Published in final edited form as:

Assessment. 2014 December ; 21(6): 669–678. doi:10.1177/1073191114524270.

Development of Abbreviated Eight-Item Form of the Penn Verbal Reasoning Test

Warren B. Bilker¹, Michael R. Wierzbicki¹, Colleen M. Brensinger¹, Raquel E. Gur¹, and Ruben C. Gur¹

¹University of Pennsylvania, Philadelphia, PA, USA

Abstract

The ability to reason with language is a highly valued cognitive capacity that correlates with IQ measures and is sensitive to damage in language areas. The Penn Verbal Reasoning Test (PVRT) is a 29-item computerized test for measuring abstract analogical reasoning abilities using language. The full test can take over half an hour to administer, which limits its applicability in large-scale studies. We previously described a procedure for abbreviating a clinical rating scale and a modified procedure for reducing tests with a large number of items. Here we describe the application of the modified method to reducing the number of items in the PVRT to a parsimonious subset of items that accurately predicts the total score. As in our previous reduction studies, a split sample is used for model fitting and validation, with cross-validation to verify results. We find that an 8-item scale predicts the total 29-item score well, achieving a correlation of .9145 for the reduced form for the model fitting sample and .8952 for the validation sample. The results indicate that a drastically abbreviated version, which cuts administration time by more than 70%, can be safely administered as a predictor of PVRT performance.

Keywords

Penn Verbal Reasoning Test; scale reduction; predictive model

Verbal reasoning capacity is an important cognitive domain that is often assessed using analogy problems (e.g., Ekstrom, French, & Harman, 1979). As stated in a recent evaluation of a large longitudinal data set, “Analogical reasoning is a core cognitive skill that distinguishes humans from all other species and contributes to general fluid intelligence, creativity, and adaptive learning capacities” (Richland & Burchinal, 2013). Clinically, deficits in verbal reasoning have been linked to damage in left temporoparietal cortex and are useful in predicting recovery (e.g., Willner et al., 1976). Therefore, tests of verbal analogical reasoning are often included in neuropsychological testing (e.g., Lezak, Howieson, Bigler, & Tranel, 2012).

© The Author(s) 2014

Corresponding Author: Warren B. Bilker, Department of Biostatistics, Perelman School of Medicine at the University of Pennsylvania, 601 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, USA. warren@upenn.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

With the advent of large-scale clinical and genomic studies and increased demand for comprehensive neurocognitive assessments to be used as endophenotypes (Green et al., 2004; R.C. Gur et al., 2011; Insel & Cuthbert, 2009; McClay et al., 2011; Reichenberg et al., 2010; Seidman et al., 2010), there has been increased need to abbreviate tests. Such efforts have been applied to several domains of cognition, resulting in abbreviated versions of specific tests (e.g., Calamia, Markon, Denburg, & Tranel, 2011; Winegarden, Yates, Moses, & Faustman, 1997) or abbreviated batteries that cover multiple domains (e.g., Cheng & Morgan, 2013; Harvey, Keefe, Patterson, Heaton, & Bowie, 2009; Hill et al., 2008; Hurford, Marder, Keefe, Reise, & Bilder, 2011; Mansbach & MacDougall, 2012; Ventura et al., 2010).

The Penn Verbal Reasoning Test (PVRT) is a multiple-choice test in which participants answer verbal analogy problems (R. C. Gur & Reivich, 1980). It has been administered in functional neuroimaging studies (R. C. Gur et al., 1982; R. C. Gur et al., 1983; R. C. Gur et al., 2000; R. C. Gur & Reivich, 1980; Roalf et al., 2013; Wharton et al., 2000), reliably producing left hemispheric temporoparietal activation, with performance linked to anatomical measures of these regions (R. C. Gur et al., 1999). The test has also been included in many large-scale genomic studies (Almasy et al., 2008; Calkins et al., 2010; R. C. Gur et al., 2012; R. E. Gur et al., 2007), demonstrating construct and criterion validity (R. C. Gur et al., 2001; R. C. Gur et al., 2010; R. C. Gur et al., 2011; R. C. Gur et al., 2012; Irani et al., 2012; Roalf et al., 2013) as well as heritability and genetic correlations (Almasy et al., 2008; Yokley et al., 2012). The items within the PVRT are multiple choice analogies, similar to “Kitten is to Cat as Puppy is to: a. Chicken; b. Cow; c. Dog; d. Milk.” The test can take more than half an hour to administer. It was originally composed of 30 items and was modified to a 29-item scale as 1 item was used for practice and never presented in tests. The total score on the test is the number of correct items. Notwithstanding Smith, McCarthy, and Anderson's (2000) call for caution in the application of abbreviated tests, abbreviation is justified for a central, narrowly defined domain and considering the need for efficient neurocognitive testing to provide endophenotypes in large-scale genomic studies.

We previously described a procedure for abbreviating a clinical rating scale, the Quality of Life Scale, from 21 to 7 items (Bilker et al., 2003). This method was modified to accommodate large item sets and used for reducing a matrix reasoning test from 60 items to 9 (Bilker et al., 2012). Here we describe the application of this modified method to reducing the number of items in the PVRT.

Method

Data

The study sample comprised 319 participants. Table 1 presents a summary of the sample demographics. Age ranged from 15 to 77, with a mean of 34.1 ($SD = 13.0$) years. There were 172 male participants (54%), and the mean number of years of education was 14.4 ($SD = 2.6$). Of the participants, 109 were healthy volunteers, 100 were patients with schizophrenia, 55 were relatives of patients, and 55 were classified as other.

We split the sample into two subsets: one to fit and the other to validate the models. The “model construction set” contained 160 participants and the “validation set” contained 159 participants. No demographic measures shown in Table 2 were significantly different between the model construction and validation sets.

The method of Bilker et al. (2003) considered all possible subsets of test items, up to about one half of the total test items. For each combination, the predicted total score of the omitted items was estimated using linear regression. The sum of predicted scores for the omitted items and the observed scores of the items included in the model was then used as the predicted total for the particular model. The Pearson correlation and the intraclass correlation between the predicted and observed total scores were used as metrics for assessing the predictive ability of each subset of test items.

Their procedure required modification to allow application to the PVRT. First, the outcomes of interest are counts, and, as shown in Figure 1, the total number of incorrect items appears to follow a Poisson distribution. Second, the PVRT contains 29 items, and as we wish to reduce the number of items to at most 20, there are over 530 million possible subtests, and thus the computational burden of applying a brute force procedure as the above is not practical.

Bilker et al. (2012) sought to rectify this issue by modifying the previous approach. First, Poisson regression replaced linear regression to better fit the count data. Second, an adaptive algorithm was employed that dramatically reduced the number of models considered in the reduction process. This modified scale reduction approach as applied to the PVRT reduction is described below.

The process of selecting a subset of items is as follows: Let A be a particular subset of items and let A^C be all items not in A . For each subset, a Poisson regression is run predicting the total number of incorrect items in A^C from the scores of the items in A . As an example, suppose A contained Items 1, 4, 7, and 13. Thus, A^C contains all items except 1, 4, 7, and 13 and the predictive Poisson model is then

$$\log(R) = \alpha + \beta_1 I_1 + \beta_4 I_4 + \beta_7 I_7 + \beta_{13} I_{13}$$

where R is the number of incorrect items in A^C ; α is the intercept; I_j is equal to 1 if item j is incorrect and 0 otherwise; and β_j is the coefficient for item j . The predicted scores of the items in A^C are then computed via: $\hat{R} = \exp(\hat{\alpha} + \hat{\beta}_1 I_1 + \hat{\beta}_4 I_4 + \hat{\beta}_7 I_7 + \hat{\beta}_{13} I_{13})$. Since $\hat{R}$ represents a count, round $\hat{R}$ to the nearest integer, denoted $\hat{R}^*$. Obtaining the predicted value of the number correct among all items is then a simple calculation. Letting $\hat{Y}$ be the predicted value of the total number correct and V be the observed number of incorrect responses from items in A , we have $\hat{Y} = 29 - (V + \hat{R}^*)$.

As the original and reduced forms of the test share items, the Pearson correlation between the total scores from both forms is spuriously high (Levy, 1967; Tellegan & Briggs, 1967). The spuriousness is due to the variance of the reduced form scores, which are contained in both forms. Thus we utilize the modified correlation coefficient proposed by Levy (1967),

which adds a negative-valued penalty term to the Pearson correlation, taking into account the shared variation (Levy, 1967). For each subset A, the modified correlation coefficient (ρ) and the intraclass coefficient (ICC) between the predicted values of total correct items with the observed number of correct items are computed for both the model construction and validation sets.

To assess the effect of the particular split on the results, we employed the two bootstrap methods presented in Bilker et al. (2012). The first considers the drop in correlation between the split sets while the second explores the particular subset of items and their frequency of inclusion in the final subtest among different splits of the data. We set an a priori rule for determining the optimal model, adding predictors until the percentage increase in correlation fell below 1% for an additional predictor.

Results

Table 3 presents the items chosen, modified correlations, and intraclass correlation coefficients for the top models for each number of predictors for the construction and validation sets. All of the models were fit using all diagnostic groups collapsed. The correlation between predicted and observed total score is above .85 for as few as four items in the construction set and five items in the validation set. The model with eight predictors was chosen as the final model using the predetermined 1% criterion. Figure 2 presents the plot of observed total score versus the predicted total score from the final model.

The items in the final model are 4, 6, 7, 11, 15, 16, 23, and 28. The coefficients for each of these items are found in Table 4. Letting I_k be equal to 1 if the subject's response for item k was incorrect and 0 if correct, predictions from our final model are computed via $\hat{Y} = 29 - (V + R^*)$, where $R^* = \exp(1.228 + 0.241*I_4 + 0.248*I_6 + 0.113*I_7 + 0.227*I_{11} + 0.200*I_{15} + 0.249*I_{16} + 0.277*I_{23} + 0.170*I_{28})$ and $V = I_4 + I_6 + I_7 + I_{11} + I_{15} + I_{16} + I_{23} + I_{28}$.

We compared the performance of the top models for 1 through 15 predictors between patients with schizophrenia and healthy controls. There were 50 patients and 55 controls in the model construction set and 50 patients and 54 controls in the validation set. The modified correlations and ICC of the observed totals with the predicted totals for patients and controls are presented in Tables 5 and 6, respectively. Correlations remained greater than .85 for as few as 3 items in the construction set and 7 items in the validation set. Considering only controls, predicted from the models based on all diagnostic groups collapsed, the correlations remained greater than .85 for as few as 9 items in the construction set and 12 in the validation set. Similar comparisons were made in the subgroups of first episode schizophrenia patients and family members: For first episode patients the correlations remained greater than .85 for as few as 4 items in the constructions set and 7 in the validation set, and for family members the correlations remained greater than .85 for as few as 7 and 10 items in the construction and validation sets, respectively. For the sake of brevity, the corresponding tables for the first episode patients and family members are not shown.

We considered the Pearson correlations among age, years of education, and gender with the full and reduced PVRT. We did not expect to find sex differences or age effects, but did expect significant correlations with education, and examined whether similar correlations were obtained with the full and reduced versions. The correlation between age and the full PVRT was $-.12$, while the correlation between age and the reduced PVRT was $-.09$. The corresponding correlations between education and the full and reduced PVRT were $r(317) = 0.58, p < .01$ and $r(317) = .55, p < .01$, respectively. To investigate whether there were gender differences between the full and reduced PVRT, we examined the mean total scores for each test stratified by gender. The mean total scores for the full PVRT were 18.2 for males and 18.3 for females. For the reduced PVRT, the mean total scores were 18.5 and 18.5, respectively. Thus, no differences in the magnitude of correlations with age, years of education, and sex were apparent between the full and reduced tests.

Since the range of the total score is bounded from below and above, floor and ceiling effects of the abbreviated test can be present. To assess the presence of these effects, we estimated the correlations within tertiles of the full PVRT based on the observed values from the full sample. For the PVRT, the tertiles are 0 to 16, 17 to 23, and 24 to 29. The modified correlations between the full and reduced PVRT for the first, second, and third tertiles were .80, .51, and .37, respectively. These suggest a ceiling effect, particularly in the third quartile, as its correlation is low.

Using a bootstrap approach, we investigated the variability in the difference in correlation in the model construction and validation sets when using our final eight-item reduced test. Using 10,000 bootstraps we calculated the drop in correlation between the model construction and validation sets. The mean drop in correlation was $\hat{R} = .0247$ ($SD = .0131$, 95% CI = .0244, .0249) and was the same drop after the Levy correction to both correlations. Thus, for the reduced PVRT, the drop in correlation was consistent across the random bootstrap data sets.

The above approach assessed our final reduced set of items only. We also assessed the reduced test along with the choice of split used to form the model construction and validation sets. We randomly split the original data 200 times (no new data were generated), and for each split we performed the following: fit the predictive model, choose the best 8 items, and estimate the correlations for the model construction and validation data sets. The mean correlation was $r = .9191$ ($SD = .0037$, 95% CI = .9186, .9196) for the model construction sets, and for the validation sets it was $r = .8760$ ($SD = .0122$, 95% CI = .8743, .8777). The high correlations and tight confidence intervals indicate that the particular split used did not influence the results.

As mentioned above, for each of the 200 random splits, an eight-item reduced test was chosen. Each reduced test may vary from our reduced test shown above. The percentage of times an item was chosen in the top eight is displayed in Table 7. Because the correlation remains high and varies little among the random splits, yet the set of eight varies, we conclude that our eight-item reduced PVRT is one among possibly numerous valid choices achieving a high correlation.

Discussion

The scale reduction approach of Bilker et al. (2012) was successfully applied to the PVRT. The procedure enabled reducing the test substantially from 29 items to 8 items. This amounts to a greater than 70% reduction in administration time, which greatly increases the feasibility of the test's use in large-scale studies. The modified correlation between the abbreviated and long forms was close to unity, $r(158) = .9145$. Validation approaches were used to further assess the quality of the prediction, and the abbreviated form performed exceedingly well. The abbreviated form also performed well when considering first episode schizophrenia patients and healthy adults subgroups. Last, there were no apparent differences in magnitude of sex differences or correlations with age or years of education between the full and reduced tests.

The methodology utilized has limitations including computational burden and the propensity to exhibit ceiling effects. Therefore, the abbreviated scale would not be suitable to detect differences among individuals performing at the high range, which is not the primary use of the PVRT. Furthermore, the reduced set of eight items is not the only valid set of reduced items, and we did not investigate the merit of the particular set relative to the other valid sets of items. However, our analyses demonstrate that the eight items chosen represent a subset that predicts the outcome of the full set of items with very good accuracy. Importantly, the abbreviated test shares many of the limitations of abbreviated tests as enumerated by Smith et al. (2000). These include lack of the ability to reliably assess factors within the large scale and the lower boundary of validity. However, our approach and results have avoided the pitfalls of failing to demonstrate comparable validity for the short and long form and the construct we assessed, verbal reasoning, is narrowly defined by the PVRT and has been unidimensional (e.g., R. C. Gur et al., 2010). Nonetheless, use of the short form should be limited to studies that do not seek to measure factors within the domain of verbal reasoning. Another limitation of abbreviated tests is that differences between the predicted and actual scores can be large even when substantial correlations have been established, and this would limit their ability to categorize impairment as accurately as the full form (Calamia et al., 2011). A final limitation of the approach is that item selection is based entirely on numerical calculations, without any consideration of theoretical relevance or interpretability. Methods that incorporate confirmatory factor analysis and item response theory (Hays, Morales, & Reise, 2000) could achieve similar efficiency while being more theoretically informative.

These limitations notwithstanding, evaluation of verbal reasoning abilities is an important part of neuropsychological assessment, and verbal analogies have traditionally served as stimuli for such tests (Ekstrom et al., 1979). The verbal reasoning test applied here has well-established psychometric properties and taps a well-validated factor (Ekstrom et al., 1979; R. C. Gur et al., 2001; R. C. Gur et al., 2010; R. C. Gur et al., 2012). Furthermore, this test has been used in multiple functional neuroimaging studies, and its processing has been mapped into a brain circuitry that includes frontal and temporoparietal components of the left hemisphere (R. C. Gur et al., 1982; R. C. Gur et al., 1983; R. C. Gur et al., 2000; R. C. Gur & Reivich, 1980). Therefore, performance deficits in the context of abilities in other domains can be linked to dysfunction of a specific brain circuitry. Computerized testing is increasingly used in large-scale genomic and clinical studies, where biomarkers and change

indices are needed (Almasy et al., 2008; Calkins et al., 2010; Green et al., 2004; Insel & Cuthbert, 2009; McClay et al., 2011; Seidman et al., 2010). In such studies verbal reasoning may not be the focus and yet its measurement is desirable if time permits. The abbreviated version of the PVRT developed here would be appropriate in such situations. It is available online for qualified investigators who want to use it in research with institutional review board oversight (<http://www.med.upenn.edu/bbl>).

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by NIH Grants T32MH065218 and P50MH096891.

References

- Almasy L, Gur RC, Haack K, Cole SA, Calkins ME, Peralta JM, Gur RE. A genome screen for quantitative trait loci influencing schizophrenia and neurocognitive phenotypes. *American Journal of Psychiatry*. 2008; 165(9):1185–1192. [PubMed: 18628350]
- Bilker WB, Brensinger C, Kurtz MM, Kohler C, Gur RC, Siegel SJ, Gur RE. Development of an abbreviated schizophrenia quality of life scale using a new method. *Neuropsychopharmacology*. 2003; 28(4):773–777. [PubMed: 12655324]
- Bilker WB, Hansen JA, Brensinger C, Richard J, Gur RE, Gur RC. Development of abbreviated nine-item forms of the Raven's Standard Progressive Matrices Test. *Assessment*. 2012; 19(3):354–369. [PubMed: 22605785]
- Calamia M, Markon K, Denburg NL, Tranel D. Developing a short form of Benton's Judgment of Line Orientation Test: An item response theory approach. *Clinical Neuropsychologist*. 2011; 25(4):670–684. [PubMed: 21469016]
- Calkins ME, Tepper P, Gur RC, Ragland JD, Klei L, Wiener HW, Gur RE. Project among African-Americans to explore risks for schizophrenia (PAARTNERS), evidence for impairment and heritability of neurocognitive functioning in families of schizophrenia patients. *American Journal of Psychiatry*. 2010; 167(4):459–472. [PubMed: 20194479]
- Cheng Y, Morgan DL. Classification accuracy and consistency of computerized adaptive testing. *Behavior Research Methods*. 2013; 45(1):132–142. [PubMed: 22956358]
- Ekstrom RB, French JW, Harman HH. Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*. 1979; 79(2):3–84.
- Green MF, Nuechterlein KH, Gold JM, Barch DM, Cohen J, Essock S, Marder SR. Approaching a consensus cognitive battery for clinical trials in schizophrenia: The NIMH-MATRICES conference to select cognitive domains and test criteria. *Biological Psychiatry*. 2004; 56(5):301–307. [PubMed: 15336511]
- Gur RC, Alsop D, Glahn D, Petty R, Swanson CL, Maldjian JA, Gur RE. An fMRI study of sex differences in regional activation to a verbal and a spatial task. *Brain and Language*. 2000; 74(2):157–170. [PubMed: 10950912]
- Gur RC, Gur RE, Obrist WD, Hungerbuhler JP, Younkin D, Rosen AD, Reivich M. Sex and handedness differences in cerebral blood flow during rest and cognitive activity. *Science*. 1982; 217(4560):659–661. [PubMed: 7089587]
- Gur RC, Gur RE, Rosen AD, Warach S, Alavi A, Greenberg J, Reivich M. A cognitive-motor network demonstrated by positron emission tomography. *Neuropsychologia*. 1983; 21(6):601–606. [PubMed: 6607426]
- Gur RC, Irani F, Seligman S, Calkins ME, Richard J, Gur RE. Challenges and opportunities for genomic developmental neuropsychology: examples from the Penn-Drexel collaborative battery. *Clinical Neuropsychologist*. 2011; 25(6):1029–1041. [PubMed: 21902564]

- Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Gur RE. Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*. 2001; 25(5):766–776. [PubMed: 11682260]
- Gur RC, Reivich M. Cognitive task effects on hemispheric blood flow in humans: Evidence for individual differences in hemispheric activation. *Brain and Language*. 1980; 9(1):78–92. [PubMed: 7357384]
- Gur RC, Richard J, Calkins ME, Chiavacci R, Hansen JA, Bilker WB, Gur RE. Age group and sex differences in performance on a computerized neurocognitive battery in children age 8–21. *Neuropsychology*. 2012; 26(2):251–265. [PubMed: 22251308]
- Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, Gur RE. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods*. 2010; 187(2):254–262. [PubMed: 19945485]
- Gur RC, Turetsky BI, Matsui M, Yan M, Bilker W, Hughett P, Gur RE. Sex differences in brain gray and white matter in healthy young adults. *Journal of Neuroscience*. 1999; 19(10):4065–4072. [PubMed: 10234034]
- Gur RE, Nimgaonkar VL, Almasi L, Calkins ME, Ragland JD, Pogue-Guile MF, Gur RC. Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. *American Journal of Psychiatry*. 2007; 164(5):813–819. [PubMed: 17475741]
- Harvey PD, Keefe RS, Patterson TL, Heaton RK, Bowie CR. Abbreviated neuropsychological assessment in schizophrenia: Prediction of different aspects of outcome. *Journal of Clinical Experimental Neuropsychology*. 2009; 31(4):462–471. [PubMed: 18720182]
- Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Medical Care*. 2000; 38(9 suppl.):II28–II42. [PubMed: 10982088]
- Hill K, Sweeney JA, Hamer RM, Keefe RS, Perkins DO, Gu H, Lieberman JA. Efficiency of the CATIE and BACS neuropsychological batteries in assessing cognitive effects of antipsychotic treatments in schizophrenia. *Journal of the International Neuropsychological Society*. 2008; 14(2):209–221. [PubMed: 18282319]
- Hurford IM, Marder SR, Keefe RS, Reise SP, Bilder RM. A brief cognitive assessment tool for schizophrenia: Construction of a tool for clinicians. *Schizophrenia Bulletin*. 2011; 37(3):538–545. [PubMed: 19776205]
- Insel TR, Cuthbert BN. Endophenotypes: Bridging genomic complexity and disorder heterogeneity. *Biological Psychiatry*. 2009; 66(11):988–989. [PubMed: 19900610]
- Levy P. The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology*. 1967; 23(1):84–86. [PubMed: 6031038]
- Lezak, MD.; Howieson, DB.; Bigler, ED.; Tranel, D. *Neuropsychological assessment*. Oxford University Press; New York, NY: 2012.
- Mansbach WE, MacDougall EE. Development and validation of the short form of the Brief Cognitive Assessment Tool (BCAT-SF). *Aging and Mental Health*. 2012; 16(8):1065–1071. [PubMed: 22852748]
- McClay JL, Adkins DE, Aberg K, Bukszár J, Khachane AN, Keefe RS, van den Oord EJ. Genome-wide pharmacogenomic study of neurocognition as an indicator of antipsychotic treatment response in schizophrenia. *Neuropsychopharmacology*. 2011; 36(3):616–626. [PubMed: 21107309]
- Reichenberg A, Caspi A, Harrington H, Houts R, Keefe RS, Murray RM, Moffitt TE. Static and dynamic cognitive deficits in childhood preceding adult schizophrenia: A 30-year study. *American Journal of Psychiatry*. 2010; 167(2):160–169. [PubMed: 20048021]
- Richland LE, Burchinal MR. Early executive function predicts reasoning development. *Psychological Science*. 2013; 24(1):87–92. [PubMed: 23184588]
- Roalf DR, Ruparel K, Gur RE, Bilker WB, Gerraty R, Elliott MA, Gur RC. Neuroimaging predictors of cognitive performance across a standardized neurocognitive battery. *Neuropsychology*. 2013 Advance online publication.
- Seidman LJ, Giuliano AJ, Meyer EC, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA. Neuropsychology of the prodrome to psychosis in the NAPLS consortium: Relationship to family

- history and conversion to psychosis. *Archives of General Psychiatry*. 2010; 67(6):578–588. [PubMed: 20530007]
- Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. *Psychological Assessment*. 2000; 12(1):102–111. [PubMed: 10752369]
- Tellegan A, Briggs PF. Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of Consulting Psychology*. 1967; 31(5):499–506. [PubMed: 6075979]
- Ventura J, Reise SP, Keefe RS, Baade LE, Gold JM, Green MF, Bilder RM. The Cognitive Assessment Interview (CAI): Development and validation of an empirically derived, brief interview-based measure of cognition. *Schizophrenia Research*. 2010; 121(1):24–31. [PubMed: 20542412]
- Wharton CM, Grafman J, Flitman SS, Hansen EK, Brauner J, Marks A, Honda M. Toward neuroanatomical models of analogy: A positron emission tomography study of analogical mapping. *Cognitive Psychology*. 2000; 40(3):173–197. [PubMed: 10790322]
- Willner AE, Rabiner CJ, Wisoff BG, Hartstein M, Struve FA, Klein DF. Analogical reasoning and postoperative outcome. Predictions for patients scheduled for open heart surgery. *Archives of General Psychiatry*. 1976; 33(2):255–259. [PubMed: 1252101]
- Winegarden BJ, Yates BL, Moses JA, Faustman WO. Development, validity and reliability analysis of short-forms of three Bentonian perceptual tests. *Archives of Clinical Neuropsychology*. 1997; 12(4):430.
- Yokley JL, Prasad KM, Chowdari KV, Talkowski ME, Wood J, Gur RC, Pogue-Geile MF. Genetic associations between neuregulin-1 SNPs and neurocognitive function in multigenerational, multiplex schizophrenia families. *Psychiatric Genetics*. 2012; 22(2):70–81. [PubMed: 22183611]

Figure 1. Histogram of the total number of incorrect items. *Note.* The distribution appears to follow that of a Poisson distribution. A Shapiro–Wilk test of the data results in a test statistic of 0.92 ($p < .001$), signifying a departure from normality.

Figure 2.

Predicted total number of correct items versus the actual total number of correct items along with the 45 degree line, for the model construction data set.

Note. Jittering was used to allow for viewing of overlapping points.

Table 1

Demographics of the Study Sample.

Variable	Group	Overall (<i>N</i> = 319)
Group	Control	109 (34.2%)
	Family	55 (17.2%)
	Other	55 (17.2%)
	Schizophrenic	100 (31.4%)
Sex	Male	172 (53.9%)
	Female	147 (46.1%)
Race	White	222 (69.6%)
	Black	79 (24.8%)
	Asian	4 (1.3%)
	Hispanic	10 (3.2%)
	Pacific Islander	4 (1.3%)
Age (years)	Mean (<i>SD</i> /range)	34.1 (12.9/15, 77)
Age at onset (years)	Mean (<i>SD</i> /range)	24.5 (7.9/15, 50)
Duration of illness (years)	Mean (<i>SD</i> /range)	8.5 (8.1/0, 35)
Education (years)	Mean (<i>SD</i> /range)	14.4 (2.6/8, 22)
Total Penn Verbal Reasoning Test correct	Mean (<i>SD</i> /range)	18.2 (6.8/2, 28)

Table 2

Demographics of the Model Building and Validation Subsets of the Study Sample.

Variable	Group	Model Build (<i>n</i> = 160)	Model Validation (<i>n</i> = 159)	<i>P</i>
Group	Control	55 (34.4%)	54 (34.0%)	.9977
	Family	28 (17.5%)	27 (17.0%)	
	Other	27 (16.9%)	28 (17.6%)	
	Schizophrenic	50 (31.3%)	50 (31.5%)	
Sex	Male	85 (53.1%)	87 (54.7%)	.7755
	Female	75 (46.9%)	72 (45.3%)	
Race	White	116 (72.5%)	106 (66.7%)	.3571
	Black	34 (21.3%)	45 (28.3%)	
	Asian	3 (1.9%)	1 (0.6%)	
	Hispanic	4 (2.5%)	6 (3.8%)	
	Pacific Islander	3 (1.9%)	1 (0.6%)	
Age (years)	Mean (<i>SD</i> /range)	33.5 (12.5/18, 72)	34.7 (13.4/15, 77)	.4370
Age at onset (years)	Mean (<i>SD</i> /range)	24.0 (7.4/15, 45)	25.1 (8.5/15, 50)	.5634
Duration of illness (years)	Mean (<i>SD</i> /range)	8.0 (7.5/0, 29)	8.9 (8.7/0, 35)	.8609
Education (years)	Mean (<i>SD</i> /range)	14.4 (2.5/10, 20)	14.5 (2.7/8, 22)	.9640
Total PVRT Correct	Mean (<i>SD</i> /range)	18.7 (6.6/2, 28)	17.7 (7.0/2, 28)	.2955

Table 3

Selected Items, Modified Correlations, and Intraclass Correlation Coefficients (ICCs) for the Top Models for Each Number of Predictors for the Construction and Validation Sets.

Number of Predictors	Items in Top Model	Model Construction Sample ($n = 160$)			Model Validation Sample ($n = 159$)		
		ρ	ICC	$M \pm SD$ (Min, Max) 18.65 $\pm$ 6.58 (2, 28)	ρ	ICC	$M \pm SD$ (Min, Max) 17.75 $\pm$ 7.00 (2, 28)
1	16	.6551	.6559	18.56 $\pm$ 4.76 (12, 22)	.6036	.5859	18.67 $\pm$ 4.73 (12, 22)
2	23, 29	.7754	.8007	18.91 $\pm$ 5.67 (6, 23)	.6961	.7212	18.23 $\pm$ 6.02 (6, 23)
3	11, 16, 23	.8368	.8665	18.73 $\pm$ 5.91 (8, 24)	.7834	.8020	18.48 $\pm$ 5.93 (8, 24)
4	11, 16, 23, 24	.8650	.8970	18.70 $\pm$ 6.11 (6, 24)	.8023	.8263	18.41 $\pm$ 6.17 (6, 24)
5	12, 16, 22, 23, 29	.8825	.9146	18.86 $\pm$ 6.17 (3, 24)	.8506	.8809	18.08 $\pm$ 6.42 (3, 24)
6	2, 6, 11, 16, 24, 26	.8952	.9285	18.84 $\pm$ 6.32 (3, 25)	.8501	.8801	18.47 $\pm$ 6.99 (3, 25)
7	4, 6, 7, 11, 15, 16, 23	.9067	.9402	18.70 $\pm$ 6.31 (4, 26)	.8865	.9192	18.08 $\pm$ 6.74 (4, 26)
8	4, 6, 7, 11, 15, 16, 23, 28	.9145	.9482	18.71 $\pm$ 6.35 (2, 26)	.8952	.9278	18.11 $\pm$ 6.77 (2, 26)
9	4, 6, 7, 11, 13, 15, 16, 18, 23	.9235	.9571	18.66 $\pm$ 6.35 (3, 26)	.8945	.9281	17.92 $\pm$ 6.78 (3, 26)
10	2, 4, 6, 11, 15, 16, 23, 24, 26, 29	.9285	.9621	18.67 $\pm$ 6.35 (3, 26)	.9015	.9343	18.09 $\pm$ 6.79 (3, 26)
11	1, 4, 6, 7, 11, 13, 15, 16, 18, 23, 27	.9336	.9674	18.61 $\pm$ 6.38 (4, 26)	.9065	.9405	17.91 $\pm$ 7.09 (1, 26)
12	3, 4, 6, 7, 11, 12, 13, 14, 15, 16, 18, 23	.9369	.9708	18.63 $\pm$ 6.44 (2, 26)	.9128	.9470	17.65 $\pm$ 6.95 (2, 26)
13	1, 4, 6, 7, 8, 11, 13, 14, 15, 16, 18, 23, 27	.9397	.9736	18.64 $\pm$ 6.41 (3, 27)	.9169	.9511	17.82 $\pm$ 7.05 (1, 27)
14	4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 22, 23	.9431	.9772	18.59 $\pm$ 6.53 (2, 27)	.9244	.9586	17.75 $\pm$ 7.09 (2, 27)
15	4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 22, 23, 29	.9463	.9804	18.67 $\pm$ 6.50 (2, 27)	.9314	.9656	17.77 $\pm$ 6.95 (2, 27)

The mean, standard deviation, minimum, and maximum of the predicted scores are also presented. The form with eight items is chosen as the final reduced form.

Table 4
Estimated Coefficients for the Poisson Model 1 for the Top Model for Each Number of Predictors.

Number of Predictors	β_1	β_2	β_3	β_5	β_6	B_7	B_8	B_9	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}	β_{18}	β_{22}	β_{23}	β_{24}	β_{26}	β_{27}	β_{28}	B_{29}	α
1																							1.957
2																	0.634					0.535	1.863
3									0.365					0.419			0.429						1.669
4									0.292					0.364			0.363	0.265					1.642
5										0.223				0.268		0.213	0.385					0.278	1.691
6		0.221			0.172				0.3					0.357				0.272	0.198				1.475
7				0.223	0.289	0.138			0.235				0.183	0.279			0.315						1.219
8				0.241	0.248	0.113			0.227				0.2	0.249			0.277				0.17		1.228
9				0.188	0.257	0.105			0.239		0.129		0.161	0.193	0.139		0.246						1.187
10		0.107		0.217	0.144				0.177				0.098	0.264			0.148	0.145	0.134			0.157	1.204
11	0.031			0.162	0.254	0.162			0.211		0.117		0.151	0.197	0.119		0.204			0.1			1.117
12			0.101	0.156	0.206	0.06			0.105	0.09	0.105	0.121	0.138	0.095	0.134		0.238						1.162
13				0.199	0.286	0.169	0.042		0.186		0.144	0.061	0.141	0.194	0.11		0.19			0.061			0.895
14	0.015			0.119	0.193	0.131	0.096	0.13	0.169	0.127	0.18	0.077	0.12	0.133	0.09	0.093	0.302						0.617
15				0.135	0.144	0.09	0.086	0.137	0.115	0.112	0.168	0.051	0.112	0.13	0.081	0.084	0.319					0.111	0.642

Table 5

Selected Items, Modified Correlations, and Intraclass Correlation Coefficients (ICCs) for the top Models for Each Number of Predictors for the Construction and Validation Sets of Schizophrenic Patients Only.

Number of Predictors	Items in Top Model	Model Construction Sample (<i>n</i> = 50)			Model Validation Sample (<i>n</i> = 50)		
		<i>p</i>	ICC	<i>M</i> ± <i>SD</i> (Min, Max)	<i>p</i>	ICC	<i>M</i> ± <i>SD</i> (Min, Max)
1	16	.7129	.6688	16.20 ± 4.99 (12, 22)	.415	.3846	16.00 ± 4.95 (12, 22)
2	23, 29	.7958	.8145	16.06 ± 6.97 (6, 23)	.6336	.6394	15.08 ± 6.39 (6, 23)
3	11, 16, 23	.8835	.9107	15.20 ± 6.82 (8, 24)	.6719	.6758	14.78 ± 5.78 (8, 24)
4	11, 16, 23, 24	.8926	.9265	15.02 ± 7.41 (6, 24)	.7456	.7536	14.78 ± 6.27 (6, 24)
5	12, 16, 22, 23, 29	.884	.919	14.86 ± 7.48 (3, 24)	.8068	.8379	13.78 ± 6.62 (3, 24)
6	2, 6, 11, 16, 24, 26	.9136	.9473	15.08 ± 7.73 (3, 25)	.8213	.8413	14.24 ± 7.32 (3, 25)
7	4, 6, 7, 11, 15, 16, 23	.9074	.9409	14.92 ± 7.29 (4, 26)	.8667	.9002	13.36 ± 6.35 (4, 26)
8	4, 6, 7, 11, 15, 16, 23, 28	.9231	.9554	14.96 ± 7.19 (2, 26)	.8812	.9164	13.18 ± 6.57 (2, 26)
9	4, 6, 7, 11, 13, 15, 16, 18, 23	.929	.9619	15.00 ± 7.32 (3, 26)	.8712	.9066	13.20 ± 6.61 (3, 26)
10	2, 4, 6, 11, 15, 16, 23, 24, 26, 29	.9298	.9644	14.70 ± 7.52 (3, 26)	.9029	.9361	13.44 ± 6.59 (3, 24)
11	1, 4, 6, 7, 11, 13, 15, 16, 18, 23, 27	.9349	.9694	14.78 ± 7.55 (4, 26)	.89	.925	13.04 ± 6.87 (1, 26)
12	3, 4, 6, 7, 11, 12, 13, 14, 15, 16, 18, 23	.9422	.9762	14.80 ± 7.45 (2, 26)	.8986	.9316	12.62 ± 6.92 (2, 26)
13	1, 4, 6, 7, 8, 11, 13, 14, 15, 16, 18, 23, 27	.9427	.9767	14.92 ± 7.72 (3, 27)	.9017	.9362	12.88 ± 6.89 (1, 25)
14	4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 22, 23	.943	.9773	14.80 ± 7.60 (2, 26)	.9059	.9383	12.56 ± 6.91 (2, 26)
15	4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 22, 23, 29	.9455	.9796	14.86 ± 7.64 (2, 26)	.9204	.9541	12.70 ± 6.73 (4, 26)

Note. The mean, standard deviation, minimum, and maximum of the predicted scores are also presented.

Table 6

Selected Items, Modified Correlations, and Intraclass Correlation Coefficients (ICCs) for the Top Models for Each Number of Predictors for the Construction and Validation Sets of Healthy Controls Only.

Number of Predictors	Items in Top Model	Model Construction Sample (<i>n</i> = 55)			Validation Sample (<i>n</i> = 54)		
		<i>p</i>	ICC	<i>M</i> ± <i>SD</i> (Min, Max)	<i>p</i>	ICC	<i>M</i> ± <i>SD</i> (Min, Max)
1	16	.497	.4679	20.18 ± 3.89 (12, 22)	.4499	.4521	20.70 ± 3.39 (12, 22)
2	23, 29	.625	.6223	20.93 ± 3.67 (6, 23)	.8515	.8849	21.39 ± 4.54 (7, 26)
3	11, 16, 23	.6925	.7163	21.51 ± 3.89 (8, 24)	.8611	.8961	21.61 ± 4.45 (7, 27)
4	11, 16, 23, 24	.7193	.7376	21.64 ± 3.38 (12, 24)	.8881	.9231	21.63 ± 4.52 (7, 27)
5	12, 16, 22, 23, 29	.8128	.8297	21.65 ± 3.42 (12, 24)	.7842	.794	21.54 ± 3.56 (9, 24)
6	2, 6, 11, 16, 24, 26	.8041	.8268	21.85 ± 3.40 (9, 25)	.7383	.7739	21.78 ± 4.32 (8, 25)
7	4, 6, 7, 11, 15, 16, 23	.8434	.8772	22.00 ± 3.95 (10, 26)	.8022	.8361	21.57 ± 4.95 (4, 26)
8	4, 6, 7, 11, 15, 16, 23, 28	.8404	.8753	22.07 ± 3.89 (10, 26)	.8206	.8566	21.65 ± 4.72 (6, 26)
9	4, 6, 7, 11, 13, 15, 16, 18, 23	.8639	.8981	22.05 ± 3.87 (9, 26)	.8122	.847	21.46 ± 4.73 (6, 26)
10	2, 4, 6, 11, 15, 16, 23, 24, 26, 29	.8899	.919	21.84 ± 3.82 (9, 26)	.8077	.8442	21.72 ± 4.63 (7, 26)
11	1, 4, 6, 7, 11, 13, 15, 16, 18, 23, 27	.8859	.9199	22.09 ± 3.83 (8, 26)	.8293	.8654	21.67 ± 4.65 (5, 26)
12	3, 4, 6, 7, 11, 12, 13, 14, 15, 16, 18, 23	.8817	.9153	22.18 ± 3.70 (11, 26)	.8515	.8849	21.39 ± 4.54 (7, 26)
13	1, 4, 6, 7, 8, 11, 13, 14, 15, 16, 18, 23, 27	.8975	.9316	22.11 ± 3.83 (7, 27)	.8611	.8961	21.61 ± 4.45 (7, 27)
14	4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 22, 23	.9205	.9538	22.05 ± 3.91 (9, 27)	.8881	.9231	21.63 ± 4.52 (7, 27)
15	4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 22, 23, 29	.927	.9603	22.11 ± 3.80 (10, 27)	.8979	.9324	21.59 ± 4.49 (8, 27)

Note. The mean, standard deviation, minimum, and maximum of the predicted scores are also presented.

Table 7

The percentage of Times Each Item Was Selected in the Top 8 Across 200 Random Splits of the Data.

In Final Model	Item	% in A
	1	33.5
	2	6.5
	3	51.0
*	4	62.5
*	6	36.0
*	7	39.0
	8	13.0
	9	35.0
	10	4.5
*	11	42.0
	12	47.5
	13	29.0
	14	32.0
*	15	51.0
*	16	47.0
	17	13.0
	18	12.0
	19	9.0
	20	34.0
	21	6.0
	22	39.0
*	23	85.0
	24	8.5
	25	1.0
	26	35.0
	27	5.0
*	28	4.5
	29	18.5

Note. The composition of the eight-item form varies across different splits.