**Supplementary Online Content**

**Development and Application of Novel Performance Validity Metrics for Computerized Neurocognitive Batteries**

**Supplementary Methods**

**Supplementary Results**

**eTable 1** Tests and domains in the Penn Computerized Neurocognitive Battery (PennCNB)

**eTable 2** Proportions correct for each item for the six PennCNB tests.

**eTable 3** 2PL model item parameters (logistic metric) for six neurocognitive tests used for simulation.

**eFigure 1** Prediction Accuracy (AUC) for the Multivariate CNB Performance Validity Estimate in Predicting True (Simulated) Careless Responding in 10% of Simulated Sample, by Proportions of Valid Responses.

**eFigure 2** Changes in the reliability (Cronbach’s alpha) of the PennCNB tests as the number of invalid responses increases (for 5% of the sample).

**eFigure 3** Receiver Operating Characteristic (ROC) Curve Analysis Predicting Manually Determined Validation Using Multivariate CNB Validity Estimates: Results from the Philadelphia Neurodevelopmental Cohort.

**eFigure 4** Receiver Operating Characteristic (ROC) Curve Analysis Predicting Manually Determined Validation Using Multivariate CNB Validity Estimates: Results from the Marine Resiliency Study-II

**eFigure 5** Sensitivity and specificity for various thresholds for the Multivariate CNB Validity Estimate (SMVE) for the Philadelphia Neurodevelopmental Cohort (PNC) using tests from the Penn Computerized Neurocognitive Battery.

**eFigure 6** Sensitivity and specificity for various thresholds for the Multivariate CNB Validity Estimate (SMVE) for the Marine Resiliency Study-II (MRS-II) using tests from the Penn Computerized Neurocognitive Battery.

**eFigure 7** Examination of Multivariate CNB Validity Estimates by Clinical Diagnoses: Results from the Philadelphia Neurodevelopmental Cohort.

**eFigure 8** Examination of Multivariate CNB Validity Estimates by Clinical Diagnoses: Results from the Marine Resiliency Study-II.

This supplementary material has been provided by the authors to give readers additional information about their work.

**SUPPLEMENTARY METHODS**

**Participants**

Calkins et al. (2015) provides extensive information on recruitment, enrollment, and procedures for the Philadelphia Neurodevelopmental Cohort (PNC). Briefly, PNC participants were drawn from a pool of 50,293 youths recruited and genotyped by the Center for Applied Genomics at the Children's Hospital of Philadelphia (CHOP) through a pediatric healthcare network of over 30 clinical community sites in the tristate area of Pennsylvania, New Jersey, and Delaware. When undergoing any blood work, patients were approached for participation in the Center for Applied Genomics recruitment pool. Participants provided a blood sample for genomic studies and access to Electronic Medical Records (EMRs). To determine preliminary eligibility for PNC participation, the EMR of each participant was screened by research staff. Potential participants were included if they were between the ages of 8–21 (though a few participants turned 22 by the time they underwent study procedures), had provided written informed consent/assent to be re-contacted for future studies, were proficient in English, and did not appear to have significant developmental delays or physical conditions that would interfere with their ability to complete study procedures. This screening yielded a pool of 19,161 eligible individuals. Participants from this pool were selected at random for invitations to participate after stratification by sex, age, and race/ethnicity. From the recruitment pool, 13,598 participants were invited (71.0% of EMR eligible), 9,498 were enrolled (64.3% of invited), and 9,421 completed the assessments. Inclusion criteria were intentionally liberal to extend generalizability (Calkins et al., 2015). Participants and their guardians (for participants <18 yrs) provided written informed consent or assent. The Institutional Review Boards at the University of Pennsylvania (Penn) and CHOP approved the protocol.

Psychopathology was assessed using a computerized, structured interview, with screen-level symptom and episode information (Calkins et al., 2015). The instrument is abbreviated and modified from the epidemiologic version of the NIMH Genetic Epidemiology Research Branch Kiddie-SADS (Merikangas et al., 2009). The screen assessed psychiatric and psychological treatment history, and lifetime occurrence of major domains of psychopathology including mood (major depressive episode, manic episode), anxiety (generalized anxiety disorder, separation anxiety disorder, specific phobia, social phobia, panic disorder, agoraphobia, obsessive-compulsive disorder, posttraumatic stress disorder), attention deficit/hyperactivity disorder (ADHD), behavioral (oppositional defiant disorder, conduct disorders), eating disorders (anorexia, bulimia), psychosis, and suicidal thinking and behavior. Each section included a screen for relevant symptoms and additional DSM-IV criteria, including symptom frequency and duration of symptoms and episodes. For all symptoms endorsed that met threshold to continue within the section, the participant was asked to rate associated distress and impairment on separate 11-point scales ranging from 0 (no bother/problems) to 10 (extremely serious bother/problems). To generate psychopathology summary variables, algorithms assigned rankings for DSM-IV indices of each psychopathology domain based on endorsement of contributing items. Psychopathology domains were considered significant if sufficient symptoms were endorsed with frequency and duration approximating DSM-IV disorder or episode criteria, accompanied by significant distress or impairment rated ≥ 5.

The participants in Marine Resiliency Study-II (MRS-II) have been described in detail previously (Acheson et al., 2015). Briefly, 1,444 U.S. Servicemembers (3 battalions; mean age = 21.9 ± 2.7 years, range 18 to 43) were assessed before and after a 7-month deployment to Afghanistan. Here, we used only baseline neurocognitive data from the MRS-II. The institutional review boards of the University of California San Diego, Penn, VA San Diego Research Service, and Naval Health Research Center approved the study. Written informed consent was obtained from participants.

For the MRS-II, psychopathology was assessed using the Clinician Administered PTSD Scale (CAPS) (Blake et al., 1995); PTSD Checklist (PCL) (Weathers et al., 1991); Beck Depression Inventory 2 (BDI-2) (Beck et al., 1996); Beck Anxiety Inventory (BAI) (Steer & Beck, 1997); Alcohol Use Disorders Identification Test (AUDIT) (Saunders et al., 1993); SF-12 Health Survey (Ware et al., 1996); Interpersonal Reactivity Index (IRI) (Davis, 1980, 1983); Army STARRS New Soldier Study Sleep Scale (NSSS) (Kessler et al., 2013), and Deployment Risk and Resilience Inventory-2 (DRRI) (Vogt et al., 2013).

**Person-Fit Metrics**

The point-biserial method capitalizes on the fact that a vector of item responses (0/1, incorrect/correct) will correlate positively with the vector of sample-level proportions correct. Thus, responses on items with a low proportion correct (difficult) will tend to be 0 (incorrect), whereas responses on items with a high proportion correct (easy) will tend to be 1 (correct). Consider a 4-item test where the sample-level proportions correct are {0.80, 0.60, 0.40, 0.20}. A response vector of {1, 1, 0, 0} would correlate positively (rpb=0.89) with the vector of proportions correct, indicating good person-fit. A response vector of {0, 0, 1, 1} would correlate negatively, indicating poor person-fit and capturing highly unusual incorrect responses to easy items and correct responses to hard items. The examinee with the negative point-biserial correlation (poor fit) would be flagged for suspicious responding.

The Kane and Brennan (1980) dependability measure, E.KB, is defined by the following equation for a test with *i* items:

where Xi is the response (0/1, incorrect/correct) to item *i*, and pi is the sample-level proportion correct for item *i*. Each item response is multiplied by its corresponding proportion correct, and these products are summed to get the numerator of Equation 1. The denominator is the maximum possible value of the numerator, given the number correct. Taking the example above, the first response vector {1, 1, 0, 0} would receive an E-KB of

because the sum of products in the numerator is the largest possible, given 2 correct responses. The second response vector {0, 0, 1, 1} from the paragraph above would receive an E.KB of

indicating worse fit to the model.

**Current PennCNB Validation Rules**

As described in the main document, after test administration, flags for potential invalid data are generated by algorithmic validity rules and assessor judgment (e.g., when an examinee clearly exhibits disengagement from a task). Experienced PennCNB data validators supervised by neuropsychologists make ultimate decisions to designate a test as invalid by integrating algorithmic rules, assessor codes and comments, and visual data inspection.

Each test described in the main paper has a specific set of algorithmic rules to flag data for review. These rules were developed by modeling and visual inspection of response patterns across all available data on the PennCNB (in pediatric and adult samples) in 2012-2013 and identifying unlikely response patterns for given samples. For the CPW and the CPF, results are algorithmically flagged if greater than 50% of responses are less than 2000 ms or the longest run of same responses is greater than 20 trials (50% of the number of recall trials). For the VOLT, results are algorithmically flagged if greater than 50% of responses are less than 2000 ms or the longest run of same responses is greater than 10 trials (50% of the number of recall trials). For the PMAT, results are algorithmically flagged if greater than 50% of responses are less than 2000 ms or the total number of correct responses is less than 5. For the PVRT, results are algorithmically flagged if the longest run of same responses is greater than 4 or if greater than two responses are less than 750 ms. For the ER40, results are algorithmically flagged if total correct responses are less than 12, correct responses to “happy” faces are less than 6, if the longest run of same responses is greater than 11, or if greater than 50% of responses are under 200 ms.

**Simulations**

In step 1, we estimated population parameters from the PNC sample. In real data from the PNC, items were calibrated using the 2-parameter logistic IRT model (2PLM), and these item parameter estimates (discrimination and difficulty) were used as the population parameters for the simulations. For example, population parameters based on the PMAT were for 24 items with a wide range of difficulty, in contrast to the population parameters based on the CPW, which were for 40 items with relatively low difficulty. Calibration of items was performed using the mirt() function in the *mirt* package (Chalmers, 2012). In each simulation, the program checked for convergence problems in mirt(), such as a direct “no convergence” warning and/or impossibly high/low parameter estimates. If such a problem was found, it re-estimated the parameters using the irt.fa() function in the *psych* package (Revelle, 2018) and re-scaled to the logistic metric to match mirt() (from D = 1.702 to D = 1.000).

**SUPPLEMENTARY RESULTS**

Figure 1 displays the relationship between simulated effort (number of valid responses across items) and the ability of the model to classify response patterns as valid or invalid. The non-monotonicity of the functions in Figure 1—i.e., the lack of smoothness along each step of the x-axis—is due to the fact that the item ordering in the simulations was preserved (same as the actual tests), and item order affects some components of the MCVE. For example, if the three easiest items on a test are the first three items, then any invalid responding beyond the first three items will not be informative; this is the case with the PMAT (see Figure 1), where the AUC for the easiest items method drops to chance level (0.50) after only three items. An alternative explanation for the non-monotonicity of the functions in Figure 1 is that we did not run enough simulations (leaving too much noise), but we confirmed this was not the case by varying the number of simulations; the asymptotic shape of the functions is not smooth. The functions could be made monotonic by permuting the order of items for each simulation, but in reality, item order will always affect these validity estimates. We therefore preserved item order to better reflect what will be encountered in actual practice.

eFigures 4 and 5 show how sensitivity and specificity vary as each individual test threshold is varied. Note that, to optimize precision for this specific exercise, these associations of threshold with sensitivity and specificity were calculated within-sample. This is different than all AUCs reported in the main paper, which were calculated out-of-sample (i.e., cross-validated).

**SUPPLEMENTARY REFERENCES**

Acheson, D. T., Geyer, M. A., Baker, D. G., Nievergelt, C. M., Yurgil, K., Risbrough, V. B., & MRS-II Team. (2015). Conditioned fear and extinction learning performance and its association with psychiatric symptoms in active duty Marines. *Psychoneuroendocrinology*, *51*, 495–505. https://doi.org/10.1016/j.psyneuen.2014.09.030

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory-II*. Harcourt Brace.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, *8*(1), 75–90. https://doi.org/10.1002/jts.2490080106

Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., Ruparel, K., Wolf, D. H., Roalf, D. R., Mentch, F. D., Qiu, H., Chiavacci, R., Connolly, J. J., Sleiman, P. M. A., Gur, R. C., Hakonarson, H., & Gur, R. E. (2015). The Philadelphia Neurodevelopmental Cohort: Constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *56*, 1356–1369. https://doi.org/10.1111/jcpp.12416

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29.

Davis, M. H. (1980). *A multidimensional approach to individual differences in empathy*.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126.

Kessler, R. C., Colpe, L. J., Fullerton, C. S., Gebler, N., Naifeh, J. A., Nock, M. K., Sampson, N. A., Schoenbaum, M., Zaslavsky, A. M., & Stein, M. B. (2013). Design of the army study to assess risk and resilience in servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research*, *22*(4), 267–275.

Merikangas, K. R., Avenevoli, S., Costello, E. J., Koretz, D., & Kessler, R. C. (2009). National Comorbidity Survey Replication Adolescent Supplement (NCS-A): I. Background and Measures. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(4), 367–379. https://doi.org/10.1097/CHI.0b013e31819996f1

Revelle, W. (2018). *psych: Procedures for personality and psychological research* (1.8.10) [Computer software]. Northwestern University. https://CRAN.R-project.org/package=psych

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction*, *88*(6), 791–804.

Steer, R. A., & Beck, A. T. (1997). Beck Anxiety Inventory. In *Evaluating stress: A book of resources* (pp. 23–40). Scarecrow Education.

Vogt, D., Smith, B. N., King, L. A., King, D. W., Knight, J., & Vasterling, J. J. (2013). Deployment risk and resilience inventory-2 (DRRI-2): An updated tool for assessing psychosocial risk and resilience factors among service members and veterans. *Journal of Traumatic Stress*, *26*(6), 710–717.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, 220–233.

Weathers, F. W., Huska, J. A., & Keane, T. M. (1991). *PCL-M for DSM-IV*. National Center for PTSD - Behavioral Science Division.

eTable 1. Tests and domains in the core Penn Computerized Neurocognitive Battery (PennCNB)

|  |  |  |
| --- | --- | --- |
| **Neurocognitive Domain** | **Cognitive Functions Assessed** | **Test** |
| Executive Control | Sustained Attention | Penn Continuous Performance Test (PCPT) |
|  | Working Memory | Letter N-Back Test (LNB) |
|  | Abstraction/Mental Flexibility | Penn Conditional Exclusion Test (PCET) |
| Episodic Memory | Verbal Episodic Memory | Penn Word Memory Test (CPW) |
|  | Face Memory | Penn Face Memory Test (CPF) |
|  | Spatial Episodic Memory | Visual Object Learning Test (VOLT) |
| Complex Cognition | Language Reasoning | Penn Verbal Reasoning Test (PVRT) |
|  | Nonverbal Reasoning | Penn Matrix Reasoning Test (PMAT) |
|  | Visuospatial Ability | Penn Line Orientation Test (PLOT) |
| Social Cognition | Emotion Identification | Penn Emotion Recognition Test (ER40) |
|  | Emotion Differentiation | Penn Emotion Differentiation Test (PEDT) |
|  | Age Differentiation | Penn Age Differentiation Test (PADT) |
| Sensorimotor | Motor Speed | Computerized Finger Tapping Test (CTAP) |
|  | Motor Praxis | Motor Praxis Test (MPT) |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| eTable 2. Proportions correct for each item for the six PennCNB tests. | | | | | | |
| Item | PMAT | VOLT | ER40 | CPF | CPW | PVRT |
| 1 | 0.962 | 0.973 | 0.423 | 0.744 | 0.819 | 0.812 |
| 2 | 0.909 | 0.952 | 0.376 | 0.237 | 0.962 | 0.557 |
| 3 | 0.973 | 0.808 | 0.624 | 0.850 | 0.984 | 0.672 |
| 4 | 0.938 | 0.793 | 0.912 | 0.909 | 0.979 | 0.888 |
| 5 | 0.930 | 0.763 | 0.947 | 0.787 | 0.981 | 0.875 |
| 6 | 0.722 | 0.863 | 0.776 | 0.876 | 0.956 | 0.495 |
| 7 | 0.971 | 0.891 | 0.705 | 0.628 | 0.949 | 0.750 |
| 8 | 0.564 | 0.619 | 0.925 | 0.775 | 0.948 | 0.802 |
| 9 | 0.637 | 0.815 | 0.988 | 0.816 | 0.859 | 0.850 |
| 10 | 0.520 | 0.756 | 0.956 | 0.693 | 0.964 | 0.961 |
| 11 | 0.431 | 0.516 | 0.997 | 0.537 | 0.919 | 0.335 |
| 12 | 0.361 | 0.957 | 0.998 | 0.801 | 0.949 | 0.802 |
| 13 | 0.549 | 0.709 | 0.894 | 0.686 | 0.873 | 0.745 |
| 14 | 0.528 | 0.754 | 0.897 | 0.952 | 0.958 | 0.766 |
| 15 | 0.464 | 0.651 | 0.895 | 0.779 | 0.916 | 0.876 |
| 16 | 0.606 | 0.660 | 0.864 | 0.555 | 0.949 |  |
| 17 | 0.512 | 0.585 | 0.733 | 0.954 | 0.979 |  |
| 18 | 0.651 | 0.794 | 0.729 | 0.933 | 0.964 |  |
| 19 | 0.520 | 0.685 | 0.802 | 0.929 | 0.900 |  |
| 20 | 0.480 | 0.960 | 0.914 | 0.778 | 0.886 |  |
| 21 | 0.567 |  | 0.496 | 0.861 | 0.974 |  |
| 22 | 0.445 |  | 0.107 | 0.846 | 0.923 |  |
| 23 | 0.291 |  | 0.662 | 0.790 | 0.926 |  |
| 24 | 0.334 |  | 0.943 | 0.797 | 0.885 |  |
| 25 |  |  | 0.829 | 0.606 | 0.893 |  |
| 26 |  |  | 0.703 | 0.838 | 0.936 |  |
| 27 |  |  | 0.924 | 0.883 | 0.927 |  |
| 28 |  |  | 0.937 | 0.719 | 0.935 |  |
| 29 |  |  | 0.882 | 0.890 | 0.886 |  |
| 30 |  |  | 0.995 | 0.513 | 0.889 |  |
| 31 |  |  | 0.996 | 0.727 | 0.872 |  |
| 32 |  |  | 0.955 | 0.775 | 0.822 |  |
| 33 |  |  | 0.906 | 0.910 | 0.904 |  |
| 34 |  |  | 0.722 | 0.601 | 0.938 |  |
| 35 |  |  | 0.871 | 0.562 | 0.888 |  |
| 36 |  |  | 0.723 | 0.677 | 0.902 |  |
| 37 |  |  | 0.499 | 0.778 | 0.947 |  |
| 38 |  |  | 0.637 | 0.571 | 0.930 |  |
| 39 |  |  | 0.801 | 0.911 | 0.823 |  |
| 40 |  |  | 0.939 | 0.499 | 0.823 |  |

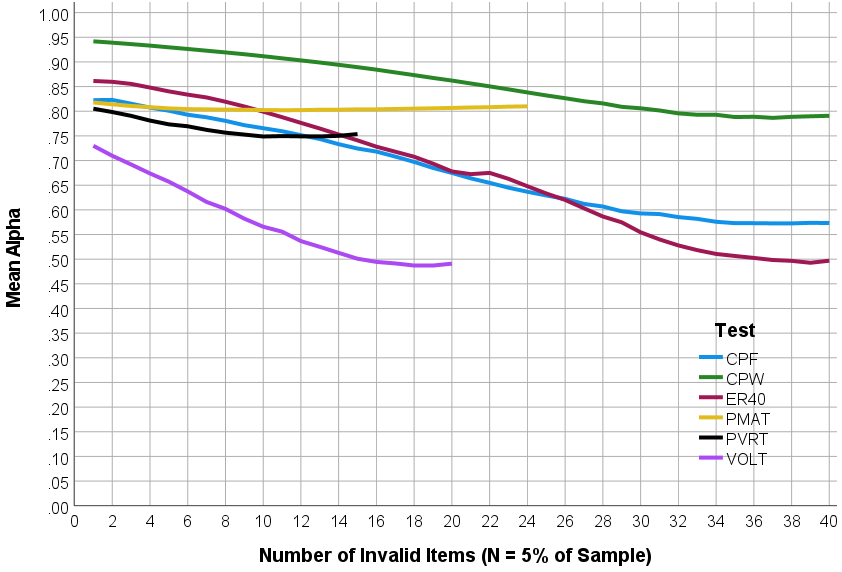
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| eTable 3. 2PL model item parameters (logistic metric) for six neurocognitive tests used for simulation. | | | | | | | | | | | | | |
|  | PMAT | | VOLT | | | ER40 | | CPF | | CPW | | PVRT | |
| Item | Disc (a) | Diff (b) | Disc (a) | Diff  (b) | Disc (a) | | Diff  (b) | Disc (a) | Diff  (b) | Disc (a) | Diff (b) | Disc (a) | Diff (b) |
| 1 | 1.76 | -2.56 | 0.00 | 81.28 | 0.03 | | 9.83 | 0.83 | -1.46 | 0.41 | -3.78 | 0.95 | -1.80 |
| 2 | 1.97 | -1.79 | 0.02 | -178.87 | 0.09 | | 5.98 | 0.72 | 1.79 | 2.43 | -2.20 | 1.23 | -0.26 |
| 3 | 1.38 | -3.21 | 0.96 | -1.76 | 0.22 | | -2.36 | 0.39 | -4.59 | 0.97 | -4.71 | 0.73 | -1.11 |
| 4 | 1.49 | -2.40 | 0.29 | -4.72 | 0.43 | | -5.65 | 0.13 | -18.40 | 2.64 | -2.45 | 1.66 | -1.76 |
| 5 | 1.62 | -2.18 | 0.06 | -18.28 | 1.22 | | -2.87 | 0.02 | -73.80 | 1.33 | -3.59 | 1.18 | -2.03 |
| 6 | 1.37 | -0.93 | 1.24 | -1.87 | 0.32 | | -3.97 | 0.00 | 17.93 | 2.34 | -2.14 | 0.33 | 0.06 |
| 7 | 0.92 | -4.21 | 1.67 | -1.78 | 0.24 | | -3.67 | 0.90 | -0.68 | 1.84 | -2.27 | 1.74 | -0.97 |
| 8 | 0.72 | -0.38 | 0.28 | -1.75 | 0.98 | | -2.95 | 0.00 | 4.24 | 2.06 | -2.14 | 2.27 | -1.08 |
| 9 | 0.88 | -0.70 | 0.46 | -3.37 | 0.53 | | -8.63 | 0.30 | -5.09 | 0.48 | -3.93 | 1.66 | -1.50 |
| 10 | 0.81 | -0.05 | 0.00 | 4.42 | 0.59 | | -5.52 | 0.00 | 7.27 | 0.82 | -4.38 | 2.24 | -2.22 |
| 11 | 1.32 | 0.37 | 0.97 | -0.08 | 1.42 | | -4.80 | 0.61 | -0.26 | 1.25 | -2.41 | 1.19 | 0.73 |
| 12 | 1.26 | 0.75 | 2.32 | -2.15 | 1.52 | | -4.81 | 0.86 | -1.85 | 0.59 | -5.25 | 1.67 | -1.23 |
| 13 | 1.09 | 0.10 | 0.07 | -12.90 | 2.06 | | -1.63 | 1.02 | -0.92 | 0.53 | -3.82 | 1.75 | -0.94 |
| 14 | 1.06 | 0.31 | 1.13 | -1.23 | 1.99 | | -1.68 | 0.16 | -19.23 | 2.62 | -2.08 | 0.86 | -1.59 |
| 15 | 1.13 | 0.72 | 0.20 | -3.17 | 1.71 | | -1.79 | 0.95 | -1.56 | 1.61 | -2.03 | 1.39 | -1.84 |
| 16 | 1.66 | 0.39 | 1.10 | -0.75 | 1.54 | | -1.65 | 0.94 | -0.28 | 0.63 | -4.92 |  |  |
| 17 | 0.94 | 0.82 | 0.73 | -0.53 | 0.17 | | -6.17 | 1.46 | -2.68 | 0.94 | -4.49 |  |  |
| 18 | 1.43 | 0.48 | 1.09 | -1.52 | 0.00 | | 13.42 | 1.11 | -2.82 | 2.45 | -2.22 |  |  |
| 19 | 1.51 | 1.04 | 0.08 | -10.14 | 0.40 | | -3.58 | 0.00 | 13.88 | 0.57 | -4.09 |  |  |
| 20 | 0.89 | 1.33 | 2.21 | -2.25 | 0.42 | | -5.87 | 0.08 | -16.22 | 1.37 | -1.94 |  |  |
| 21 | 1.47 | 1.13 |  |  | 0.00 | | -0.16 | 0.38 | -4.93 | 0.95 | -4.26 |  |  |
| 22 | 1.79 | 1.58 |  |  | 0.00 | | -27.34 | 0.07 | -22.92 | 0.62 | -4.24 |  |  |
| 23 | 1.81 | 2.14 |  |  | 0.16 | | -4.13 | 1.26 | -1.35 | 1.80 | -2.02 |  |  |
| 24 | 1.70 | 2.19 |  |  | 0.40 | | -7.27 | 1.08 | -1.54 | 0.64 | -3.44 |  |  |
| 25 |  |  |  |  | 0.48 | | -3.42 | 0.00 | 6.78 | 0.67 | -3.42 |  |  |
| 26 |  |  |  |  | 0.33 | | -2.64 | 0.99 | -1.96 | 2.10 | -1.97 |  |  |
| 27 |  |  |  |  | 0.83 | | -3.36 | 1.31 | -1.97 | 1.70 | -2.09 |  |  |
| 28 |  |  |  |  | 0.88 | | -3.44 | 0.00 | 9.27 | 1.85 | -2.09 |  |  |
| 29 |  |  |  |  | 0.27 | | -7.62 | 0.21 | -9.98 | 0.62 | -3.54 |  |  |
| 30 |  |  |  |  | 0.94 | | -6.12 | 0.84 | -0.07 | 1.65 | -1.77 |  |  |
| 31 |  |  |  |  | 1.29 | | -5.03 | 0.04 | -28.30 | 0.56 | -3.65 |  |  |
| 32 |  |  |  |  | 0.18 | | -16.78 | 0.94 | -1.54 | 0.40 | -3.97 |  |  |
| 33 |  |  |  |  | 1.52 | | -2.01 | 1.26 | -2.29 | 1.81 | -1.81 |  |  |
| 34 |  |  |  |  | 1.06 | | -1.10 | 0.00 | 9.75 | 2.09 | -2.01 |  |  |
| 35 |  |  |  |  | 1.23 | | -1.94 | 0.00 | 5.94 | 1.20 | -2.13 |  |  |
| 36 |  |  |  |  | 1.27 | | -0.98 | 0.00 | 3.44 | 0.59 | -4.00 |  |  |
| 37 |  |  |  |  | 0.13 | | 0.02 | 0.88 | -1.64 | 1.94 | -2.19 |  |  |
| 38 |  |  |  |  | 0.00 | | 6.81 | 0.66 | -0.48 | 1.74 | -2.09 |  |  |
| 39 |  |  |  |  | 0.22 | | -6.27 | 1.16 | -2.43 | 0.48 | -3.34 |  |  |
| 40 |  |  |  |  | 0.56 | | -5.09 | 0.00 | -0.02 | 0.57 | -2.86 |  |  |

eFigure 1. Prediction Accuracy (AUC) for the Multivariate CNB Performance Validity Estimate in Predicting True (Simulated) Careless Responding in 10% of Simulated Sample, by Proportions of Valid Responses.

**Chart, scatter chart

Description automatically generated**

eFigure 2. Changes in the reliability (Cronbach’s alpha) of the PennCNB tests as the number of invalid responses increases (for 5% of the sample).

****

eFigure 3. Receiver Operating Characteristic (ROC) Curve Analysis Predicting Manually Determined Validation Using Multivariate CNB Validity Estimates: Results from the Philadelphia Neurodevelopmental Cohort.A close up of a map

Description automatically generated

*Note.* CPF = Penn Face Memory Test; CPW = Penn Word Memory Test; VOLT = Visual Object Learning Test; ER40 = Penn Emotion Recognition Test; PMAT = Penn Matrix Analysis Test; PVRT = Penn Verbal Reasoning Test.

eFigure 4. Receiver Operating Characteristic (ROC) Curve Analysis Predicting Manually Determined Validation Using Multivariate CNB Validity Estimates: Results from the Marine Resiliency Study-II

Chart, line chart

Description automatically generated

*Note.* CPF = Penn Face Memory Test; CPW = Penn Word Memory Test; VOLT = Visual Object Learning Test; ER40 = Penn Emotion Recognition Test; PMAT = Penn Matrix Analysis Test; PVRT = Penn Verbal Reasoning Test.

eFigure 5. Sensitivity and specificity for various thresholds for the Multivariate CNB Validity Estimate (MCVE) for the Philadelphia Neurodevelopmental Cohort (PNC) using tests from the Penn Computerized Neurocognitive Battery

A close up of a device

Description automatically generated

eFigure 6. Sensitivity and specificity for various thresholds for the Multivariate CNB Validity Estimate (MCVE) for the Marine Resiliency Study-II (MRS-II) using tests from the Penn Computerized Neurocognitive Battery

**A close up of text on a white background

Description automatically generated**

eFigure 7.Examination of Multivariate CNB Validity Estimates by Clinical Diagnoses: Results from the Philadelphia Neurodevelopmental Cohort.

Chart, box and whisker chart

Description automatically generated with medium confidence

*Note.* CPF = Penn Face Memory Test; CPW = Penn Word Memory Test; VOLT = Visual Object Learning Test; ER40 = Penn Emotion Recognition Test; PMAT = Penn Matrix Analysis Test; PVRT = Penn Verbal Reasoning Test.

eFigure 8.Examination of Multivariate CNB Validity Estimates by Clinical Diagnoses: Results from the Marine Resiliency Study-II.

Graphical user interface, chart, application, table, Excel, box and whisker chart

Description automatically generated

*Note.* CPF = Penn Face Memory Test; CPW = Penn Word Memory Test; VOLT = Visual Object Learning Test; ER40 = Penn Emotion Recognition Test; PMAT = Penn Matrix Analysis Test; PVRT = Penn Verbal Reasoning Test.