UnderStandingAmericaStudy

UAS 594 NUMERACY SCORING PROTOCOL

USC Dornsife Center for Economic and Social Research

8/21/2024

CONSTRUCTION OF NUMERACY SCALE SCORES

The numeracy scale consists of 8 items taken from Weller et al. (2013). For each item, the participant is asked to solve a problem designed to measure "the ability to understand, manipulate, and use numerical information, including probabilities" (p. 198). Items are scored dichotomously as correctly solved or incorrect.

Numeracy scale scores for UAS panel respondents are derived using a two-parameter logistic Item Response Theory (IRT) model. In this IRT model, the probability of correctly solving a test item is viewed as a function of a test taker's ability level and the difficulty and discrimination parameters of the test item. The difficulty parameter measures the ability level at which there is a 50% chance of answering the item correctly, whereas the discrimination parameter measures how sensitive this probability is to differences in the ability level. The two-parameter logistic model allows both the difficulty and discrimination parameters to differ across test items.

IRT scoring requires sufficient unidimensionality of the test items, which was evaluated with confirmatory factor analysis for binary outcome variables. Common criteria for adequate model fit include a root mean squared error of approximation (RMSEA) less than .06, Tucker-Lewis index (TLI) greater than .95, and comparative fit index (CFI) greater than .95. A one-factor model provided an excellent fit to the data, RMSEA = .020 (90% CI = .011/.029), TLI = .996, CFI = .997, supporting sufficient unidimensionality of the 8-item numeracy scale.

Item difficulty and discrimination parameters were calibrated based on a weighted sample of 2,159 UAS respondents, with weights ensuring that the demographic variables race, sex, age, education, and household income in the survey sample match their population counterparts. Because weights are unavailable for the LA County subsample, these were not included in the estimation. However, we compute scores for all respondents, regardless of whether they have a weight or not, provided that they answer at least one item. The item difficulty and discrimination parameters are shown in Table 1.

The final IRT-based scaled scores are converted into T-scores, where 50 is the mean and 10 is the SD of a census-weighted sample of the general United States population. The T-score metric has widespread use in psychological testing and has been adapted, for example, by the Patient-Reported Outcomes Measurement Information System (PROMIS®, Cella et al., 2010). A numeracy score of 50 means that the person's numeric ability is equal to that of the average person in the

general population, a score of 60 means that the person's ability is one standard deviation above average, and a score of 40 means that the person's ability is one standard deviation below average.

Table 1: Item parameters of the numeracy scale

Item	Difficulty	Discrimination
Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up as an even number?	-0.61	1.39
In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS?	-0.47	1.46
In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?	0.87	2.57
If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?	-1.27	1.42
If the chance of getting a disease is 20 out of 100, this would be the same as having how much of a percent chance of getting the disease?	-.088	2.24
Suppose you have a close friend who has a lump in her breast and must have a mammogram… The table below summarizes all of this information. Imagine that your friend tests positive, what is the likelihood that she has a tumor?	1.89	1.97
A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	1.47	2.38
In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	0.90	2.90

REFERENCES

Weller, J. S., Dieckman, N. F., Tusler, M., Mertz, C. K., Burns, W. J., and E. Peters. (2013). "Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach." Journal of Behavioral Decision-Making 26, 198-212.

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Hays, R. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. Journal of Clinical Epidemiology, 63, 1179-1194