UnderStandingAmericaStudy

UAS42 - SCORING EXPLANATION

USC Dornsife Center for Economic and Social Research

8/10/2016

SCORING EXPLANATION

The cognitive measures (number series, picture vocabulary, verbal analogies) were taken from the Woodcock–Johnson Tests of Cognitive Abilities® (Mather and Jaffe, 2016). The tests were designed to measure the respondent's quantitative reasoning (number series) and lexical knowledge (picture vocabulary, verbal analogies). Each measure consists of 15 items, which are scored dichotomously as correctly solved or incorrect.

Cognitive test 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 .90, and comparative fit index (CFI) greater than .90. A one factor model provided a good fit to the data for the number series and verbal analogies tests [number series: RMSEA = .041 (90% CI = .038/.044), TLI = .90, CFI = .91; verbal analogies: RMSEA = .022 (90% CI = .018/.026), TLI = .98, CFI = .98]. Adequacy of model fit was somewhat less consistent for the picture vocabulary test [RMSEA = .042 (90% CI = .039/.046), TLI = .80, CFI = .83], but was considered sufficient to support unidimensionality.

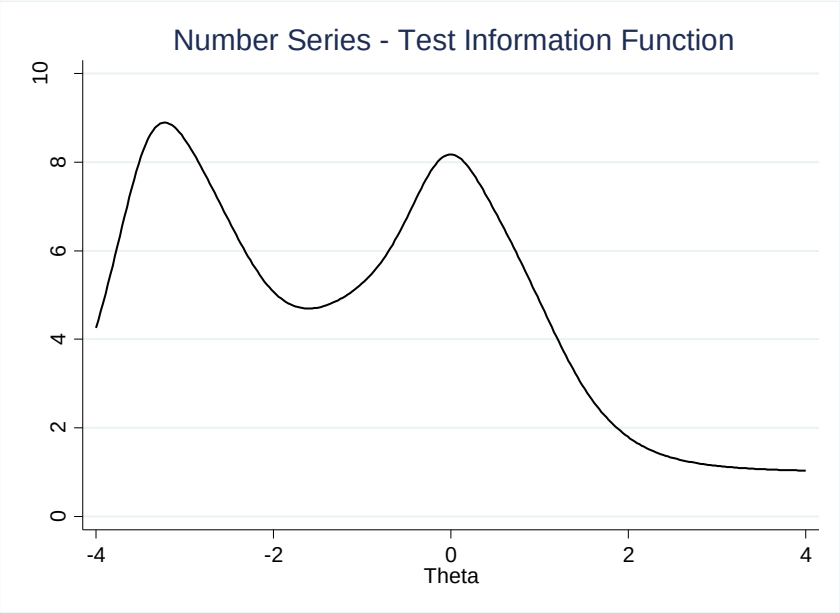
Item difficulty and discrimination parameters were calibrated based on weighted samples of 3,258 UAS respondents, with weights ensuring that the demographic variables race, sex, age, education, and household income in the survey sample match their population counterparts. The estimated item parameters are shown in Tables 1-3. Below each table is the estimated test information function for the measure; the information function shows the reliability of the measure (where an information > 5 corresponds with a reliability >.80) across the latent ability continuum theta (where 0 is the general population mean and 1 is 1 SD above the mean). 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 only exception is that respondents who completed the

picture vocabulary and verbal analogies tests in Spanish do not receive scores on these measures. The reason is that the parameters for items of these lexical knowledge tests may differ between languages, but the sample size of Spanish test takers is currently too small to evaluate this.

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 score of 50 means that the person's cognitive 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 number series test (N=3,258)

Item	Difficulty	Discrimination
nsa_11	-3.404889	3.559662
nsa_12	-3.450795	2.26631
nsa_13	-2.713029	1.709391
nsa_21	-3.043672	2.594486
nsa_22	-2.104694	1.592658
nsa_23	-0.9996097	1.120357
nsa_31	-2.658152	2.626019
nsa_32	-1.168207	2.061662
nsa_33	-0.512248	1.789896
nsa_41	-0.485506	0.9929408
nsa_42	-0.0651054	1.107131
nsa_43	0.5897162	2.468995
nsa_51	-0.0437221	3.395932
nsa_52	-0.2538728	1.929131
nsa_53	0.8100692	2.375817


Number Series - Test Information Function
0
2
4
6
8
10
-4
-2
0
2
4
Theta

REFERENCES

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

Mather, N, Jaffe, L.E. (2016). *Woodcock-Johnson IV: Reports, Recommendations, and Strategies.* Jossey-Bass: Hoboken, NJ.