Published in final edited form as:

J Clin Exp Neuropsychol. 2015 June ; 37(5): 530–537. doi:10.1080/13803395.2015.1037253.

Practice effects distort translational validity estimates for a Neurocognitive Battery

Ibtihal Ibrahim², Salwa Tobar², Mai Elassy², Hader Mansour^{1,2}, Kehui Chen^{1,4}, Joel Wood¹, Ruben C. Gur⁵, Raquel E. Gur⁵, Wafaa El Bahaei², and Vishwajit Nimgaonkar^{1,3,*}

¹Department of Psychiatry, University of Pittsburgh School of Medicine, Western Psychiatric Institute and Clinic, Pittsburgh, Pennsylvania, USA

²Department of Psychiatry, Mansoura University School of Medicine, Mansoura, Egypt.

³Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

⁴Department of statistics, Pittsburgh, Pennsylvania, USA.

⁵Neuropsychiatry Section, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.

Abstract

Introduction—With the globalization of biomedical research and the advent of “precision medicine”, there is increased need for translation of neuropsychological tests, such as computerized batteries that can be incorporated in large-scale genomic studies. Estimates of translational validity are obtained by administering the test in the original and the translated versions to **bilingual individuals**. We investigated the translation of a neuropsychological battery from English to Arabic and how practice effects influence translational validity estimates.

Methods—The Penn computerized neurocognitive battery (Penn CNB) includes tests that were validated with functional neuroimaging and provides measures of accuracy and speed of performance in several cognitive domains. To develop an Arabic version of the CNB, the English version was translated into Arabic, then back translated and revised. The Arabic and the original English versions were administered in a randomized crossover design to bilingual participants (N=22).

Results—Performance varied by cognitive domain, but generally improved at the second session regardless of the language of the initial test. When performance on the English and Arabic version was compared, significant positive correlations were detected for accuracy in 8/13 cognitive domains and for speed in 4/13 domains ($r=0.02$ to 0.97). When the practice estimates **using linear models** were incorporated, the translational validity estimates improved substantially (accuracy, $r=0.50$ - 0.96 , speed, $r=0.63$ - 0.92 , all correlations, $p=0.05$ or better).

Conclusion—While crossover designs control for order effects on average performance, practice effects, regardless of language, still need to be removed to obtain estimates of translational

* **Corresponding Author:** Vishwajit L. Nimgaonkar, MD, PhD WPIC, Room. 441 3811 O'Hara Street Pittsburgh, PA 15213 Phone: (412) 246-6353 Fax: (412) 246-6350 nimga+@pitt.edu.

validity. When practice effect is controlled for, the Arabic and English versions of the Penn-CNB are well correlated and the Arabic version is suitable for use in research.

Keywords

Arabs; Cognition; Reaction time; Psychometrics; Linear Models; Translations

INTRODUCTION

Computer based testing for neurocognitive functions has been increasing because it enables flexibility in test choice, provides accurate estimates of response time and time to completion, and eases scoring, data storage and access (**Cernich, Brennana, Barker, & Bleiberg, 2007**). To enable accurate cognitive testing for individuals with severe psychiatric disorders, Gur et al described the Penn neurocognitive battery (Penn CNB) that assesses accuracy as well as speed for a range of cognitive domains (**R. C. Gur et al., 2001b; R. C. Gur et al., 2010**). The Penn CNB reliably identified sex differences & age gradients among research participants (R. C. Gur et al., 2014; R. C. Gur et al., 2001a; R. C. Gur et al., 2001b; R. C. Gur et al., 2012; R. C. Gur et al., 2010; Roalf et al., 2014). It has acceptable construct validity (**R. C. Gur et al., 2010**) and psychometric properties (**Moore, Reise, Gur, Hakonarson, & Gur, 2014**) and the well-documented cognitive impairments in schizophrenia are also detectable (**R. C. Gur et al., 2001a; R. E. Gur et al., 2006**). The Penn CNB has been successfully implemented in large-scale genetic studies that have enabled estimates of heritability as well as gene mapping studies of cognitive domains (Aliyu et al., 2006; Greenwood et al., 2007; Gur RE et al., 2007a; Gur RE et al., 2007b). It is publicly available and can be administered through the internet (<https://pennncnp.med.upenn.edu>). Thus, large-scale normative and disease-specific data on thousands of individuals are available. It has been translated into several languages (Deshpande et al., 1998; Kryspin-Exner, Gur, Hoheisel, Klein, & Six, 2003), but not to Arabic. Its translation to Arabic will make it accessible for studies in 57 Arabic speaking countries (<http://www.numberof.net/number-of-arabic-speakers/>; retrieved 1/31/2015).

The most robust design for testing comparability between an original and a translated version is to administer both to bilingual individuals, balancing for order of languages. Notably, however, balancing the order of languages will control for language order differences, but not for practice effects that will influence either language. Practice effects on measures of cognitive ability are prevalent (Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007), and potentially moderated by age (Mitrushina & Satz, 1991), neurological status (Shatz, 1981), differences among tests (Temkin, Heaton, Grant, & Dikmen, 1999), the interval between testing (Theisen, Rapport, Axelrod, & Brines, 1998) and the number of test administration (Wilson, Watson, Baddley, Emslie, & Evans, 2000). Failure to take practice effects into consideration may compromise the validity of an assessment (Calamia, Markon, & Tranel, 2012).

We developed and implemented an Arabic version of the CNB, and report on its translational validity while examining distortions due to practice effects (**Inzelberg et al., 2007**).

METHODS

Study design

The English version of the Penn CNB was translated into Arabic by bi-lingual faculty level psychiatrists at Mansoura University Hospital (MUH), Mansoura, Egypt. Subsequently, the Arabic version was back translated by a different team of psychiatrists (**independent back translation**). The original and the back-translated versions were compared for discrepancies that were corrected. Back-translation procedures have become a widely accepted standard in cross-cultural research to rule out linguistic biases (**Brislin, 1986**). Next, the revised Arabic and the original English versions of the Penn CNB were administered to Arab individuals - **living in the United States of America**- proficient in English and Arabic. **The** drawback of this single group design is that a testing effect may be present (Sireci & Swaminathan, 1996), but it provides the most robust control of the extraneous variables inherent to between-participant designs (Sanchez J. I., Spector, & Cooper, 2006). Counterbalancing and analyzing order of language administration also helps establish the absence of significant language effects (Sanchez J. I. et al., 2006). Therefore, half of the participants completed the Arabic versions of the CNB first, followed by the English version, while the rest completed the English version first. The study was approved by the Institutional Review Boards (IRB) at Mansoura University and the University of Pittsburgh. Verbal consent was obtained from all participants, per Mansoura University and the University of Pittsburgh IRB regulations.

Session of the Penn CNB

The battery was administered in a quiet space convenient for each participant. The battery was re-administered at the same place and at approximately the same time of day as the initial test. The following tests were included.

1. Motor Praxis Test (MPRAXIS).
2. Penn Continuous Performance Test – Number (PCPT-n)
3. Penn Face Memory Test (CPF).
4. Penn Face Memory Test Delayed Memory (CPFd).
5. Penn Conditional Exclusion Task (PCET).
6. Short Computerized Finger-Tapping Task (sCTAP).
7. Short Visual Object Learning Test (sVOLT).
8. Short Visual Object Learning Test Delayed Memory (sVOLTd).
9. Measured Emotion Differentiation Test (MEDF).
10. Penn Matrix Reasoning Test (PMAT).
11. Short Penn Line Orientation Test (sPLOT).
12. Age Differentiation Test (ADT).
13. Penn Emotion Recognition Task (ER40).
14. Short Fractal N-Back (SFNB2).

MPRAXIS was used primarily to familiarize participants with the computer, so it was not included for data analysis.

Statistical analysis

The Statistical Package for Social Sciences (SPSS, version 19.0 for Windows) and R version 2.15. 1 was used for all analyses.

RESULTS

Demographic details

There were 22 participants (16 women and 6 men). All the participants were at least high school graduates (years of education, mean $\pm$ standard deviation, 17.59 ± 2.08 years, range 12-20). The mean age of participants was 30.18 ± 9.14 years (range 18-48). The mean interval between the first and second session was 14.91 ± 1.5 days (range 13 to 19 days).

Completion rates

Each participant completed the battery in less than 1.5 hours without breaks, except for one participant.

Mean difference

Paired t-tests showed no statistically significant differences between the English and Arabic versions for any of the cognitive tests either with regard to accuracy or completion times, regardless of the order of administration (Supplementary Table 1s, 2s, 3s).

Correlations between English and Arabic versions of the Penn CNB

(i) Translational validity unadjusted for practice effects

Accuracy: A significant positive correlation was detected for 8 out of the 13 comparisons between the English and Arabic versions ($p = 0.05$, Table 1). Overall, the range of correlation values was extensive, **from nil to near unity** ($r = 0.02$ to 0.97).

Speed: Only 4 statistically significant correlations were observed when speed variables were compared for each cognitive domain ($p = 0.05$ or better, Table 1). The range of correlations was as broad as that observed for accuracy ($r = -0.03$ to 0.8). There is **no accuracy** component for the Short Computerized Finger-Tapping Task, as participants are asked tap as many times as possible in 60 seconds **and thus the score indicates only speed**.

(ii) Translation validity adjusted for practice effects: As expected from well-established practice effects, for most of the cognitive tests individuals performed better at the second session compared with the first session. As expected from regression to the mean effects (Stigler, 1997), the change in accuracy estimates between the second and the first session ($Ac2 - Ac1$) appeared to be a linear function of performance at the first session ($Ac1$, Figure 1, left panel). Thus, individuals who performed better at the first session showed smaller improvement compared with individuals who performed worse. Speed typically improved at the second session (i.e., $Sp2$, the second observation, was numerically smaller than $Sp1$, the

first observation.) For clarity, the difference in speed estimates between the first and the second session (Sp1-Sp2) were plotted against the speed estimates for the first session (Figure 1, right panel). The trends for the speed measures also suggested the presence of practice effects. Notably, the relationship of Ac2 – Ac1 versus Ac1 for the Penn Continuous Performance Test was the converse of the trends for the other cognitive tests (pCPT-n, Figure 2, left panel). When an outlier was removed from the analyses (Figure 2, bottom left of left panel), the slope Ac2-Ac1 appeared to be consistent with the other cognitive tests (Figure 2, right panel).

Therefore, the practice effect was modeled as follows. For subject i , we fit the linear model on the other $n-1$ subjects, and then predicted the learning effect for subject i . Then we computed the correlations between the Arabic and English versions, after removing the practice effect predicted from the linear models. This method is similar to the work of Crawford and Garthwaite (**Crawford & Garthwaite, 2006**). Table 2 shows correlations of the Arabic and English versions after adjusting for the predicted learning effects. The correlations improved substantially for several variables and the range of values appeared smaller compared to the unadjusted correlations. All of the correlations were now statistically significant ($p = 0.05$ or better) and in an acceptable range for reliability.

DISCUSSION

The dawn of the age of genomics and “precision medicine” has increased the demands for rapid large-scale assessment of continuous variables that can serve as “endophenotypic markers” of vulnerability for neurocognitive disorders (**Gottesman & Gould, 2003**). Large-scale genomics studies are increasingly conducted across sites and often across international sites, requiring translations of instructions and stimulus materials to multiple languages. Arabic is spoken by over 400 million people on all major continents (**World-Arabic-Language-Day., 2012**), and a translation of the Penn CNB, which is being used in multiple genomic studies worldwide, would permit inclusion of Arabic speakers in such studies. Our finding that the Arabic translation produces data that are reliable against the original English version is encouraging and supports the adoption of the Penn CNB for neurocognitive studies in Arabic speaking populations.

Another lesson from this study would be generally applicable to efforts at validating a translation or producing two forms of a battery, even when in the same language. When administering two versions repeatedly to the same individuals, controlling for version order can eliminate version effect, but correlations between the versions may still be attenuated by practice effects. Indeed, the identical means obtained for the two versions in our study indicated lack of version effect. However, considering the correlations between the raw scores of the tests without adjusting for practice effects falsely underestimated the translational validity of the tests. Upon adjustment for the practice estimates, the correlations between the English and the Arabic administrations improved substantially. Our results are consistent with another study in which test-specific practice effects on performance in cognitive domains were estimated through repetitive testing using the English version of a cognitive battery over one year. Improvement from baseline to second testing approximately 2-3 weeks later accounted for the largest proportion of change in all cognitive domains, most

prominently for executive functions and learning/memory (Bartels, Wegrzyn, Wiedl, Ackermann, & Ehrenreich, 2010). After three months and upon reduced testing frequency over one year, the acquired cognitive level stabilized. A reliable prediction of expected practice effects under healthy good-performance conditions is thus deemed essential for accurate effect size estimation and for selection of a suitable test set (Bartels et al., 2010).

To avoid over fitting, we adjusted the correlations for the practice effect by utilizing the predicted value, not the fitted value, but other relevant variables could impact the magnitude of the practice effect estimates. Evaluation of a larger sample is needed in order to obtain more accurate estimates of the variability. The practice effect may also be overestimated in our sample due to the short interval between the first and second session, the relatively young sample (range: 18-48 years) and relatively high educational achievements of the participants (average: 17 years). Practice effects were not detected when the Penn CNB was administered after a 1-2 year interval (R. C. Gur et al., 2001 a); (Roalf et al., 2013), but practice-induced score gains are well documented for other cognitive tests when they were re-administered following a shorter time lag (Hausknecht et al., 2007). For some tests practice effects were found between the first and second test administration only (Ivnik et al., 2000). In a meta-analysis of the practice effects in neuropsychological assessment, increased age was associated with decrease in score gain at retesting, the longer the retest interval the lower the score gain in retesting and a clinical group showed less score gain than healthy groups (Calamia et al., 2012). Since the order of language version was randomized among subjects, we do not expect the language order to be associated with the initial performance. A limitation of our study is the small sample size, necessitated by the limited availability of bilingual participants at the research sites. Sanchez and his colleagues noted that studies of bilinguals are scarce (Sanchez J. I. et al., 2006), probably because large samples of bilinguals are not readily available.

Cross-cultural studies including translations of instruments became inevitable and require the use of appropriate statistical analysis techniques to help properly measure or control for score gains due to practice effects (Calamia et al., 2012). There are multiple factors that could influence the validity of scores derived from an instrument in different cultural settings and languages (Geisinger, 1994; Hambleton, 2001; Van de Vijver & Hambleton, 1996). Our initial study reports on an Arabic version of the Penn CNB that can be used for a range of investigations requiring assessment of cognitive function. When testing the validity of other translated versions of cognitive tests, we recommend adjustment for practice effects, especially when intervals between testing sessions are short.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This study was funded in part by support from NIMH, the Fogarty International Center (MH 63480, MH093246, D43TW008302, D43 TW009114) and the Stanley Medical Research Institute (07R-1712).

REFERENCES

- Aliyu MH, Calkins ME, Swanson CL Jr, Lyons PD, Savage RM, May R, Group PS. Project among African-Americans to explore risks for schizophrenia (PAARTNERS): recruitment and assessment methods. *Schizophr Res*. 2006; 87(1-3):32–44. doi: 10.1016/j.schres.2006.06.027. [PubMed: 16887335]
- Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci*. 2010; 11:118. doi: 10.1186/1471-2202-11-118. [PubMed: 20846444]
- Brislin, RW. The wording and translation of research instruments.. In: Lonner, WJ.; Berry, JW., editors. *Field Methods in Cross-Cultural Research*. Sage; Newbury Park, CA: 1986.
- Calamia M, Markon K, Tranel D. Scoring Higher the Second Time Around: Meta-Analyses of Practice Effects in Neuropsychological Assessment. *The Clinical Neuropsychologist*. 2012; 26(4):543–570. doi: 10.1080/13854046.2012.680913. [PubMed: 22540222]
- Cernich AN, Brennana DM, Barker LM, Bleiberg J. Sources of error in computerized neuropsychological assessment. *Arch Clin Neuropsychol*. 2007; 22(Suppl 1):S39–48. doi: 10.1016/j.acn.2006.10.004. [PubMed: 17097851]
- Crawford J, Garthwaite P. Comparing Patients' Predicted Test Scores From a Regression Equation With Their Obtained Scores: A Significance Test and Point Estimate of Abnormality With Accompanying Confidence Limits. *Neuropsychology*. 2006; 20(3):259–271. doi: 10.1037/0894-4105.20.3.259. [PubMed: 16719619]
- Deshpande SN, Mathur MN, Das SK, Bhatia T, Sharma S, Nimgaonkar VL. A Hindi version of the Diagnostic Interview for Genetic Studies. *Schizophr Bull*. 1998; 24(3):489–493. [PubMed: 9718640]
- Geisinger KF. Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*. 1994; 6(4):304–312.
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 2003; 160(4):636–645. [PubMed: 12668349]
- Greenwood TA, Braff DL, Light GA, Cadenhead KS, Calkins ME, Dobie DJ, Schork NJ. Initial heritability analyses of endophenotypic measures for schizophrenia: the consortium on the genetics of schizophrenia. *Arch Gen Psychiatry*. 2007; 64(11):1242–1250. [PubMed: 17984393]
- Gur RC, Calkins ME, Satterthwaite TD, Ruparel K, Bilker WB, Moore TM, Gur RE. Neurocognitive growth charting in psychosis spectrum youths. *JAMA Psychiatry*. 2014; 71(4):366–374. doi: 10.1001/jamapsychiatry.2013.4190. [PubMed: 24499990]
- Gur RC, Ragland JD, Moberg PJ, Bilker WB, Kohler C, Siegel SJ, Gur RE. Computerized neurocognitive scanning: II. The profile of schizophrenia. *Neuropsychopharmacology*. 2001a; 25(5):777–788. doi: 10.1016/S0893-133X(01)00279-2. [PubMed: 11682261]
- 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*. 2001b; 25(5):766–776. 25 (5), 766-776. doi: 10.1016/S0893-133X(01)00278-0. [PubMed: 11682260]
- Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C. Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*. 2001 a; 25(5):766–776. 766-776. [PubMed: 11682260]
- 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. doi: 10.1037/a0026712. [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. *J Neurosci Methods*. 2010; 187(2):254–262. doi: 10.1016/j.jneumeth.2009.11.017. [PubMed: 19945485]
- Gur RE, Calkins ME, Gur RC, Horan WP, Nuechterlein KH, Seidman LJ, WS S. The Consortium on the Genetics of Schizophrenia: neurocognitive endophenotypes. *Schizophr Bull*. 2007a; 33:49–68. doi: 10.1093/schbul/sbl055. [PubMed: 17101692]

- Gur RE, Nimgaonkar VL, Almasy L, Calkins ME, Ragland JD, MF P-G, RC G. Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. *Am J Psychiatry*. 2007b; 164(5):813–819. [PubMed: 17475741]
- Gur RE, Kohler CG, Ragland JD, Siegel SJ, Lesko K, Bilker WB, Gur RC. Flat affect in schizophrenia: relation to emotion processing and neurocognitive measures. *Schizophr Bull*. 2006; 32(2):279–287. doi: 10.1093/schbul/sbj041. [PubMed: 16452608]
- Hambleton RK. The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*. 2001; 17(3):164–172.
- Hausknecht JP, Halpert JA, Di Paolo NT, Moriarty Gerrard MO. Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. *J Appl Psychol*. 2007; 92(2): 373–385. doi: 10.1037/0021-9010.92.2.373. [PubMed: 17371085]
- Inzelberg R, Schechtman E, Abufel A, Masarwa M, Mazarib A, Strugatsky R, Friedland RP. Education effects on cognitive function in a healthy aged Arab population. *Int Psychogeriatr*. 2007; 19(3):593–603. 19 (3), 593–603. doi: 10.1017/S1041610206004327. [PubMed: 17052375]
- Ivnik RJ, Smith GE, Petersen RC, Boeve BF, Kokmen E, Tangalos EG. Diagnostic accuracy of four approaches to interpreting neuropsychological test data. *Neuropsychology*. 2000; 14:163–177. doi: 10.1037//0894-4105.14.2.163. [PubMed: 10791857]
- Kryspin-Exner I, Gur R, Hoheisel B, Klein M, Six N. Neurobehavioral Probes: Adaptierung und Erweiterung von Verfahren zur computergestützten neuropsychologischen Diagnostik. *Verhaltenstherapie und Verhaltensmedizin*. 2003; 24:27–51.
- Mitrushina M, Satz P. Effect of repeated administration of a neuropsychological battery in the elderly. *J Clin Psychol*. 1991; 47(6):790–801. doi: 10.1002/1097-4679(199111)47:6<790. [PubMed: 1757583]
- Moore TM, Reise SP, Gur RE, Hakonarson H, Gur RC. Psychometric Properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology*. 2014 doi: 10.1037/neu0000093.
- Roalf DR, Gur RC, Almasy L, Richard J, Gallagher RS, Prasad K, Gur RE. Neurocognitive Performance Stability in a Multiplex Multigenerational Study of Schizophrenia. *Schizophr Bull*. 2013; 39(5):1008–1017. doi: DOI 10.1093/schbul/sbs078. [PubMed: 22927671]
- Roalf DR, Ruparel K, Gur RE, Bilker W, Gerraty R, Elliott MA, Gur RC. Neuroimaging predictors of cognitive performance across a standardized neurocognitive battery. *Neuropsychology*. 2014; 28(2):161–176. doi: 10.1037/neu0000011. [PubMed: 24364396]
- Sanchez JI, Spector P. E. a. Cooper CL. FREQUENTLY IGNORED METHODOLOGICAL ISSUES IN CROSS-CULTURAL STRESS RESEARCH. *Handbook of Multicultural Perspectives on Stress and Coping*. International and Cultural Psychology. 2006:187–201.
- Schatz MW. WAIS practice effects in clinical neuropsychology. *J Clin Neuropsychol*. 1981; 3(2):171–179. doi: 10.1080/01688638108403123. [PubMed: 7276204]
- Sireci, SG.; Swaminathan, H. Evaluating Translation Equivalence: So What's the Big Dif?. Paper presented at the Annual Meeting of the Northeastern Educational Research Association; Ellenville, NY. October 1996; 1996.
- Stigler SM. Regression toward the mean, historically considered. *Statistical Methods in Medical Research*. 1997; 6(2):103–114. doi: 10.1191/096228097676361431. PMID 9261910. [PubMed: 9261910]
- Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*. 1999; 5:357–369. doi: 10.1017/S1355617799544068. [PubMed: 10349298]
- Theisen ME, Rapport LJ, Axelrod BN, Brines DB. Effects of practice in repeated administrations of the Wechsler Memory Scale-Revised in normal adults. *Assessment*. 1998; 5:85–92. doi: 10.1177/107319119800500110. [PubMed: 9458344]
- Van de Vijver F, Hambleton RK. Translating tests: Some practical guidelines. *European Psychologist*. 1996; 1(2):89–99.
- Wilson BA, Watson PC, Baddley AD, Emslie H, Evans JJ. Improvement or simply practice? The effects of twenty repeated assessments on people with and without brain injury. *Journal of the International Neuropsychological Society*. 2000; 6:469–479. doi: 10.1017/S1355617700644053. [PubMed: 10902416]

World-Arabic-Language-Day. W. A. L. D. UNESCO. , editor. Building Knowledge Societies. 2012. (pp. Published in March 2013 by the Sector for External Relations and Public Information of the United Nations Educational, Scientific and Cultural Organization (UNESCO).). <http://www.unesco.org/new/en/unesco/events/prizes-and-celebrations/celebrations/international-days/world-arabic-language-day/>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Change in accuracy and speed between the first and second session of the Penn CNB as a function of performance at the first session
The slopes of change in accuracy (left panel) and speed are shown (right panel). Ac1, Ac2: accuracy of the test in the first and second test sessions, respectively; Sp1, Sp2: speed of the test in the first and second test sessions respectively. For explanation of abbreviations of cognitive tests, please see second column of Table.

Figure 2. Accuracy estimates for the Penn Continuous Performance test

Accuracy estimates for the first and second sessions of the Penn continuous Performance test are shown (Ac1, Ac2, respectively) for the entire sample (left panel). In the right panel they are shown after removal of an outlier in the left lower panel. pcpt: Penn continuous Performance test.

Table 1

Correlations for performance in the Arabic and English versions of the Penn CNB.

Test	Abbreviation	Accuracy		Speed	
		Correlation	p	Correlation	P
Penn Face Memory Test Delayed Memory	CPFd	0.02	0.936	0.03	0.896
Penn Face Memory Test Memory Test	CPF	0.07	0.769	−0.03	0.885
Penn Conditional Exclusion Task	PCET	0.14	0.539	0.41	0.058
Short Visual Object Learning Test	sVOLT	0.33	0.131	0.42	0.053
Penn Emotion Recognition Task	ER40	0.37	0.090	0.32	0.144
Short Visual Object Learning Test Delayed	sVOLTd	0.44	0.041	0.27	0.232
Short Fractal N-Back	SFNB2	0.50	0.017	0.38	0.085
Age Differentiation Test	ADT	0.67	0.001	0.60	0.003
Penn Matrix Reasoning Test	PMAT	0.70	<0.001	0.68	0.001
Measured Emotion Differentiation Test	MEDF	0.79	<0.001	−0.08	0.735
Short Penn Line Orientation Test	sPLOT	0.80	<0.001	0.49	0.021
Short Computerized Finger-Tapping Task	sCTAP	0.82	<0.001	-----	-----
Penn Continuous Performance Test – Number	pCPT-n	0.97	<0.001	0.8	0.000

Table 2

Correlations after adjusting for practice effects.

Test	Abbreviations	Accuracy		Speed	
		Correlation	P	Correlation	P
Penn Face Memory Test Delayed Memory	CPFd	0.61	0.003	0.90	<0.001
Penn Face Memory Test	CPF	0.71	<0.001	0.89	<0.001
Penn Conditional Exclusion Task	PCET	0.58	0.005	0.63	0.002
Short Visual Object Learning Test	sVOLT	0.50	0.017	0.74	<0.001
Penn Emotion Recognition Task	ER40	0.75	<0.001	0.79	<0.001
Short Visual Object Learning Test Delayed Memory	sVOLTd	0.74	<0.001	0.90	<0.001
Short Fractal N-Back	SFNB2	0.84	<0.001	0.90	<0.001
Age Differentiation Test	ADT	0.82	<0.001	0.92	<0.001
Penn Matrix Reasoning Test	PMAT	0.73	<0.001	0.76	<0.001
Measured Emotion Differentiation Test	MEDF	0.88	<0.001	0.84	<0.001
Short Penn Line Orientation Test	sPLOT	0.83	<0.001	0.84	<0.001
Short Computerized Finger-Tapping Task	sCTAP	0.91	<0.001	-----	-----
Penn Continuous Performance Test – Number	pCPT-n	0.96	<0.001	0.84	<0.001