher than the Outlet-loyal Right.

Predictive Value of Amount of Quantification in Outlets

We assigned quantification scores to outlets based on the average number of QR content codes (Data Why and How, Descriptive Statistics, Graphics and Tables, Probability, and Inferential Statistics) assigned to sampled stories from the outlet. As news stories could receive all, some, or none of these codes, the score for each story could fall between 0 and 5. Mean scores for each outlet varied from 1.06 (FOX) to 2.48 (*New York Times*). Notably, FOX, *Huffington Post,* and MSNBC were on average less QR-heavy than other outlets. The mean score for each outlet is available in Supplement B.

We then weighted the quantification score for each outlet with respondents' reported frequency of use for the outlet to generate an individual score for each respondent, indicating the extent to which they would encounter the need for QR during news use. The first weighting scheme we tested assigned outlet weights proportional to the approximate number of times in a week someone reported using the outlet (e.g., 0.5 for "occasionally," 7 for "daily"). In this model, an individual's "QR Use" score did not predict QR assessment scores, and its interaction with Education had a near-zero explanatory effect (0.1% of total variance).

Taking into account the conclusion of Prior (2009) that self-reports of news use result in overestimates, we also tested a model where we weighted the news use frequency responses as simply levels from "occasionally" (1) to "daily" (4). In this model, an individual's "QR Use" score explained only 0.4% of total QR assessment

score variance. The interaction term with Education explained only 0.7% of variance. MANCOVA tables for both models are available in Supplement C.

Put simply, the amount of quantification in news outlets that individuals use was not particularly useful for predicting their QR assessment scores. We note several limitations of our approach: even if respondents accurately reported their frequency of use for a given news outlet, this would not tell us how much news they look at each time. For instance, someone might watch a 60-minute news broadcast once a week, while another person might scan the headlines on an outlet's website every day but only occasionally click through to read a full story. Similarly, people who use a given outlet may do so for specific types of coverage—perhaps someone reads *USA Today* only for entertainment news, while checking a different outlet for political coverage. In this case, they would encounter very light QR demands from their use of *USA Today*, which would not be accounted for in our analysis. Asking about use frequency for each combination of category and outlet would have placed undue burden on survey respondents, and potentially introduce more error due to inaccurate estimates of news use.

Discussion and Conclusions

This study aimed to test the common assumption that people's news habits and repertoires are related to their ability with numbers. We did so as part of a collaboration between journalists and social scientists to support this reasoning through the news. We found that the relationship may not be as direct as previously assumed.

RQ1: What differences in quantitative reasoning (QR) can be detected among adults?

The results on our 10-item quantitative reasoning questionnaire successfully differentiate adults' ability to interpret and reason with the types of numbers that appear frequently in news stories. We caution that participants in our study were, on average, more educated than the overall US population, so the distribution of scores may not represent the news-related QR skills of US adults as a whole.

RQ2: What patterns of news use can people be grouped into?

Based on a relatively brief set of items on news topics and outlets, this study found that adults' news repertoires can be categorized into six groups, which we label based on their most salient features: News Grazers, the Outlet-loyal Left, Omnivores, News Avoiders, the Outlet-loyal Right, and Business-focused news users. These repertoires are similar to those identified in studies using different methods and different questions (e.g., Edgerly 2015; Dvir-Gvirsman 2020).

However, in order to more fruitfully compare the news habits of audiences across different contexts, we see a need for future research that can tie together the many strands of research on media repertoires. Scholars have segmented repertoires

along a range of criteria, including outlets (present study), topics (present study), medium (Edgerly 2015; Swart et al. 2017), ideological orientation (Edgerly 2015), social media sharing behaviors (Dvir-Gvirsman 2020), geographic focus (Swart et al. 2017), reasons to use news (Swart et al. 2017), and reasons to consider news important (Swart et al. 2017). We hope researchers interested in news repertoires will continue to refine the concept and identify the criteria that matter most broadly. We also encourage researchers in this vein to recognize that "news" may not be a natural category for users (cf. Armstrong et al. 2015), and to think about information sourcing more broadly.

RQ3: To what extent do news habits predict QR, after controlling for individual differences (in education, math anxiety, need for closure)?

The segmentation described above (i.e., news repertoire) was the variable we considered that explained the largest amount of variance in QR. However, it still explained only about 2% of variance, with 7% of additional variance explained by our model. Four variables (the interaction between education and news repertoire, math anxiety, education by itself, and whiteness) all contributed to the model, while need for closure—often conceptualized as discomfort with uncertainty—contributed very little. Furthermore, knowing how much quantification was available in different outlets explained less variability than the repertoires alone.

There are two potential reasons for this result. Perhaps a different way of operationalizing repertoires or quantification in news sources would identify far more variance in QR scores. The literature on news repertoires has not previously considered relationships to quantitative content and quantitative reasoning, but it may also simply be the case that exposure to quantification in news sources is far less critical to adults' QR than the field has long assumed.

We speculate that at least three other factors may affect adults' QR. First, the presence (or absence) of quantification in everyday activities may be a better predictor of adults' quantitative reasoning. Many jobs that do not require high levels of formal education still have high QR demands. For example, grocers need to estimate how much milk to order based on estimated demand and supply chain factors, and farmers need to plan what to plant when based on current and projected weather conditions, as well as estimated demand for various agricultural products. Everyday activities also include hobbies: for example, we would expect that fans of tabletop role-playing games should be well-acquainted with basic probability.

Additional characteristics of news habits may also be germane. Who do people discuss news with? Those conversations play a considerable role in sense-making, and we simply do not know how they impact QR. How deeply people engage with news statistics and graphics is another variable. Two authors of this paper are currently conducting a series of focus groups with adults across the QR spectrum. We have consistently seen that, given enough time, adults can identify the central argument of even highly complex graphics. However, participants have also

consistently said that they would spend far less time considering images if they encountered them on their own.

Finally, there may be an affective dimension to quantitative reasoning that deserves further consideration. Beyond anxiety, adults may have a range of affective responses toward math and quantification, and those who find math difficult may equally react with skepticism or with uncritical trust in numbers. Perceived relevance and strength of preconceptions about a particular topic may also influence quantitative reasoning in keeping with research on confirmation bias.

While adults are frequently exposed to quantitative content in news, this study calls into question whether that exposure affects their QR skills. News providers should be aware that their audiences span a wide range of quantitative ability, even if they target a specific level. We hope to see further exploration of the factors associated with QR skills in adults, and how non-education-focused activities such as engagement with news can best support the development of these skills.

Acknowledgment

We thank the many colleagues who helped us think through this work, and in many cases read earlier drafts of the paper. These include project advisors Jim Corter, Danny Bernard Martin, Caitlin Petre, Jonathan Stray, Nikki Usher, and Darryl Yong; John Fraser at Knology; Travis Daub, Vanessa Dennis, Molly Finnegan, Isabella Isaacs-Thomas, Erica Hendry, Chloe Jones, Megan McGrew, Miles O'Brien, and James Williams at *PBS NewsHour*; and two anonymous reviewers. We also thank members of the News Nerds group on Slack for suggestions on sampling stories from news outlets, and Tim Jacobbe for discussing potential adaptations of LOCUS with us early in the process.

This work was supported by the National Science Foundation under Grant No. DRL-1906802. 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 NSF.

References

Armstrong, Cory L., Melinda J. McAdams, and Jason Cain. 2015. "What Is News? Audiences May Have Their Own Ideas." *Atlantic Journal of Communication* 23 (2): 81–98. https://doi.org/10.1080/15456870.2015.1013102

Bretz, Frank, Thorsten Hothorn, and Peter Westfall. 2010. *Multiple Comparisons Using R.* Boca Raton, Florida, USA: Chapman & Hall/CRC Press.

Barchas-Lichtenstein, Jena, John Voiklis, Laura Santhanam, Nsikan Akpan, Shivani Ishwar, Bennett Attaway, Patti Parson, & John Fraser. 2021. "Better

News About Math: A Research Agenda." *Numeracy* 14 (1): Article 4. https://doi.org/10.5038/1936-4660.14.1.1377

Barchas-Lichtenstein, Jena, John Voiklis, Bennett Attaway, Laura Santhanam, Patti Parson, Uduak Grace Thomas, Isabella Isaacs-Thomas, Shivani Ishwar, & John Fraser. 2022. "Number Soup: Case Studies of Quantitatively Dense News." *Journalism Practice.* https://doi.org/10.1080/17512786.2022.2099954

Crettaz von Roten, Fabienne, and Yves de Roten. 2013. "Statistics in Science and in Society: From a State-of-the-Art to a New Research Agenda." *Public Understanding of Science* 22 (7): 768–784. https://doi.org/10.1177/0963662513495769

Cusatis, Christine, and Renee Martin-Kratzer. 2009 "Assessing the State of Math Education in ACEJMC-accredited and Non-accredited Undergraduate Journalism Programs." *Journalism & Mass Communication Educator* 64 (4): 355–377. https://doi.org/10.1177/107769580906400402

Cushion, Stephen, Justin Lewis, and Robert Callaghan. 2017. "Data Journalism, Impartiality and Statistical Claims." *Journalism Practice* 11 (10): 1198–1215. https://doi.org/10.1080/17512786.2016.1256789

Dvir-Gvirsman, Shira. 2020. "Understanding News Engagement on Social Media: A Media Repertoire Approach." *New Media & Society*: 1461444820961349. https://doi.org/10.1177/1461444820961349

Edgerly, Stephanie. 2015. "Red Media, Blue Media, and Purple Media: News Repertoires in the Colorful Media Landscape." *Journal of Broadcasting & Electronic Media* 59 (1): 1–21. https://doi.org/10.1080/08838151.2014.998220

Erikson, Susan L. 2012. "Global Health Business: The Production and Performativity of Statistics in Sierra Leone and Germany." *Medical Anthropology* 31 (4): 367–384. https://doi.org/10.1080/01459740.2011.621908

Fraley, Chris, and Adrian E. Raftery. 2002. "Model-based Clustering, Discriminant Analysis, and Density Estimation." *Journal of the American Statistical Association* 97 (458): 611–631. https://doi.org/10.1198/016214502760047131

Franklin, Christine, Gary Kader, Denise Mewborn, Jerry Moreno, Roxy Peck, Mike Perry, and Richard Scheaffer. 2007. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.

Gal, Iddo. 2002. "Adults' Statistical Literacy: Meanings, Components, Responsibilities." *International Statistical Review* 70 (1): 1–25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gershon, Ilana. 2010. "Media Ideologies: An Introduction." *Journal of Linguistic Anthropology* 20 (2): 283–293. https://doi.org/10.1111/j.1548-1395.2010.01070.x

Gorman, Sara E., and Jack M. Gorman. 2021. *Denying to the Grave: Why We Ignore the Science that Will Save Us.* Oxford University Press. https://doi.org/10.1093/oso/9780197547458.001.0001

Harring, Jeffrey R., and Flaviu A. Hodis. 2016. "Mixture Modeling: Applications in Educational Psychology." *Educational Psychologist* 51 (3–4): 354–367. https://doi.org/10.1080/00461520.2016.1207176

Harrison, Steven. 2016. "Journalists, Numeracy and Cultural Capital." *Numeracy* 9 (2): Article 3. https://doi.org/10.5038/1936-4660.9.2.3

Hasebrink, Uwe, and Hanna Domeyer. 2012. "Media Repertoires as Patterns of Behaviour and as Meaningful Practices: A Multimethod Approach to Media Use in Converging Media Environments." *Participations* 9 (2): 757–779.

Karaali, Gizem, Edwin H. Villafane Hernandez, and Jeremy A. Taylor. 2016. "What's in a Name? A Critical Review of Definitions of Quantitative Literacy, Numeracy, and Quantitative Reasoning." *Numeracy* 9 (1): Article 2. https://doi.org/10.5038/1936-4660.9.1.2

Kaufman, Leonard, and Peter J. Rousseeuw. 1990. "Partitioning around Medoids (program pam)." *Finding Groups in Data: An Introduction to Cluster Analysis* 344: 68–125.

Liboiron, Max, Justine Ammendolia, Katharine Winsor, Alex Zahara, Hillary Bradshaw, Jessica Melvin, Charles Mather, Natalya Dawe, Emily Wells, France Liboiron, Bojan Fürst, Coco Coyle, Jacquelyn Saturno, Melissa Novacefski, Sam Westscott, and Grandmother Liboiron. 2017. "Equity in Author Order: A Feminist Laboratory's Approach." *Catalyst: Feminism, Theory, Technoscience* 3 (2): 1–17. https://doi.org/10.28968/cftt.v3i2.28850

Maher, Jessica M., Jonathan C. Markey, and Diane Ebert-May. "The Other Half of the Story: Effect Size Analysis in Quantitative Research." *CBE - Life Sciences Education* 12 (3): 345–351. https://doi.org/10.1187/cbe.13-04-0082

Maier, Scott R. 2002. "Numbers in the News: A Mathematics Audit of a Daily Newspaper." *Journalism Studies* 3 (4): 507–519. https://doi.org/10.1080/1461670022000019191

Maier, Scott R. 2003. "Numeracy in the Newsroom: A Case Study of Mathematical Competence and Confidence." *Journalism & Mass Communication Quarterly* 80 (4): 921–936. https://doi.org/10.1177/107769900308000411

Martin, Daniel B., Maisie L. Gholson, and Jacqueline Leonard. 2010. "Mathematics as Gatekeeper: Power and Privilege in the Production of Knowledge." *Journal of Urban Mathematics Education* 3 (2): 12–24.

Martin, Justin D. 2017. "A Census of Statistics Requirements at US Journalism Programs and a Model for a 'Statistics for Journalism' Course." *Journalism & Mass Communication Educator* 72 (4): 461–479. https://doi.org/10.1177/1077695816679054

Mitchell, Amy, Jeffrey Gottfried, Elisa Shearer, and Kristine Lu. 2017. "How Americans Encounter, Recall and Act upon Digital News." Accessed November 22, 2021. https://www.pewresearch.org/journalism/2017/02/09/part-i-an-analysis-of-individuals-online-news-habits-over-the-course-of-one-week/

Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2 (2): 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nguyen, An, and Jairo Lugo-Ocando, eds. 2016. "Special Issue: Data and Statistics in Journalism and Journalism Education." *Journalism* 17 (1). https://doi.org/10.1177/1464884915593234

Oughton, Helen. 2018. "Disrupting Dominant Discourses: A (Re) Introduction to Social Practice Theories of Adult Numeracy." *Numeracy* 11 (1): Article 2. https://doi.org/10.5038/1936-4660.11.1.2

Parker, Kim, Juliana M. Horowitz, Anna Brown, Richard Fry, D'Vera Cohn, and Ruth Igielnik. 2018. "Demographic and Economic Trends in Urban, Suburban and Rural Communities." Report, Pew Research Center. https://www.pewresearch.org/social-trends/2018/05/22/demographic-and-economic-trends-in-urban-suburban-and-rural-communities/ (accessed November 22, 2021).

Paulos, John Allen. 1988. *Innumeracy: Mathematical Illiteracy and its Consequences*. Macmillan.

Paulos, John Allen. 1995. *A Mathematician Reads the Newspaper*. Anchor Books.

Pletzer, Belinda, Guilherme Wood, Thomas Scherndl, Hubert H. Kerschbaum, and Hans-Christoph Nuerk. 2016. "Components of Mathematics Anxiety: Factor Modeling of the MARS30-Brief." *Frontiers in Psychology* 7: 91. https://doi.org/10.3389/fpsyg.2016.00091

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research." *Journal of Experimental Social Psychology* 70: 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9780691210544

Prior, Markus. 2009. "The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure." *Public Opinion Quarterly* 73 (1): 130–143. https://doi.org/10.1093/poq/nfp002

Prolific. 2019. "Representative Samples FAQ." Accessed June 21, 2021. https://researcher-help.prolific.co/hc/en-gb/articles/360019238413

Richardson, Frank C., and Richard M. Suinn. 1972. "The Mathematics Anxiety Rating Scale: Psychometric Data." *Journal of Counseling Psychology* 19 (6): 551. https://doi.org/10.1037/h0033456

Roets, Arne, and Alain Van Hiel. 2011. "Item Selection and Validation of a Brief, 15-item Version of the Need for Closure Scale." *Personality and Individual Differences* 50 (1): 90–94. https://doi.org/10.1016/j.paid.2010.09.004

Roets, Arne, Arie W. Kruglanski, Malgorzata Kossowska, Antonio Pierro, and Ying-yi Hong. 2015. "The Motivated Gatekeeper of Our Minds: New Directions in Need for Closure Theory and Research." In *Advances in Experimental Social Psychology, vol. 52*, edited by Mark Zanna and James Olson, 221–283. Academic Press. https://doi.org/10.1016/bs.aesp.2015.01.001

Seifert, Colleen M. 2014. "The Continued Influence Effect: The Persistence of Misinformation in Memory and Reasoning Following Correction." *Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences*: 39–71.

Suinn, Richard M., and Elizabeth H. Winston. 2003. "The Mathematics Anxiety Rating Scale, a Brief Version: Psychometric Data." *Psychological Reports* 92 (1): 167–173. https://doi.org/10.2466/pr0.2003.92.1.167

Swart, Joëlle, Chris Peters, and Marcel Broersma. 2017. "The Ongoing Relevance of Local Journalism and Public Broadcasters: Motivations for News Repertoires in the Netherlands." *Participations: Journal of Audience & Reception Studies* 14 (2): 268–282.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. "Estimating the Number of Clusters in a Data Set via the Gap Statistic." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2): 411–423. https://doi.org/10.1111/1467-9868.00293

US Census Bureau. 2021. "Bachelor's Degree Attainment in the United States: 2005 to 2019. American Community Survey Briefs. ACSBR-009." https://www.census.gov/content/dam/Census/library/publications/2021/acs/acsbr-009.pdf (accessed November 22, 2021).

Utts, Jessica. 2003. "What Educated Citizens Should Know about Statistics and Probability." *The American Statistician* 57 (2): 74–79. https://doi.org/10.1198/0003130031630

Vacher, H. L. 2014. "Looking at the Multiple Meanings of Numeracy, Quantitative Literacy, and Quantitative Reasoning." Numeracy 7 (2): Article 1. https://doi.org/10.5038/1936-4660.7.2.1

Van Witsen, Anthony. 2018. "How Journalists Establish Trust in Numbers and Statistics: Results from an Exploratory Study." *Iowa State University*

Summer Symposium on Science Communication. https://doi.org/10.31274/sciencecommunication-181114-8

Van Witsen, Anthony. 2020. "How Daily Journalists Use Numbers and Statistics: The Case of Global Average Temperature." *Journalism Practice* 14 (9): 1047–1065. https://doi.org/10.1080/17512786.2019.1682944Verto Analytics. 2018. "Most Popular News Brands in the United States as of June 2018, by Monthly Users." *Statista - The Statistics Portal*. https://www.statista.com/statistics/875811/most-popular-us-news-brands-ranked-by-audience/ (accessed June 28, 2021).

Voiklis, John, Jena Barchas-Lichtenstein, Bennett Attaway, Uduak G. Thomas, and Shivani Ishwar. 2022. "Surveying the Landscape of Numbers in US News." *Numeracy* 15 (1): Article 2. https://doi.org/10.5038/1936-4660.15.1.1406

Webster, Donna M., and Arie W. Kruglanski. 1994. "Individual Differences in Need for Cognitive Closure." *Journal of Personality and Social Psychology* 67 (6): 1049. https://doi.org/10.1037/0022-3514.67.6.1049

Whitaker, Douglas, Steven Foti, and Tim Jacobbe. 2015. "The Levels of Conceptual Understanding in Statistics (LOCUS) Project: Results of the Pilot Study." *Numeracy* 8 (2): Article 3. https://doi.org/10.5038/1936-4660.8.2.3

Appendix: Quantitative Reasoning Items

Participants saw the following items, and the answer choices for each item, in random order. We have ordered them here from easiest to most difficult, as measured by the number of correct responses. The correct answer is highlighted.

Variability1, Based on Utts' "Variability Is Normal"

Please read about a research study and answer the question by choosing the option that makes sense to you.

Many research studies have found that smoking cigarettes causes lung cancer. However, Chris has an uncle who smoked from age 16 to age 98 and died peacefully in his sleep. Which of the following is true?

Answer Text	% Selected
Research studies cannot tell us very much about specific individuals.	90%
Chris is less likely to get lung cancer than other smokers.	6%
Those research studies cannot be trusted.	2%
Chris is more likely to get lung cancer than other smokers.	2%

Causation1, LOCUS Basic Item

Please read about a research study and answer the question by choosing the option that makes sense to you.

A recent study on habits of adolescents and younger adults reported that adolescent boys typically watch 2 hours more TV per week than girls, and younger adult men typically watch 4 hours more TV per week than younger adult women. Based on these data, can we conclude that middle-aged men typically watch 6 hours more TV per week than middle-aged women?

Answer Text	% Selected
No, because data were not collected from middle-aged adults.	86%
Yes, it is reasonable to predict TV habits of middle-aged adults from younger adults.	8%
Yes, because 6 hours is the next number in the pattern.	4%
No, because middle-aged adults can afford other entertainment and don't watch as much TV.	3%

SpuriousCorrelation, Based on Utts' Correlation/Causation

Please read about a research study and answer the question by choosing the option that makes sense to you.

School officials are considering methods to prevent bullying. Researchers found that bullying incidents often spike on days when students wear their school colors. There are several occasions when students wear their school colors including field trips, school assemblies, fund raising drives, game days with rival schools, among others. The principal bans the wearing of school colors, but

researchers found that the spikes in bullying incidents continue. Which of the following do you think is the most likely reason the ban failed to prevent bullying?

Answer Text	% Selected
The spikes in bullying are likely because of the occasions for wearing school colors rather than simply wearing the colors.	64%
Students can always find something else to bully each other about.	30%
The students resent the banning of school colors and, so, continue to bully each other.	4%
The bullies likely come from the rival schools.	2%

Causation2, LOCUS Intermediate Item

Please read about a research study and answer the question by choosing the option that makes sense to you.

A study was conducted to investigate whether washing with soap and water or using hand sanitizer removes more bacteria from a person's hands. Volunteers were recruited from a community center and randomly assigned to a group that washed their hands with soap and water or to a group that used hand sanitizer. When they were finished, each volunteer pressed his or her hands into specially prepared petri dishes. After several days, the number of bacteria colonies was counted on each petri dish. Which of the following statements best describes the random assignment in this study?

Answer Text	% Selected
The random assignment was important because it tends to create groups that are similar with respect to other variables that might affect bacteria growth.	60%
The random assignment was important so that these results could be applied to all community center users.	28%
Including random assignment was incorrect because people should be divided into the two groups based on their usual method of cleaning their hands.	10%
The random assignment was unnecessary because using volunteers makes the study worthless.	2%

NotSignificant, Based on Utts' "Absence of Significance"

Please read about a research study and answer the question by choosing the option that makes sense to you.

A study of 201 American adults found that they checked personal email on weekdays more often than on weekend days, but this result was not statistically significant. Which of the following conclusions is valid?

Answer Text	% Selected
There might be a difference in how many times adults check their email on weekends and weekdays, but we can't tell from this study.	57%
On average, American adults check their email more often on weekdays than weekend days.	31%
On average, American adults check their email the same amount on weekdays and weekend days.	9%
The study was not conducted correctly.	3%

Variability2, LOCUS Advanced Item

Please read about a research study and answer the question by choosing the option that makes sense to you.

A survey of 625 randomly selected people was conducted to determine their opinion about music. The survey reported that 36 percent of the people surveyed preferred country/western music. The survey estimate had a margin of error of 4 percentage points. A margin of error is reported because …

Answer Text	% Selected
Sample proportions vary from sample to sample.	54%
The people doing the survey may have recorded results incorrectly.	19%
People may misunderstand the survey questions.	18%
People may intentionally respond incorrectly.	9%

Bias, LOCUS Intermediate Item/Utts' "Sources of Bias"

Please read about a research study and answer the question by choosing the option that makes sense to you.

A committee of 27 neighbors wants to estimate the proportion of eligible voters in their school district who intend to vote in the upcoming school board election. They decide to base their estimate on 270 eligible voters. Which of the following plans would allow the committee to generalize from these 270 people to the population of all eligible voters?

Answer Text	% Selected
Mail surveys to 270 randomly selected eligible voters and follow-up with those who do not respond.	51%
Mail surveys to all eligible voters and take the first 270 who respond.	28%
Have each of the 27 neighbors randomly select 10 neighbors to participate in the survey.	12%
Survey 270 people visiting a local grocery store on the Saturday before the election.	10%

Variability3, Utts' "Variability Is Normal"

Please read about a research study and answer the question by choosing the option that makes sense to you.

A library surveyed 452 members and reported that they read 200 pages per week on average. Pat is a member of this library and reads, on average, 100 pages per week. Which of the following pieces of information would allow you to determine whether Pat's behavior is typical of this library's members?

Answer Text	% Selected
No additional information is needed, because Pat reads half as many pages as the average library member.	46%
How much the number of pages read varied between the members surveyed.	42%
How many total members the library has.	7%
How much Pat enjoys reading.	5%

Significance, Utts' "Statistical/Practical Significance"

Please read about a research study and answer the question by choosing the option that makes sense to you.

In a study, 2352 randomly selected adults were shown a picture of either a goat or a cat and asked how relaxed they felt. On a scale of 1 (Extremely tense) to 7 (Extremely relaxed), the average rating for cats was 5.68 and the average rating for goats was 5.57. The researchers determined this difference to be statistically significant. Which of the following is a valid conclusion?

Answer Text	% Selected
For practical purposes both pictures are about equally relaxing, because the difference in ratings was small.	39%
The picture of a cat is much more relaxing than the picture of a goat, because the difference in ratings was statistically significant.	34%
The results were not actually statistically significant, because the difference in ratings was small.	19%
No conclusions can be drawn, because the difference in ratings might be a coincidence.	7%

Causation3, LOCUS Advanced Item

Please read about a research study and answer the question by choosing the option that makes sense to you.

A 13-year study of 1328 adults randomly selected from a population carefully monitored the personal habits and health conditions of participants. Personal habits included tobacco use and coffee consumption. Health conditions included incidence of stroke. Which of the following questions about this population **CANNOT** be answered using data from this study?

Answer Text	% Selected
What percentage of the population are coffee drinkers?	43%
Does coffee consumption cause a reduction in the incidence of stroke?	25%
Are coffee drinkers more likely to smoke than adults who do not drink coffee?	23%
Do coffee drinkers have fewer strokes than adults who do not drink coffee?	8%