**The Numeric Understanding Measures (NUM): Developing and Validating Adaptive and Non-Adaptive Numeracy Scales**

**SUPPLEMENTAL MATERIALS**

Michael C. Silverstein, Pär Bjälkebring, Brittany Shoots-Reinhard, Ellen Peters

Michael C. Silverstein, MS\*

Center for Science Communication Research

Department of Psychology

University of Oregon

1715 Franklin Blvd.

Eugene, OR 97403

United States

Email: msilver2@uoregon.edu

https://orcid.org/0000-0001-5497-3578

Pär Bjälkebring, PhD

Department of Psychology, University of Gothenburg

Center for Science Communication Research, University of Oregon

Email: pop-par.bjalkebring@psy.gu.se

https://orcid.org/0000-0002-4430-4044

Brittany Shoots-Reinhard, PhD

Department of Psychology, The Ohio State University

Center for Science Communication Research, University of Oregon

Email: shoots-reinhard.1@osu.edu

https://orcid.org/0000-0002-2844-4995

Ellen Peters, PhD

Center for Science Communication Research

School of Journalism and Communication

Department of Psychology

University of Oregon

Email: ellenpet@uoregon.edu

https://orcid.org/0000-0003-0702-6169

\*Address correspondence to:

Michael C. Silverstein

1715 Franklin Blvd, Room 146

Eugene, OR 97403

msilver2@uoregon.edu

**Appendix A: Item Parameters for 66 Candidate Items in the Calibration Study**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Item** | **Single Factor Loading (**λ) | **Loading Standard Error** | **Discriminability (**α) | **Discriminability Standard Error** | **Difficulty (**β) | **Difficulty Standard Error** |
| In a box of pens, 25% have blue ink, 25% have red ink, and 50% have black ink. 10% of the blue and black pens don’t work and 90% of the red pens don’t work. What percentage of pens that don’t work are red? \_\_ % | 0.41 | 0.18 | 0.76 | 0.23 | 4.91 | 1.6 |
| In a box of cookies, 25% have chocolate chips, 25% have raisins, and 50% are plain. 40% of the chocolate chip and plain cookies aren’t fresh, and 30% of the raisin cookies aren’t fresh. What percentage of cookies that aren’t fresh are raisin cookies?  \_\_ % | 0.49 | 0.11 | 0.95 | 0.17 | 2.51 | 0.42 |
| If the probability of getting the common cold is 60% in 1 year, what is the probability of getting the common cold in 2 years? \_\_\_\_\_% | 0.75 | 0.12 | 1.92 | 0.41 | 2.28 | 0.37 |
| In a lake 20% of fish are red. A red fish is poisonous with a probability of 20%. A fish that is not red is poisonous with a probability of 15%. What is the probability that a poisonous fish in the lake is red? \_\_\_\_\_\_\_% | 0.69 | 0.09 | 1.64 | 0.24 | 1.89 | 0.24 |
| In the Broadway Direct Lottery, the chance of winning a seat to a Broadway show is 2 in 4,000. What percent of tickets of Broadway Direct Lottery win a seat? \_\_\_\_\_\_\_% | 0.63 | 0.1 | 1.38 | 0.22 | 0.83 | 0.17 |
| If the chance of getting struck by lightning is 5 out of 2500, this would be the same as having what percent chance of getting struck by lightning? \_\_ \_\_% | 0.74 | 0.09 | 1.86 | 0.28 | 0.74 | 0.14 |
| Out of 1,000 people in a city, 500 are members of a society. Out of these 500 members in the society, 300 are women. Out of the 500 people in the city that are not in the society, 100 are women. What is the probability from 0-100% that a randomly drawn woman is a member of the society? \_\_\_\_\_% | 0.74 | 0.07 | 1.89 | 0.24 | 0.71 | 0.12 |
| Imagine we are throwing a loaded five-sided die (the sides of which show 1, 2, 3, 4, 5). The probability that the die shows a 5 is twice times as high as the probability of each of the other numbers. On average, out of 120 throws how many times would the die show the number 1? \_\_\_\_\_\_\_\_out of 120 throws. | 0.58 | 0.09 | 1.22 | 0.17 | 0.61 | 0.16 |
| In the OHIO LOTTERY, the chance of winning is 4 in 8,000. What percent of tickets of OHIO LOTTERY would you expect to win the lottery? \_\_\_\_\_\_\_% | 0.63 | 0.1 | 1.39 | 0.21 | 0.55 | 0.15 |
| The town of Jamesville has a pole that is red, blue and green standing in the center of town. One-third of a pole is painted red, one-half of the pole is painted blue, and three feet of the pole is painted green. What is the height of the pole? \_\_\_\_\_\_ feet | 0.83 | 0.07 | 2.51 | 0.37 | 0.53 | 0.12 |
| Out of 300 fruits, 200 are apples and 100 are bananas. Out of the 200 apples, 90 are green. Out of the 100 bananas, 30 are green. What is the probability that a randomly picked green fruit will be an apple?   \_\_\_\_ % | 0.76 | 0.08 | 1.99 | 0.28 | 0.52 | 0.13 |
| Out of 300 animals, 200 are dogs and 100 are hamsters. Out of the 200 dogs, 90 are brown. Out of the 100 hamsters, 30 are brown. What is the probability that a randomly picked brown animal will be a dog?   \_\_\_\_ % | 0.83 | 0.06 | 2.51 | 0.35 | 0.52 | 0.11 |
| If you are playing with a shuffled deck of 52 cards (with four Aces) and you draw 12 random cards, including the Ace of clubs and the Ace of diamonds, what is the percent chance that the next randomly picked card will be another Ace? \_\_\_% | 0.51 | 0.11 | 1 | 0.17 | 0.5 | 0.21 |
| If you are playing with a shuffled deck of 52 cards (with 12 face cards) and you draw 12 random cards, including four of the face cards, what is the percent chance that the next randomly picked card will be a face card? \_\_\_% | 0.66 | 0.09 | 1.48 | 0.21 | 0.47 | 0.16 |
| Emily is being treated for stage 2 breast cancer. The chance that the cancer will come back is 10% over 10 years. If Emily takes a new medicine, this chance will decrease by 10%. If 100 women like Emily take this medicine, how many are now expected to have breast cancer come back within 10 years? \_\_\_ out of 100 women | 0.76 | 0.07 | 1.97 | 0.27 | 0.45 | 0.11 |
| Out of 1,300 people in a small town, 700 are members of a choir. Out of these 700 members in the choir, 300 are men. Out of the 600 inhabitants that are not in the choir, 100 are men. What is the probability from 0-100% that a randomly drawn man is a member of the choir? \_\_\_\_\_% | 0.83 | 0.07 | 2.57 | 0.4 | 0.45 | 0.11 |
| If 50 people in a town of 100,000 people catch a virus, what percent of the town has the virus? ­­\_\_\_ \_\_\_% of the town | 0.58 | 0.1 | 1.22 | 0.19 | 0.44 | 0.16 |
| Allenton College has a column that is green, white, and yellow (the school’s colors) standing in front of the campus library. One-third of the column is painted green, one-half of the column is painted white, and four feet of the column is painted yellow. What is the height of the column? \_\_\_\_\_\_ feet | 0.84 | 0.06 | 2.68 | 0.41 | 0.41 | 0.12 |
| A sandwich and a cookie cost $5.50. The cookie costs $5 less than the sandwich. How much does the cookie cost? \_\_\_\_\_ cents | 0.74 | 0.08 | 1.88 | 0.26 | 0.18 | 0.12 |
| If the chance of winning the lottery is .01%, how many people would be expected to win the lottery out of 100,000 people? \_\_\_\_\_ people  | 0.41 | 0.11 | 0.77 | 0.15 | 0.09 | 0.2 |
| If John can drink one jug of water in 6 days, and Mary can drink one jug of water in 12 days, how long would it take them to drink one jug of water together? \_\_\_\_\_ days | 0.62 | 0.09 | 1.36 | 0.19 | 0.09 | 0.14 |
| Imagine that you have a five-sided die (the sides of which show 1, 2, 3, 4, 5), and we throw it 150 times. On average, out of these 150 throws how many times would this five-sided die show an odd number (1, 3, 5)? \_\_\_\_\_ throws | 0.79 | 0.07 | 2.17 | 0.3 | 0 | 0.1 |
| Imagine a city where the chance of rain on any given day in April is twice as high as the chance that it does not rain. On average, how many days in April (which has 30 days) would it rain? \_\_\_\_\_ days | 0.69 | 0.08 | 1.64 | 0.23 | -0.06 | 0.12 |
| If the chance of getting a disease is 5 out of 250, this would be the same as having what percent chance of getting the disease? \_\_\_\_% | 0.69 | 0.08 | 1.64 | 0.22 | -0.16 | 0.12 |
| In a field containing 1000 squirrels, 40% of squirrels are striped and a striped squirrel is rabid with a probability of 20%, on average, how many squirrels are there in the field that are rabid and have stripes? \_\_\_ squirrels | 0.79 | 0.07 | 2.17 | 0.29 | -0.2 | 0.11 |
| The chance of getting a paper cut is 0.0001 in one year. Out of 500,000 people, about how many of them are expected to get a paper cut in one year? ­­\_\_\_\_\_\_ people | 0.52 | 0.1 | 1.02 | 0.16 | -0.27 | 0.16 |
| In a forest, 70 trees are pine trees and 30 trees are oak trees. If a wildfire destroys 10% of the pine trees and 20% of the oak trees, how many total pine and oak trees are left undestroyed? \_\_\_ trees | 0.67 | 0.09 | 1.53 | 0.21 | -0.28 | 0.13 |
| Suppose that you are running a lottery as a fundraiser and you plan on selling 1000 tickets. If 90% of the earnings are awarded to the winner and the other 10% is donated to charity, what is the minimum amount you could charge per ticket to raise $200 for the donation? $\_\_\_\_\_ (dollars) | 0.66 | 0.09 | 1.51 | 0.21 | -0.28 | 0.13 |
| Imagine you are throwing a fair six-sided die (the sides of which show 1, 2, 3, 4, 5, 6) 120 times. On average, how many times would you expect this die to show a number less than 5 (1, 2, 3 or 4)? \_\_\_\_\_\_ out of 120 throws. | 0.79 | 0.07 | 2.16 | 0.29 | -0.38 | 0.12 |
| One nickel (which equals $0.05) weighs 4 grams. How much does $1.05 in nickels weigh? \_\_\_\_\_\_ grams | 0.51 | 0.12 | 1 | 0.19 | -0.44 | 0.21 |
| Imagine you are throwing a fair six-sided die 150 times (the sides of which show 1, 2, 3, 4, 5, 6). On average, how many times would you expect this die show a number less than 3 (1 or 2)? \_\_\_\_\_\_ out of 150 throws. | 0.66 | 0.09 | 1.49 | 0.21 | -0.49 | 0.16 |
| If the chance of slipping on a rainy day is 60 out of 400, this would be the same as having what percent chance of slipping on a rainy day? \_\_\_\_\_% | 0.75 | 0.08 | 1.96 | 0.28 | -0.63 | 0.14 |
| The doctor told Doug not to take more than 5 grams (g) of Ibuprofen a day. Each Ibuprofen pill is 700 milligrams (mg; 1000 mg = 1 g). What is the greatest number of whole pills that Doug can take in one day? \_\_\_\_ pills | 0.74 | 0.09 | 1.88 | 0.31 | -0.71 | 0.41 |
| If it takes 25 monkeys 25 minutes to eat 25 bananas, how long would it take 100 monkeys to eat 100 bananas? \_\_\_\_\_\_\_ minutes | 0.61 | 0.1 | 1.3 | 0.2 | -0.76 | 0.18 |
| In a field, there is a colony of mushrooms. Every day, the colony doubles in size. If it takes 28 days for the colony to cover the entire field, how long would it take for the colony to cover half of the field? \_\_\_\_\_\_\_\_ days | 0.74 | 0.08 | 1.9 | 0.26 | -0.82 | 0.14 |
| Imagine that you just made a post to a social media website. Your chance of a response from one of your friends is 5% in the first minute, and doubles every minute after (i.e., it's 10% after 2 minutes, 20% after 3 minutes etc.). What is the chance after 5 minutes? \_\_\_\_\_\_% | 0.67 | 0.1 | 1.53 | 0.24 | -0.87 | 0.18 |
| If the chance of getting a disease is 60 out of 300, this would be the same as having what percent chance of getting the disease? \_\_\_\_\_\_ % | 0.8 | 0.07 | 2.24 | 0.31 | -1.09 | 0.13 |
| Dan is treated after a heart attack. The chance that he will have another heart attack is 10% over 10 years. If Dan starts a new diet and exercise regimen, this chance will decrease by 50%. If 100 men like Den starts this new diet and exercise regimen, how many are now expected to have another heart attack within 10 years? \_\_\_ out of 100 men | 0.63 | 0.1 | 1.31 | 0.2 | -1.09 | 0.21 |
| In the LUCKY WINNER LOTTERY, the chances of winning a $60 prize are 100 in 1000. How many people would you expect to win a $60.00 prize if 10,000 people each buy a single ticket from LUCKY WINNER? \_\_\_\_\_ people | 0.66 | 0.1 | 1.48 | 0.24 | -1.16 | 0.21 |
| If 70% of basketball players on a college basketball team are over six feet tall and there are 20 players on the team, how many players on the team are shorter than six feet tall? \_\_\_ players | 0.8 | 0.07 | 2.31 | 0.32 | -1.19 | 0.19 |
| In a fast food sweepstakes, the chance of winning a $5 gift card with the purchase of a soft drink is 1%. How many people, on average, would you expect to win a $5 gift card if 4,500 people each buy a single soft drink? \_\_\_\_ people | 0.68 | 0.09 | 1.58 | 0.24 | -1.27 | 0.2 |
| Suppose that you are buying a gallon of milk at the grocery store. There are two options for the same brand of milk: buying 4 quarts at $2.50 per quart or buying 1 gallon for $8.00. What is the cost per quart (1 gallon=4 quarts) of the better priced milk? $\_\_\_\_\_\_ per quart | 0.73 | 0.08 | 1.84 | 0.25 | -1.41 | 0.17 |
| A pair of shoes costs $32 to buy (excluding tax). If the shoes are on sale for 25% off, how much will these shoes cost (excluding sales tax)? $\_\_\_\_\_\_\_ dollars | 0.6 | 0.1 | 1.29 | 0.19 | -1.45 | 0.22 |
| Imagine that you just sent an email to your boss. Your chance of a response from your boss is 4% in the first minute, and doubles every minute after the first (i.e., it's 8% after 2 minutes, 16% after 3 minutes etc.). What is the chance after 5 minutes? \_\_\_\_\_\_% | 0.74 | 0.09 | 1.86 | 0.3 | -1.52 | 0.19 |
| In the Lucky Star Lottery, the chance of winning a $10 prize is 1%. How many people would you expect to win a $10.00 prize if 500 people each buy a single ticket to the Lucky Star Lottery? \_\_\_\_ people | 0.71 | 0.1 | 1.71 | 0.28 | -1.53 | 0.2 |
| A shirt costs $24 to buy (excluding tax). There is currently a 25% off sale that applies to this shirt. How much with this shirt cost you (excluding sales tax)? $\_\_\_\_\_\_\_ (dollars). | 0.63 | 0.1 | 1.39 | 0.23 | -1.55 | 0.26 |
| A medical study will either give people medicine A or medicine B. Each person has an equal chance to get medicine A or B. If there are 536 people in the study, about how many are expected to get medicine A? \_\_\_ people | 0.76 | 0.08 | 2 | 0.3 | -1.58 | 0.24 |
| Suppose you go to a farmer stand that only accepts cash and does not charge tax. You have $15 in cash. If you buy 2 watermelons for $3 each and 4 apples for $1.50 each, how many bananas can you afford to buy if they each cost $0.75? \_\_\_\_\_ bananas | 0.59 | 0.11 | 1.25 | 0.21 | -1.68 | 0.28 |
| If a class of 200 people includes 50 men, this would be the same as the class being what percent men? \_\_\_\_% | 0.74 | 0.1 | 1.86 | 0.31 | -1.68 | 0.21 |
| One dime (which equals $0.10) weighs 2 grams. How much does $1.30 in dimes weigh? \_\_\_\_\_\_ grams | 0.7 | 0.1 | 1.66 | 0.29 | -1.82 | 0.26 |
| In the New Jersey State Lottery, the chances of winning a $50.00 prize are 1 in 100. How many people would you expect to win a $50.00 prize if 1,700 people each buy a single ticket from New Jersey State Lottery? \_\_\_\_\_ people | 0.7 | 0.11 | 1.67 | 0.29 | -1.82 | 0.26 |
| Suppose that you are buying a gallon of orange juice at the grocery store. There are two options for the same brand of orange juice: buying 4 quarts at $3.50 per quart or buying 1 gallon for $12.00. What is the cost per quart (1 gallon=4 quarts) of the better priced orange juice? $\_\_\_\_\_\_ per quart | 0.58 | 0.12 | 1.2 | 0.23 | -1.83 | 0.33 |
| What is 75% of $300? $\_\_\_\_\_\_\_ (dollars) | 0.63 | 0.11 | 1.37 | 0.24 | -1.9 | 0.31 |
| What is 63% of 1000? \_\_\_\_\_\_\_\_ | 0.65 | 0.11 | 1.46 | 0.26 | -1.97 | 0.28 |
| Suppose you go to a food stall that only accepts cash and does not charge tax. You have $10 in cash. If you buy 2 hot dogs for $2 each and 3 burgers for $1.50 each, how many mini corn dogs can you afford to buy if they each cost $0.50? \_\_\_\_\_ mini corn dogs | 0.49 | 0.13 | 0.95 | 0.19 | -2.02 | 0.41 |
| If there are 993 students signed up to take an introductory biology course and there are 3 sections of the class being offered, how many students should be in each section if all the sections of the class are the same size? \_\_\_ students | 0.71 | 0.11 | 1.7 | 0.31 | -2.1 | 0.3 |
| The doctor told Doug not to take more than 3 grams (g) of Ibuprofen a day. Each Ibuprofen pill is 500 milligrams (mg; 1000 mg = 1 gram). What is the greatest number of pills that Doug can take in one day? \_\_\_\_ pills | 0.62 | 0.13 | 1.35 | 0.27 | -2.12 | 0.37 |
| What is 457 + 563? \_\_\_\_\_. | 0.49 | 0.13 | 0.96 | 0.19 | -2.31 | 0.5 |
| A company will either assign people to project A or project B. Each person has an equal chance to get project A or B. If there are 350 people in the company, about how many are expected to work on project A? \_\_\_ people | 0.66 | 0.12 | 1.51 | 0.29 | -2.33 | 0.36 |
| You would like to buy a goldfish for $0.75. However, you do not want your goldfish to be lonely when you are not around, so you decide to get a second goldfish for the same price. Along with fish food for $2.00, what is the total price for the two goldfish and food? $\_\_\_\_\_\_ (dollar) | 0.54 | 0.13 | 1.08 | 0.21 | -2.34 | 0.46 |
| What is 43% of 100? \_\_\_\_\_\_\_\_ | 0.68 | 0.13 | 1.57 | 0.32 | -2.43 | 0.41 |
| If Person A's chances of winning the lottery is 4% in twenty years, and Person B's chances is triple that of A's, what are B's chances? \_\_\_\_ % in twenty years | 0.62 | 0.13 | 1.35 | 0.27 | -2.43 | 0.4 |
| What is 63% of $100? $\_\_\_\_\_\_\_ | 0.66 | 0.13 | 1.5 | 0.32 | -2.45 | 0.43 |
| What is 1024 x 3? \_\_\_\_ | 0.58 | 0.13 | 1.2 | 0.24 | -2.53 | 0.49 |
| If Person A's risk of getting a disease is 7% in twenty years, and Person B's risk is double that of A's, what is B's risk of getting the disease in twenty years? \_\_\_\_ % in twenty years | 0.8 | 0.1 | 2.24 | 0.46 | -2.66 | 0.41 |
| What is 74% of 100 people? \_\_\_\_\_\_\_\_ people | 0.75 | 0.11 | 1.93 | 0.39 | -2.69 | 0.38 |

**Appendix B: Construction of the Adaptive Numeric Understanding Measure (A-NUM)**

The adaptive measure was developed and simulated using CatR (Magis et al., 2018). All adequately fitted items were used from the question bank for the development of the A-NUM. The measure was set up to assume that participants’ numeracies are distributed normally about average ability (z = 0). The scale used expected a posteriori for intermittent ability estimates as well as for final estimates and used the global-discrimination index to select items (Kaplan et al., 2015). The measure began with a fixed item chosen for its average difficulty and high discriminability (α = 2.17, β = 0.00) and ended once the participant had responded to four items. The resulting measure used 13 unique questions of which each participant responded to four (constituting seven unique measures) and categorized participants into one of nine categories or 16 unique EAP estimates (ranging between z = -2.57 and z = 2.21).

Once the measure was established, a simulation study was conducted. 5,000 thetas were generated from a standard normal distribution. Measure responses were then simulated for all of the thetas using catR (Magis et al., 2018) and the theta estimates (using EAP) were compared to the true thetas. The ability estimates were highly correlated with the actual abilities (*r* = .85, *p* < .001). Moreover, the error in the estimated ability was unassociated with true ability (*r* = .003, *n.s.*) meaning that the scale does not directionally bias estimates. However, as expected from a four-item measure, the estimates of ability tended to regress toward average and the RMSE of the estimates tended to be large at the extremes meaning that the scale is less accurate at very high and very low abilities.

**Figure S1.** Results of the Simulation Study of the A-NUM

The graphs show associations between actual ability and estimated ability (A), directional bias in the measure (B), uncertainty in the measure (C), and difference between actual and estimated ability (D). They demonstrate that the A-NUM is accurate (A), regressive in its estimates (B), has less information at the extreme of the ability spectrum (C), but is an unbiased estimator of ability (D).



**C)**

**D)**

**B)**

**A)**

**Figure S2.** Item Information Curves for the Items Making up the A-NUM.

As seen from this the spread of item curves, the items of the A-NUM spans a wide range of ability.



**Figure S3.** Test Information Curves for the Items Making up the A-NUM

As seen from this graph, the A-NUM is most informative of ability around average ability and drops off at the high and low ends. Additionally, the test is generally more informative of ability for people who are below average than above average.



**Appendix C: Construction of the Four-Item Numeric Understanding Measure (4-NUM)**

After removing the items used in A-NUM to avoid any overlap, the 4-NUM items were chosen to be highly discriminating (ranging from α = 1.62 to α = 2.16) and to cover a wide range of ability levels with item difficulties ranging between β = -1.41 and β = 1.58 (this difficulty range includes about 86.4% of the population). Similar to the A-NUM, a simulation study was conducted for the 4-NUM using catIRT (Nydick & Nydick, 2013). 5,000 thetas were generated from a standard normal distribution. Measure responses were then simulated for all of the thetas using catIRT (Nydick & Nydick, 2013) and the theta estimates (using EAP) were compared to the true thetas. The ability estimates were highly correlated with the actual abilities (*r* = .80, *p* < .001), moreover, the error in the estimated ability was unassociated with the true ability (*r* = -.01, *n.s.*). Moreover, the EAP estimates were highly correlated with sum scores (*r* > .99).

**Figure S4.** Results of the Simulation Study of the Four-Item Numeric Understanding Measure (4-NUM)

The graphs show associations between actual ability and estimated ability (A), difference between actual and estimated ability (B), and uncertainty in the measure (C). They demonstrate that the 4-NUM is accurate (A), but is an unbiased estimator of ability (B), but has less information at the extremes of the ability spectrum (C),



**C)**

**B)**

**A)**

**Figure S5.** Item Information Curves for the Items Making up the 4-NUM

As seen from this the spread of item curves, the items of the 4-NUM spans a wide range of ability.



**Figure S5.** Test Information Curves for the Items Making up the 4-NUM

As seen from this graph, the 4-NUM is most informative of ability around average ability and drops off at the higher and lower ability levels.



**Appendix D: Comparative Predictive Validity Analyses with and without Covariates**

**Table S1.** Comparative Validity Analyses Regressing Decision Performance on Numeracy Measures for the Probability Interpretation Task

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **A-NUM** | **4-NUM** | **1-NUM** | **BNT** | **Weller** |
|  | ***Predictors*** | **OR** | **p** | **OR** | **p** | **OR** | **p** | **OR** | **p** | **OR** | **p** |
| **No Covariates** | **Intercept** | 0.09 | < .001 | 0.38 | < .001 | 0.62 | < .001 | 0.31 | < .001 | 0.14 | < .001 |
| **Numeracy** | 1.69 | < .001 | 1.74 | < .001 | 3.00 | < .001 | 1.70 | < .001 | 1.46 | < .001 |
|  | **Tjur R2** | .12 | .09 | .07 | .09 | .11 |
| **With Covariates** | **Intercept** | 0.05 | < .001 | 0.15 | < .001 | 0.17 | < .001 | 0.11 | < .001 | 0.08 | < .001 |
| **Numeracy** | 1.62 | < .001 | 1.62 | < .001 | 2.56 | < .001 | 1.59 | < .001 | 1.42 | < .001 |
|  | **Raven’s Matrices** | 0.97 | .056 | 1.00 | .94 | 1.04 | .49 | 1.02 | .76 | .96 | .48 |
|  | **Vocabulary** | 1.14 | .01 | 1.18 | < .001 | 1.12 | < .001 | 1.18 | < .001 | 1.16 | < .001 |
|  | **Tjur R2** | .13 | .11 | .10 | .11 | .13 |

**Table S2.** Comparative Validity Analyses Regressing Decision Performance on Numeracy Measures for the Benefit Perception Task

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **A-NUM** | **4-NUM** | **1-NUM** | **BNT** | **Weller** |
|  | ***Predictors*** | **OR** | **p** | **OR** | **p** | **OR** | **p** | **OR** | **p** | **OR** | **p** |
| **No Covariates** | **Intercept** | 0.17 | < .001 | 0.24 | < .001 | 0.31 | < .001 | 0.18 | < .001 | 0.15 | < .001 |
| **Numeracy** | 1.17 | .02 | 1.20 | .02 | 1.28 | .19 | 1.31 | < .001 | 1.17 | .002 |
|  | **Tjur R2** | .01 | .01 | .003 | .02 | .02 |
| **With Covariates** | **Intercept** | 0.14 | < .001 | 0.19 | < .001 | 0.19 | < .001 | 0.16 | < .001 | 0.14 | < .001 |
| **Numeracy** | 1.16 | .048 | 1.18 | .06 | 1.20 | .35 | 1.30 | .002 | 1.18 | .004 |
|  | **Raven’s Matrices** | 0.97 | .58 | 0.97 | .63 | 1.00 | .96 | 0.96 | .49 | 0.95 | .36 |
|  | **Vocabulary** | 1.06 | .27 | 1.07 | .20 | 1.08 | .11 | 1.05 | .30 | 1.05 | .32 |
|  | **Tjur R2** | .01 | .01 | .007 | .02 | .02 |

**Table S3.** Comparative Validity Analyses Regressing Decision Performance on Numeracy Measures for the Framing Task without Covariates

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **A-NUM** | **4-NUM** | **1-NUM** | **BNT** | **Weller** |
| ***Predictors*** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** |
| **Intercept** | 4.11 | < .001 | 4.19 | < .001 | 4.16 | < .001 | 4.15 | < .001 | 4.16 | < .001 |
| **Frame** | 1.03 | < .001 | 0.80 | < .001 | 0.73 | < .001 | 0.85 | < .001 | 0.90 | < .001 |
| **Numeracy** | 0.02 | .47 | 0.01 | .80 | 0.09 | < .34 | 0.03 | .48 | 0.01 | .72 |
| **Frame\*****Numeracy** | -0.08 | .08 | -0.10 | .06 | -0.12 | .34 | -0.07 | .17 | -0.04 | .19 |
| ***Random Effect*** |  |  |  |  |
| ***σ2*** | 0.53 | 0.53 | 0.53 | 0.53 | 0.53 |
| ***τ00 Participants*** | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| ***τ00 Students*** | 1.20 | 1.20 | 1.20 | 1.20 | 1.20 |
| **ICC** | 0.76 | 0.76 | 0.76 | 0.76 | 0.76 |
| **Marginal/****Conditional R2** | .05/.78 | .05/.78 | .05/.78 | .05/.78 | .05/.78 |

**Table S4.** Comparative Validity Analyses Regressing Decision Performance on Numeracy Measures for the Framing Task with Covariates

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **A-NUM** | **4-NUM** | **1-NUM** | **BNT** | **Weller** |
| ***Predictors*** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** |
| **Intercept** | 4.41 | < .001 | 4.52 | < .001 | 4.59 | < .001 | 4.51 | < .001 | 4.47 | < .001 |
| **Frame** | 1.02 | < .001 | 0.87 | < .001 | 0.74 | < .001 | 0.84 | < .001 | 0.93 | < .001 |
| **Raven’s Matrices** | -0.02 | .22 | -0.02 | .41 | -0.02 | .21 | -0.02 | .26 | -0.02 | .27 |
| **Vocab** | -0.05 | .002 | -0.05 | .004 | -0.05 | .001 | -0.05 | .002 | -0.05 | .002 |
| **Numeracy** | 0.02 | .47 | 0.04 | .27 | 0.15 | .10 | 0.06 | .14 | 0.04 | .16 |
| **Frame\*****Numeracy** | -0.07 | .09 | -0.10 | .052 | -0.12 | .35 | -0.07 | .18 | -0.05 | .13 |
| ***Random Effect*** |  |  |  |  |
| ***σ2*** | 0.53 | 0.53 | 0.53 | 0.53 | 0.53 |
| ***τ00 Participants*** | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 |
| ***τ00 Students*** | 1.20 | 1.20 | 1.20 | 1.20 | 1.20 |
| **ICC** | 0.76 | 0.76 | 0.76 | 0.76 | 0.76 |
| **Marginal/****Conditional R2** | .06/.78 | .06/.78 | .06/.78 | .05/.78 | .05/.78 |

**Table S5.** Comparative Validity Analyses Regressing Decision Performance on Numeracy Measures for the Bets Task

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **A-NUM** | **4-NUM** | **1-NUM** | **BNT** | **Weller** |
|  | ***Predictors*** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** |
| **No Covariates** | **Intercept** | 9.52 | < .001 | 8.43 | < .001 | 7.80 | < .001 | 8.01 | < .001 | 8.82 | < .001 |
| **Condition** | 0.03 | .98 | 1.29 | .14 | 2.64 | < .001 | 1.54 | .13 | 1.59 | .24 |
| **Numeracy** | -0.44 | .049 | -0.56 | .06 | -0.82 | .21 | -0.26 | .35 | -0.26 | .13 |
| **Numeracy\*****Condition** | 0.74 | .02 | 1.22 | .002 | 1.86 | .048 | 0.86 | .03 | 0.36 | .13 |
|  | **R2/Adj. R2** | .10/.09 | .10/.10 | .09/.08 | .10/.09 | .09/.09 |
| **With Covariates** | **Intercept** | 12.18 | < .001 | 12.14 | < .001 | 11.55 | < .001 | 11.82 | < .001 | 11.77 | < .001 |
| **Condition** | 0.25 | .87 | 1.45 | .09 | 2.64 | < .001 | 1.46 | .15 | 1.65 | .22 |
| **Numeracy** | -0.10 | .67 | -0.14 | .64 | -0.25 | .71 | 0.05 | .86 | -0.04 | .81 |
|  | **Raven’s Matrices** | -0.21 | .14 | -0.25 | .09 | -0.19 | .18 | -0.24 | .10 | -0.20 | .18 |
|  | **Vocabulary** | -0.49 | < .001 | -0.48 | < .001 | -0.47 | < .001 | -.50 | < .001 | -0.47 | < .001 |
|  | **Numeracy\* Condition** | 0.71 | .02 | 1.15 | .004 | 1.87 | .04 | 0.89 | .02 | 0.36 | .12 |
|  | **R2/Adj. R2** | .12/.11 | .13/.12 | .12/.11 | .13/.12 | .12/.11 |

**Table S6.** Comparative Validity Analyses Regressing Decision Performance on Numeracy Measures for the Risk Consistency Task

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **A-NUM** | **4-NUM** | **1-NUM** | **BNT** | **Weller** |
|  | ***Predictors*** | **Estimate** | **p** | **Estimate** | **P** | **Estimate** | **p** | **Estimate** | **p** | **Estimate** | **p** |
| **No Covariates** | **Intercept** | 3.99 | < .001 | 4.42 | < .001 | 4.76 | < .001 | 4.53 | < .001 | 4.01 | < .001 |
| **Numeracy** | 0.23 | < .001 | 0.35 | < .001 | 0.64 | < .001 | 0.23 | .002 | 0.20 | < .001 |
|  | **R2/Adj. R2** | .03/.02 | .04/.04 | .02/.02 | .02/.01 | .03/.03 |
| **With Covariates** | **Intercept** | 3.35 | < .001 | 3.67 | < .001 | 3.68 | < .001 | 3.54 | < .001 | 3.39 | < .001 |
| **Numeracy** | 0.15 | .02 | 0.27 | < .001 | 0.46 | .01 | 0.14 | .08 | 0.14 | .003 |
|  | **Raven’s Matrices** | 0.13 | .01 | 0.12 | .02 | 0.14 | .005 | 0.15 | .004 | 0.12 | .03 |
|  | **Vocabulary** | 0.05 | .29 | 0.04 | .33 | .06 | .17 | 0.06 | .16 | 0.04 | .32 |
|  | **R2/Adj. R2** | .04/.03 | .05/.04 | .04/.04 | .03/.03 | .04/.04 |

**Appendix E: Exploratory Analyses of Individual Evaluations in Framing Task**

To further explore responses in this task, evaluations of the individual students were predicted in separate multiple linear regressions from the framing condition (positive vs. negative), numeracy as assessed by the four measures, and the frame-by-numeracy interaction controlling for vocabulary and Raven’s matrices with a Sidak correction for multiple test (α-level = .01). Table S7 describes the 5 student performances used as stimuli for the framing task.

**Table S7.** Stimuli For the Five Students Described in the Framing Task

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Number** | **Student Label** | **200-Level Grade (correct)** | **300-Level Grade****(correct)** | **400-Level Grade****(correct)** | **Average****(correct)** | **A-NUM-by-Frame Interaction Effect Size (Partial R2)**  |
| **1** | Mike | 78% | 54% | 67% | 66.3% | < .001 |
| **2** | **Molly** | **85%** | **82%** | **95%** | **87.3%** | **.016\*** |
| **3** | Jesse | 93% | 67% | 89% | 83.0% | .008 |
| **4** | Bernie | 76% | 77% | 84% | 79.0% | .003 |
| **5** | Linda | 65% | 93% | 77% | 78.3% | < .001 |

*Note*. A-NUM-by-Frame Interaction is based on a model controlling for vocab and Raven’s Matrices. \* indicates p < .01.

Evaluations of one student (the student with the highest average grade: Molly; 87.3% average correct) were significantly predicted by the frame × numeracy interaction when numeracy was measured by the A-NUM (*b* = -0.15, *SE =* 0.05, *t = -*3.07, *p* = .002) and the Weller (*b* = -0.14, *SE =* 0.05, *t = -*2.97, *p* = .003), controlling for vocabulary and Raven’s matrices. The frame × numeracy did not attain significance at the .01 α-level when numeracy was measured by the 4-NUM (*b* = -0.13, *p* = .03), BNT (*b* = -0.11, *p* = .05), nor 1-NUM (*b* = -0.23, *p* = .11). None of the other student evaluations attained significance at the .01 α-level using any of the numeracy measures. This suggests a potential boundary condition for the numeracy-by-frame interaction related to the numerical proximity of the positive and negative frames (i.e., the magnitude of difference between positive and negative frames). Numeracy-by-frame interactions may be larger when the difference between positive and negative frames is larger. This boundary condition would be consistent with the proposed mechanism for the numeracy × frame effect: that the highly numerate are more likely to convert between positive and negative frames (Peters et al., 2006). Higher grades result in larger differences between positive and negative frames than lower grades closer to 50%. For example, converting 10% incorrect to 90% correct may be more impactful on evaluations than converting 48% incorrect to 52% correct. This argument implies that the use of higher average grades across the stimuli might have resulted in a replication of the original effect.

**Appendix F: Power Analyses for Predictive Validity Tasks**

Using G\*Power (Faul et al., 2007), we investigated how the use of different numeracy measures influences the power of a study to find a numeracy effect. Based on the observed effect sizes, G\*Power was used to calculate the sample sizes required to find each of the predictive validity using each of the measures with and without controlling for vocabulary and Raven’s matrices. To estimate required sample sizes for the framing task, we used the effect size for the successful stimulus. The required sample sizes were calculated with a target power of .90 and an α-level of .05.

**Table S8.** Required Sample Sizes for Each Measure to Attain Significance based on Power Analyses (90% power at α-level = .05)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Task** |  | **A-NUM** | **4-NUM** | **1-NUM** | **BNT** | **Weller** |
| Probability Interpretation | No Covariates | 88 | 119 | 53 | 124 | 95 |
| With Covariates | 99 | 148 | 65 | 159 | 106 |
| Benefit Perception | No Covariates | 806 | 928 | 709 | 443 | 443 |
| With Covariates | 928 | 1,082 | 1,284 | 446 | 424 |
| Framing Effect\* | No Covariates | 576 | 1,305 | 1,743 | 1,305 | 649 |
| With Covariates | 649 | 1,305 | 2,619 | 1,493 | 692 |
| Bets Task | No Covariates | 1,155 | 645 | 1,493 | 1,305 | 2,619 |
| With Covariates | 1,153 | 742 | 1,493 | 1,153 | 2,619 |
| Risk Consistency | No Covariates | 396 | 268 | 449 | 610 | 301 |
| With Covariates | 1,159 | 545 | 947 | 2,093 | 742 |
| Maximum Required Sample | No Covariates | 1,155 | 1,305 | 1,743 | 1,305 | 2,619 |
| With Covariates | 1,159 | 1,305 | 2,619 | 2,093 | 2,619 |
| Average Required Sample | No Covariates | 604.2 | 653.0 | 889.4 | 757.4 | 821.4 |
| With Covariates | 797.6 | 764.4 | 1,281.6 | 1.068.8 | 916.6 |

*\*Based on effect on evaluation of the individual student with highest average grade (see Appendix E)*

**Appendix G: Discriminant Validity With Big 5 Personality Traits**

Some researchers have argued that personality includes intelligence (DeYoung, 2011; DeYoung, Quilty, Peterson, & Gray, 2014). Openness to experience has demonstrated consistent positive associations with general intelligence (r = .30). We hypothesize that the new numeracy measures would be positively associated with openness to experience (sometimes referred to as intellect[[1]](#footnote-2)) while being relatively unassociated with extraversion, conscientiousness, emotional stability, and agreeableness, consistent with previous findings (Cokely et al., 2012).

*Big 5 Personality Traits.* Participants responded to the 30 items of The Big Five Inventory–2 Short Form (BFI-2-S) using a 5-point Likert scale from 1 (Disagree strongly) to 5 (Agree strongly; Soto & John, 2017). Items for each personality trait were averaged together.

The new numeracy measures were weakly positively correlated with openness and was generally unrelated with the other personality factors with extraversion and conscientiousness as occasional, unexpected exceptions (correlations that were also shared by the established measures; see Table S9 for those results). Interestingly, participants lower in neuroticism (negative emotionality) reported lower subjective numeracy and numeric self-efficacy than those higher in neuroticism; neuroticism was unrelated to numeracy. It may be that those higher in neuroticism have more negative self-views generally, irrespective of their actual level of ability.

**Table S9.** Means, Standard Deviations, and Correlations for all Constructs used for Convergent/Discriminant Validity

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Numeracy measures |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1. A-NUM (Categories) | 4.70 | 1.49 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2. 4-NUM | 1.84 | 1.17 | .72\*\* |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3. 1-NUM | 0.48 | 0.50 | .75\*\* | .58\*\* |  |  |  |  |  |  |  |  |  |  |  |  |
| 4. BNT | 2.32 | 1.19 | .63\*\* | .66\*\* | .53\*\* |  |  |  |  |  |  |  |  |  |  |  |
| 5. Weller | 5.30 | 1.98 | .71\*\* | .66\*\* | .52\*\* | .60\*\* |  |  |  |  |  |  |  |  |  |  |
| 6. Subjective Numeracy | 4.61 | 0.96 | .44\*\* | .39\*\* | .38\*\* | .34\*\* | .40\*\* |  |  |  |  |  |  |  |  |  |
| 7. Numeric Self-Efficacy | 4.38 | 1.23 | .44\*\* | .42\*\* | .37\*\* | .36\*\* | .40\*\* | .91\*\* |  |  |  |  |  |  |  |  |
| 8. Numeric Preference | 4.84 | 0.96 | .31\*\* | .24\*\* | .29\*\* | .23\*\* | .28\*\* | .84\*\* | .53\*\* |  |  |  |  |  |  |  |
| Intelligence measures |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9. Ravens Matrices | 5.34 | 1.70 | .38\*\* | .36\*\* | .26\*\* | .30\*\* | .39\*\* | .20\*\* | .15\*\* | .20\*\* |  |  |  |  |  |  |
| 10. Vocabulary | 6.49 | 2.03 | .35\*\* | .29\*\* | .22\*\* | .26\*\* | .31\*\* | .14\*\* | .13\*\* | .11\*\* | .15\*\* |  |  |  |  |  |
| Big-5 personality measuresmeasures |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 11. Openness | 3.83 | 0.88 | .11\*\* | .14\*\* | .07 | .07 | .15\*\* | .13\*\* | .14\*\* | .09\* | .10\* | .25\*\* |  |  |  |  |
| 12. Conscientiousness | 3.97 | 0.84 | -.03 | -.11\* | -.04 | -.10\* | -.07 | .17\*\* | .17\*\* | .11\*\* | -.00 | .09\* | .23\*\* |  |  |  |
| 13. Extraversion | 3.03 | 0.93 | -.10\* | -.10\* | -.08 | -.09\* | -.09\* | .22\*\* | .25\*\* | .12\*\* | -.09\* | -.01 | .34\*\* | .47\*\* |  |  |
| 14. Agreeableness | 3.85 | 0.78 | .00 | -.02 | -.01 | -.05 | -.04 | .06 | .08 | .03 | -.01 | .06 | .21\*\* | .43\*\* | .30\*\* |  |
| 15. Neuroticism | 2.41 | 1.03 | -.03 | -.03 | -.04 | -.03 | -.04 | -.18\*\* | -.24\*\* | -.05 | .03 | -.09\* | -.16\*\* | -.52\*\* | -.52\*\* | -.43\*\* |

*Note.* *M* and *SD* are used to represent mean and standard deviation, respectively. \* indicates *p* < .05. \*\* indicates *p* < .01. A-NUM: Adaptive Numeric Understanding Measure; 4-NUM: Four-Item Numeric Understanding Measure; 1-NUM: Single-Item Numeric Understanding Measure; BNT: Berlin Numeracy Test; Weller: Rasch-Based Numeracy Scale

**References**

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. Judgment and Decision Making, 7(1), 25–47.

DeYoung, C. G. (2011). Intelligence and personality. In R. J. Sternberg & S. B. Kaufman. (Eds.), *The Cambridge handbook of intelligence* (pp. 711–737). New York, NY: Cambridge University Press.

DeYoung, C. G., Quilty, L. C., Peterson, J. B., & Gray, J. R. (2014). Openness to experience, intellect, and cognitive ability. *Journal of Personality Assessment, 96*, 46 –52. doi:10.1080/00223891.2013.806327

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <https://doi.org/10.3758/BF03193146>

Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, *39*(3), 167–188. <https://doi.org/10.1177/0146621614554650>

Magis, D., Raiche, G., & Magis, M. D. (2018). Package ‘catR.’ *R Package, Version*, *3*.

Nydick, S. W., & Nydick, M. S. W. (2013). Package ‘catIrt.’ *R Package, Version*, *3*.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*(5), 407–413.

1. The preregistration mistakenly said conscientiousness [↑](#footnote-ref-2)