Published in final edited form as:

Clin Neuropsychol. 2019 April ; 33(3): 557–570. doi:10.1080/13854046.2018.1488995.

Validation of Computerized Episodic Memory Measures in a Diverse Clinical Sample Referred for Neuropsychological Assessment

Robyn M. Busch, PhD^{1,2}, Olivia Hogue, MPH⁵, Lisa Ferguson, M.S.^{1,2}, Michael W. Parsons, PhD^{1,4}, Cynthia S. Kubu, PhD^{1,3}, and Darlene P. Floden, PhD^{1,3}

¹Department of Psychiatry & Psychology, Neurological Institute, Cleveland Clinic

²Epilepsy Center, Neurological Institute, Cleveland Clinic

³Center for Neurological Restoration, Neurological Institute, Cleveland Clinic

⁴Brain Tumor and Neuro-Oncology Center, Neurological Institute, Cleveland Clinic

⁵Quantitative Health Sciences, Cleveland Clinic

Abstract

Objective: To examine the convergent and discriminant validity of two brief computerized episodic memory measures in a large, diverse clinical sample of adults undergoing neuropsychological assessment.

Method: Computerized measures of word and face memory were administered to 233 adults (age 30 and over) who also completed comprehensive neuropsychological testing.

Results: Moderate correlations were observed between the computerized memory tests and a wide range of traditional neuropsychological measures of episodic memory (e.g., word-list learning, story recall, face recognition, design recall). Select measures of visuospatial processing speed and language were also related to performance on the computerized tasks. In contrast, the computerized memory tests showed weak correlations with tests in other cognitive domains (i.e., visuospatial skills, attention/working memory, executive function, motor dexterity, academic skills) and self-report screening measures of mood and anxiety. Similar to traditional measures of episodic memory, the computerized memory measures were sensitive to effects of age and gender.

Corresponding Author: Robyn M. Busch PhD, Department of Psychiatry & Psychology, Cleveland Clinic, 9500 Euclid Avenue, P57, Cleveland, OH 44195, Phone: (216) 444-9042, Fax: (216) 444-5625, buschr@ccf.org.

Other Authors:

Olivia Hogue, MPH, Quantitative Health Sciences, 9500 Euclid Ave., JN3, Cleveland, OH 44195, 216-445-9320, hogueo@ccf.org
Lisa Ferguson, MS, Department of Psychiatry & Psychology, 9500 Euclid Ave., P57, Cleveland, OH 44195, 216-444-5111, fergusl3@ccf.org

Michael Parsons, PhD, Psychology Assessment Center, Massachusetts General Hospital, One Bowdoin Square, Suite 701, Boston, MA 02114, 617-726-3647, mwp@parsons.mgh.harvard.edu

Cynthia Kubu, PhD, Department of Psychiatry & Psychology, 9500 Euclid Ave., P57, Cleveland, OH 44195, 216-445-6848, kubuc@ccf.org

Darlene Floden, PhD, Department of Psychiatry & Psychology, 9500 Euclid Ave., P57, Cleveland, OH 44195, 216-444-1298, flodend@ccf.org

Disclosures

The authors have no conflicts of interest to disclose.

Conclusions: Computerized measures of word and face memory showed good convergent and discriminant validity in this diverse clinical sample supporting the construct validity of these measures. This indicates that it may be feasible to measure memory function in clinical settings using brief, well-designed computerized memory measures.

Keywords

Neuropsychological Tests; Episodic Memory; Computerized Assessment; Psychometrics; Validation Studies

Introduction

It is well recognized that computerized cognitive assessment measures have a number of potential advantages over examiner-administered neuropsychological tests, including automated and accurate scoring, precise reaction time measurements, and the ability to assess large numbers of individuals within a brief time frame (Bauer et al., 2012). As such, there has been a marked increase in the development of computerized cognitive measures and batteries for use in clinical practice and research. While several neuropsychological domains (e.g., visuomotor processing speed, visuospatial skills) easily lend themselves to computer adaptation, other functions, such as memory and language, are more difficult to translate from the traditional administration procedures relying on an examiner into a more automated, computerized format.

The Penn Computerized Neuropsychological Battery (PennCNP; <https://penncnp.med.upenn.edu/>) was developed as a research tool for assessing cognition in functional imaging studies. The subtests that comprise the PennCNP were developed within a cognitive neuroscience framework and assess a wide range of cognitive functions. The PennCNP has demonstrated good internal consistency and moderate to high inter-item correlations within cognitive domains, including episodic memory (Gur et al., 2010), and confirmatory factor analysis has provided support for the theory used to construct the PennCNP (Moore, Reise, Gur, Hakonarson, & Gur, 2015). While the PennCNP has been used in research for many years (daSilva, Richard, & Barrett, 2009), the potential clinical utility of this web-based automated measure has not been thoroughly examined. For this study, we administered a subset of the full battery that included tests of verbal (Penn Word Memory Test) and visual (Penn Face Memory Test) episodic memory, as well as a computerized version of the Judgment of Line Orientation test (Benton, Varney, & Hamsher, 1978) to serve as a distractor task prior to delayed recall.

Gur and colleagues examined the reliability and validity of the PennCNP in several studies. An early study examined performance on the PennCNP and a traditional neuropsychological battery (administered in a counter-balanced order) in a group of 92 healthy, normal adults (Gur, Ragland, Moberg, Turner et al., 2001). Reliability was assessed using Cronbach's alpha with moderate ($r=0.60$) to good (0.83) values reported for the Penn Face Memory Test and Penn Word Memory Test, respectively. The PennCNP memory subtests were sensitive to age and gender effects (similar to what has been documented in standard neuropsychological measures) supporting the criterion validity. The PennCNP memory summary score was

significantly correlated with a memory summary score based on traditional tests ($r=0.53$, $p=0.0001$) but not with the other summary scores reflecting executive function ($r=0.18$, $p>0.05$), intellectual ability ($r=0.14$, $p>0.05$), and sensorimotor function ($r=0.07$, $p>0.05$), providing evidence of convergent and discriminant validity. Another psychometric study based on a large sample of healthy adults documented reasonable inter-item correlations for the accuracy measures of the Penn Word Memory Test (immediate $n=392$, $r=0.71$; delayed $n=390$, $r=0.73$) and Penn Face Memory Test (immediate $n=536$, $r=0.70$; delayed $n=531$, $r=0.75$) with generally good evidence of convergent and discriminant validity based on the pattern of correlations among all measures in the PennCNP (Gur et al., 2010). Together, these studies support the reliability and construct validity of the PennCNP memory measures in healthy adults.

The validity of the PennCNP has also been examined in patients with schizophrenia compared to healthy control subjects (Gur, Ragland, Moberg, Bilker et al., 2001). The PennCNP was found to be sensitive to schizophrenia diagnosis, with evidence of differential impairments in executive function and memory in patients with schizophrenia compared to normal control subjects similar to findings based on traditional neuropsychology tests. Likewise, the PennCNP was sensitive to clinical variables relevant to schizophrenia (e.g., symptom severity, disease duration). However, unlike the studies conducted in normal controls, performance on the PennCNP memory measures was significantly correlated with the broad cognitive summary scores derived from standard neuropsychological measures (r range = $0.40-0.53$, $p<0.05$) with no evidence of discriminant validity in patients with schizophrenia.

In summary, the existing data support the reliability and validity of the PennCNP memory measures in healthy adults. Early studies examining the convergent validity of the PennCNP memory subtests in patients with schizophrenia are adequate, though the data regarding discriminant validity in this population is less compelling. The psychometric properties of these computerized measures have not yet been examined in other clinical populations. Hence, the primary goal of the current study was to examine the construct validity of the memory subtests of the PennCNP in a large, diverse clinical sample of adults undergoing neuropsychological assessment as part of clinical care.

We assessed construct validity by examining convergent and discriminant (or divergent) validity of the PennCNP Word and Face Memory Tests. Construct validity is supported when the correlations are higher for those tests assessing the same process or function (convergent validity) versus those correlations between tests assessing different processes or functions (discriminant or divergent validity). We hypothesized that the PennCNP Word and Face Memory Tests would demonstrate good construct validity in a large sample of adult patients referred for neuropsychological assessment.

Method

Participants

This prospective, IRB-approved investigation included patients, aged 30 and over, who were referred for clinical neuropsychological evaluations at a large, tertiary medical center. All

patients included in the sample were English-speaking, had valid clinical neuropsychological testing results (per formal symptom validity testing and/or clinical judgment of the Board Certified staff neuropsychologist who was responsible for the clinical evaluation of the patient), and provided informed consent to complete the selected PennCNP subtests.

Measures

PennCNP Subtests

The following subtests from the PennCNP were administered to all participants: Penn Word Memory Test, Penn Face Memory Test, and Computerized Judgment of Line Orientation. The Penn Word Memory Test (Gur et al., 1993; Gur et al., 1997) is comprised of a series of 20 target words that are individually presented on the computer screen. Participants are then asked to identify the 20 target words that are mixed with 20 distractors (Words Immediate). This latter procedure is repeated following a delay with 20 new distractors (Words Delay). The Penn Face Memory Test (Gur et al., 1993; Gur et al., 1997) is similarly designed. A series of 20 faces is presented on the computer screen, and participants are asked to identify the 20 target faces that are mixed with 20 distractors immediately following the initial presentation (Faces Immediate) as well as following a delay (Faces Delay). Participants receive an immediate and delayed score for both measures comprised of the number of correctly recognized targets and rejected foils. Response times are recorded for both correct and incorrect responses. The Computerized Judgment of Line Orientation (Gur et al., 1982) is a computerized version of the original paper-and-pencil test (Benton et al., 1978) that is administered during the delay period for the memory subtests. In this task, participants are shown two lines at different orientations/angles and are asked to match them to corresponding lines on another array. This measure is scored by summing correct responses. In addition, patients answered a brief questionnaire that included questions regarding their relative comfort with using a computer and their subjective cognitive complaints.

Participants were offered a parking voucher to compensate them for their time in completing the PennCNP for research. As noted previously, the PennCNP was developed for research and has not been validated or normed for clinical use. As such, the Board Certified neuropsychologist staffing the clinical cases was blinded to PennCNP results.

Neuropsychological Measures

Patients in this study completed the neuropsychological measures deemed necessary by the staff neuropsychologist to address the clinical referral question. As such, there was variability in the cognitive measures used to validate the PennCNP memory subtests. Nevertheless, all clinical evaluations included neuropsychological measures that assessed a wide range of cognitive abilities including verbal and nonverbal memory, visuospatial functioning, attention/working memory, processing speed, language, executive functioning, and motor functioning. Self-report screening measures of mood and anxiety were also completed by most patients. The PennCNP subtests and clinical neuropsychological measures were administered in a counterbalanced order whenever possible given clinical demands. Twenty-eight percent of participants completed the PennCNP prior to the clinical

evaluation, and 72% of patients completed the computerized measure following their clinical evaluation.

Analyses

Descriptive statistics and average time to complete testing were calculated for the sample. Demographic characteristics of patients who participated in the study were compared with patients who did not consent and with patients who consented but withdrew before completing the study. Group differences were examined using independent-samples t-tests for normally-distributed continuous variables, Wilcoxon rank-sum tests for continuous variables with non-normal distributions, and chi-squared tests for categorical variables.

The effects of time of administration and demographic variables on PennCNP subtest performance were assessed using independent-samples t-tests, univariate analysis of variance with Tukey's post-hoc test, and Pearson product-moment correlations. Independent-samples t-tests, Wilcoxon rank-sum tests, and correlation analyses were used to assess associations between demographic factors and standard neuropsychological tests. Pearson product-moment correlations were used for linear relationships and Spearman's rank-order correlations were used for monotonic relationships or non-normal distributions.

Correlation analyses were used to assess the strength of the associations between PennCNP subtests and clinical neuropsychological tests to evaluate convergent and discriminant validity. Scatterplots, Cook's D plots, and leverage by studentized residual plots were used to visually screen for bivariate outliers. In order to reduce the likelihood of interpreting spurious findings due to multiple correlations, correlation coefficients measuring associations between PennCNP subtests and clinical neuropsychological tests were interpreted based on effect size, rather than p values, using the convention set forth by Cohen for use in behavioral sciences (Cohen, 1969): small effect = 0.10, medium effect = 0.30, and large effect = 0.50. For all tests comparing demographic variables, a significance level of 0.05 was assumed. Analyses were conducted using SAS Studio v. 3.5, and raw scores were used for both computerized and traditional neuropsychological measures.

Results

Participants.

Two hundred thirty three out of 654 eligible consecutive patients completed the PennCNP tests. See Figure 1 for details of study enrollment and Table 1 for included sample characteristics. Patients who consented were younger ($t_{653} = -2.53$; $p = .012$) and more educated ($Z = 2.43$; $p = .015$) than those who declined participation or who were not approached. There were no significant demographic differences between those who consented and completed research and those who consented but withdrew.

Performance characteristics of PennCNP.

Participants completed the three PennCNP subtests included in this study in an average of 13.8 (SD=4.5) minutes. Descriptive statistics for patient performance on these subtests are

provided in Table 1. There were no significant differences in PennCNP performance based on administration time (before or after neuropsychological testing).

Women achieved higher scores than men on Words Delayed ($t_{232} = 2.09$, $p = .038$), Faces Immediate ($t_{232} = 3.28$, $p = .001$), and Faces Delayed ($t_{232} = 3.02$, $p = .003$). Men achieved higher scores on Computerized Judgment of Line Orientation ($t_{232} = 4.61$, $p < .001$). Patient age was negatively correlated with Words Delayed ($r = -.14$, $p = .031$, 95% CI $[-.27, -.02]$) and Faces Delayed ($r = -.25$, $p < .001$, 95% CI $[-.37, -.13]$), and patient education was positively correlated with Computerized Judgment of Line Orientation ($r = .21$, $p = .001$, 95% CI $[.08, .33]$). There were no other performance differences when considering demographic variables.

There was no association between accuracy and reaction time for any of the PennCNP memory subtests.

Performance characteristics of Traditional Memory Measures.

Women achieved higher scores than men on the learning trials (total raw score) of the Hopkins Verbal Learning Test ($t_{94} = 2.02$, $p = .047$) and the California Verbal Learning Test ($t_{36} = 2.32$, $p = .026$) as well as the delayed recognition trial of the latter measure ($t_{35} = 2.37$, $p = .023$). Men achieved higher scores than women on Visual Reproduction I of the WMS-IV ($t_{55} = -2.31$, $p = .025$) and Judgment of Line Orientation ($t_{230} = -4.96$, $p < .001$). Patient age was negatively correlated with performance on several verbal and nonverbal memory measures, including list learning total and delayed recall scores on the Rey Auditory Verbal Learning Test and Hopkins Verbal Learning Test, Logical Memory II of the WMS-IV, immediate, delayed, and recognition scores of the BVMT, and Visual Reproduction I and II from the WMS-IV ($r/p = -.21 - -.39$, $p < .001 - .035$). Education was positively correlated with Logical Memory I from the WMS-IV ($r = .28$, $p < .001$, 95% CI $[.13, .42]$) and Judgment of Line Orientation ($r = .20$, $p = .002$, 95% CI $[.05, .34]$).

Convergent validity.

Table 2 displays the associations between the computerized measures and the standard examiner-administered memory tests, and highlights those with moderate and large effect sizes. The Penn Word Memory Test was positively correlated with all neuropsychological measures of verbal memory, including various measures of word-list learning (i.e., Rey Auditory Verbal Learning Test, California Verbal Learning Test, and Hopkins Verbal Learning Test) and story recall (i.e., Logical Memory subtests from WMS-III and WMS-IV) (See Figures 2A and 2B). Most of these correlations had moderate effect sizes ($r/p = .30 - .49$). The Penn Word Memory Test also demonstrated moderate correlations ($r = .31 - .43$) with select measures of visual memory (i.e., Brief Visual Memory Test, Faces subtests from WMS-III). The pattern of correlations did not change when the effect of age was accounted for.

The Penn Face Memory Test was strongly associated with the Faces subtests from the WMS-III ($r = 0.46 - 0.54$) and performance on the Brief Visual Memory Test ($r = .41 - .50$), with moderate to large effect sizes (See Figures 2C and 2D). Again, the pattern of correlations did not change when the effect of age was accounted for. The Penn Face

Memory Test was also positively correlated with many neuropsychological measures of verbal memory, with most effect sizes in the moderate range ($r = .30 - .47$).

The Computerized Judgment of Line Orientation was strongly, positively correlated with the standard version of the same measure ($r = .66$).

Discriminant validity.

Table 3 displays the associations between computerized measures and non-memory standard neuropsychological tests. The Penn Word Memory Test and the Penn Face Memory Test showed moderate relationships with select measures of processing speed ($r = .30 - .41$) and verbal fluency ($r = .30 - .37$). Relationships between the Penn Memory Tests and neuropsychological measures assessing attention/working memory, visuospatial skills, executive functioning, motor functioning, and single-word reading were typically weak. Likewise, small effect sizes were noted in correlations between self-reported depression or anxiety symptoms and the computerized cognitive measures ($r = -.16 - .08$).

The Computerized Judgment of Line Orientation was moderately correlated with measures of working memory, nonverbal memory, and single-word reading ($r = .33 - .42$). Relationships with all other neuropsychological measures, including tasks assessing verbal memory, processing speed, language, executive functioning, and manual dexterity, were small ($r = -0.24 - 0.23$).

Finally, the PennCNP Word and Face Memory Tests showed weak relationships with the PennCNP computerized Judgment of Line Orientation Test ($r = .07 - .23$).

Discussion

The Words and Faces subtests of the PennCNP have good convergent validity (based on presence of medium or large effect sizes) with traditional neuropsychological measures of verbal and visual memory in a large sample of adult patients referred for cognitive assessment. There were generally moderate correlations between computerized and traditional memory measures. The strength of correlations reported between the PennCNP memory subtests and standard, examiner-administered neuropsychological tests are on par with those of many traditional neuropsychological measures assessing episodic memory. For example, in the standardization sample for the Wechsler Memory Scale IV (WMS-IV), correlations between the two verbal memory tests, Logical Memory (LM) and Verbal Paired Associates (VPA), are .44 and .42 for immediate and delayed recall, respectively (Wechsler, 2009), Table 4.1). Similarly, different tests of visual memory (Designs and Visual Reproduction) have correlations of .47 and .38 at the immediate and delayed time points. These data indicate that it may be feasible to measure memory function in clinical settings using well-designed computerized memory measures.

The computerized memory tests also appear to have reasonable discriminant validity in a diverse adult neuropsychology referral sample. Specifically, the Words and the Faces subtests showed weak correlations with performance on standard measures of visuospatial matching, attention/working memory function, executive function, naming, word reading, or

upper extremity motor function. There were, however, moderate associations with processing speed and verbal fluency. This suggests that faster processing and/or better strategic retrieval are beneficial for performance on computerized measures. This is also the case for the traditional memory measures, which showed similar correlations with fluency and processing speed tests. Faster processing conceivably results in more opportunity to meaningfully encode stimuli that are presented at a constant rate by the computer. Faster speed of processing may also produce higher scores on fluency measures and allow better use of strategies for retrieving. Finally, small but consistent correlations among apparently different tests have been demonstrated from the earliest days of psychological testing, and may represent a core capability of the individual, referred to as *g* (Spearman, 1904).

The data also provide support for discriminant validity based on generally weak correlations between the PennCNP memory tests and the computerized Judgment of Line Orientation Test. These data indicate that computerized versus standard examiner administration was much less of a factor than differences in task domain.

Computerized memory performance for both PennCNP Faces and Words Tests (delayed recognition) were significantly but weakly correlated with age, similar to standard neuropsychological measures. Gur and colleagues (2001a) also found small age effects on these tests in a sample of normal adults. We view the similarities between absolute values of the correlations between age and performance on the computerized and standard neuropsychological tests positively because it suggests that computerized memory tests do not overwhelmingly disadvantage older adults who may have less familiarity with computers. Indeed, while older adults in our sample tended to report less comfort with computers ($p = -.29$, $p < .001$, 95% CI $[-.41, -.16]$), computerized test performance was not associated with ratings of comfort with computer use ($r = .02$ to $.15$).

Education and self-reported depression and anxiety did not significantly influence computerized memory performance or traditional memory test scores. The absence of education effects was unexpected given that many standard tests show education effects or stratify norms by education (Strauss, Sherman, & Spreen, 2006). However, the relatively high level of education in our sample may have limited our ability to examine the impact of education on performance.

One aspect of test performance that does seem to change with administration route is sex differences in memory performance. Women performed better than men on both the standard and computerized (delay) verbal tasks. This is consistent with a general tendency for women to achieve higher scores on measures of verbal episodic memory (Strauss et al., 2006). Women also outperformed men on the nonverbal computerized tests (immediate and delay), similar to Gur and colleagues' normative sample study (Gur et al., 2001). We did not find sex differences on the traditional nonverbal memory tests in our sample, as is typically the case in published norms (Strauss et al., 2006).

As noted above, an important limitation of our study arises from the relatively high education level in our sample. Recruiting bias likely contributed to this. While we recruited a majority of eligible patients (64%), those who declined to participate were less educated

(median = 13 yrs, interquartile range = 4). This may have reduced our ability to identify significant relationships between CNP performance and education. Those who declined to participate were also slightly older (mean = 61.35 yrs, SD = 12.83) than patients who participated. Our experience in conducting the study was that older patients were concerned that they would be too fatigued at the end of a full neuropsychological evaluation. Finally, the higher education of our sample may increase the likelihood that patients were more comfortable using a computer than may be the case in the general population. Future work on computerized cognitive measures should specifically recruit older adults and those with lower educational attainment to evaluate the effect of these potential sources of bias.

We must also acknowledge the limitations often associated with use of a convenience sample. In this study, we recruited patients who were undergoing neuropsychological testing as part of their clinical care. Clinical referral questions were very diverse (e.g., memory complaints, epilepsy surgery, depression) and patients were evaluated by 1 of 7 different clinical neuropsychologists who often made use of different neuropsychological batteries. While every patient completed a comprehensive neuropsychological evaluation, with at least one measure in each cognitive domain, a number of different neuropsychological measures were used to assess the same cognitive domain/function. This resulted in relatively small sample sizes for some measures, and sampling bias may have distorted validity coefficients to some extent (Ree, Carretta, Earles, & Albert, 1994). Further, the diversity in cognitive measures and varied sample sizes precluded confirmatory factor analysis, which is the gold standard for establishing convergent and discriminant validity. Future studies will be needed in this regard.

There is a clear recognition within neuropsychology that computer-assisted cognitive testing has several benefits over traditional examiner administered measures. Computerized measures can be self-administered, obviating the need for examiner presence throughout test administration. In addition, computerized measures reduce errors in administration and scoring, and data points can be saved digitally for easy entry to electronic medical records or research databases. Shorter test administration is also a significant advantage. The computerized memory measures used in this study were completed in an average of 13.8 minutes, whereas traditional neuropsychological measures of episodic memory typically take 30–60 minutes to complete (Strauss et al., 2006). Moreover, it is possible that using a computerized format for memory testing, in particular, may provide additional value over traditional formats. For example, the reaction time data that can be captured electronically for individual items may be useful in characterizing aspects of memory performance such as confidence judgments (i.e., trials with long reaction times suggesting low confidence in the results of memory search). Computerized memory testing may also permit use of more complex and naturalistic stimuli than traditional test formats allow (i.e., video vignettes, complex scenes). Using technology to collect and evaluate cognitive processing will better reflect the preferences and experience of younger generations who perform many of their daily tasks with the aid of computers and other forms of technology. Computerized testing is likely to be the ‘tip of the iceberg’ as advances in virtual reality will enable even more elaborate and realistic methods to assess real-world cognitive functioning.

While there are clear benefits to computerized testing, there are also important barriers and factors that must be considered in adopting new technologies for use in clinical practice, such as technical issues, privacy/data security, and cultural/experiential factors (Bauer et al., 2012). This is certainly the case for memory assessment, which is more difficult to translate into computerized environments given that most traditional memory testing relies on oral responses. Further, computerized memory tests are limited in their ability to assess free recall, making it difficult to determine whether a memory failure is related to difficulties with encoding, retrieval, or consolidation. Indeed, correlations between the PennCNP and traditional memory measures were smaller for free recall than recognition scores. As such, use of computerized memory measures in a clinical setting may require additional testing in specific populations or settings to fully understand the idiosyncrasies that accompany memory performance in a computerized format. Unfortunately, the clinical diversity of our cohort precluded us from examining test performance separately by diagnostic category. Certainly, clinical use would also require establishment of good test-retest reliability. The lack of test-retest data for the PennCNP memory subtests precluded the correction of correlation coefficients for attenuation (the potentially weakening effect that can arise from measurement error). Extensive normative data would also be required prior to clinical implementation of any new measure. Nonetheless, our results suggest that the PennCNP Word and Face memory assessments show reasonable validity in a diverse patient population referred for neuropsychological assessment. The PennCNP measures were quick to administer and well tolerated by patients with a wide range of cognitive abilities. It will be increasingly important for neuropsychologists to develop and embrace computerized cognitive screening measures with solid reliability and validity as healthcare shifts to a population-based model and the demand for psychometrically sound, easily administered neurocognitive measures increases.

Acknowledgements

Primary support for this research was provided by the Cleveland Clinic Neurological Institute and a number of Centers within the Neurological Institute, including the Epilepsy Center, the Center for Neurological Restoration, the Lou Ruvo Center for Brain Health, the Rose Ella Burkhardt Brain Tumor & Neuro-Oncology Center, and the Center for Behavioral Health. Additional support was provided by the National Institute of Neurological Disorders and Stroke (D.P.F., grant number 1K23NS091344 – 01A1). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

The authors would like to thank the patients who participated in this study. We would also like to thank the staff members and psychometrists in the Section of Neuropsychology at Cleveland Clinic who assisted with recruitment, clinical coding, and test administration. Finally, the authors would like to thank Ruben Gur, PhD and his team at the University of Pennsylvania for permitting us to use the PennCNP subtests for this study.

References

- Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, & Naugle RI (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist*, 26(2), 177–196. [PubMed: 22394228]
- Benton AL, Varney NR, & Hamsher KD (1978). Visuospatial judgment: A clinical test. *Archives of Neurology*, 35, 364–367. [PubMed: 655909]
- Cohen J (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- daSilva F, Richard J, & Barrett F (2009). *PennCNP General Administration Manual 2.10*. Philadelphia, PA: University of Pennsylvania.

- Gur RC, Gur RE, Obrist WD, Hungerbuhler JP, Younkin D, Rosen AD, et al. (1982). Sex and handedness differences in cerebral blood flow during rest and cognitive activity. *Science*, 217, 659–661. [PubMed: 7089587]
- Gur RC, Jaggi JL, Ragland JD, Resnick SM, Shtasel D, Muenz L, et al. (1993). Effects of memory processing on regional brain activation: cerebral blood flow in normal subjects. *International Journal of Neuroscience*, 72, 31–44. [PubMed: 8225798]
- Gur RC, Ragland D, Mozley PD, Smith R, Alvani A, Bilker W, et al. (1997). Lateralized changes in regional cerebral blood flow during performance of verbal and facial recognition tasks: Correlations with performance and “effort”. *Brain and Cognition*, 33, 388–414. [PubMed: 9126402]
- Gur RC, Ragland JD, Moberg PJ, Bilker WB, Kohler C, Siegel SJ, et al. (2001). Computerized neurocognitive scaling. II. The profile of schizophrenia. *Neuropsychopharmacology*, 25(5), 777–788. [PubMed: 11682261]
- Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, et al. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*, 25, 766–776. [PubMed: 11682260]
- Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, et al. (2010). A cognitive neuroscience based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods*, 187(2), 254–262. [PubMed: 19945485]
- Moore TM, Reise SP, Gur RE, Hakonarson H, & Gur RC (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology*, 29(2), 235–246. [PubMed: 25180981]
- Ree MJ, Carretta TR, Earles JA, & Albert W (1994). Sign changes when correcting for range restriction: A note on Pearson’s and Lawley’s selection formulas. *Journal of Applied Psychology*, 79, 298–301.
- Spearman C (1904). The Proof and Measurement of Association Between Two Things. *American Journal of Psychology*, 15, 72–101.
- Strauss E, Sherman EMS, & Spreen O (2006). *A Compendium of Neuropsychological Tests*. New York: Oxford University Press.
- Wechsler D (2009). *WMS-IV Technical and Interpretive Manual*. San Antonio, TX: Pearson Inc.

Figure 1.
Flowchart Summarizing Patient Eligibility and Study Enrollment

Figure 2.
Scatterplots Depicting the Relationships between Performance on the PennCNP Word Memory Test and Traditional Verbal Memory Measures
Imm = Immediate; Del = Delayed

Figure 3.
Scatterplots Depicting the Relationships between Performance on the PennCNP Face Memory Test and Traditional Visual Memory Measures

Table 1.**Characteristics of Study Sample**

Age (years)	Mean (SD): 57.69 (13.15) Range: 30–87
Education (years)	Median (IQR): 14 (4) Range: 8–21
Sex	Male = 112 (48.07%) Female = 121 (51.93%)
Race	Caucasian = 212 (90.99%) African American = 5 (2.15%) Other = 16 (6.87%)
Patient Group	General Adult Clinic = 96 (41.2%) Dementia Clinic = 66 (28.33%) Brain Tumor Clinic = 26 (11.16%) Movement Disorders Clinic = 25 (10.73%) Epilepsy Clinic = 14 (6.01%) Other = 6 (2.58%)
Neuropsychological Characterization	Cognitively Normal = 77 (33.05%) Memory Impaired = 90 (38.63%) Executive Function Impaired = 78 (33.48%) Attention/WM Impaired = 44 (18.88%) Language Impaired = 29 (12.45%) Visuospatial Impaired = 14 (6.01%) Processing Speed Impaired = 49 (21.03%)
PennCNP Memory Performance	Penn Word Memory Test – Immediate Mean (SD): 32.2 (4.0), Range: 19–40 Penn Word Memory Test – Delayed Mean (SD): 30.1 (4.2), Range: 20–40 Penn Face Memory Test – Immediate Mean (SD): 31.0 (4.3), Range: 14–39 Penn Face Memory Test – Delayed Mean (SD): 31.7 (4.6), Range: 18–40 Computerized Judgment of Line Orientation Mean (SD): 20.9 (5.1), Range: 5–30

Table 2.

Correlations Between PennCNP Subtests and Episodic Memory Measures

	n	Penn Word Memory Test		Penn Face Memory Test		JOLO [§] (95% CI)
		Immediate(95% CI)	Delayed(95% CI)	Immediate(95% CI)	Delayed(95% CI)	
Verbal Memory						
RAVLT Total	102	.356 (.174,.514)	.374 (.194,.529)	.266 (.076,.437)	.261 (.071,.433)	.149 (-.046,.333)
RAVLT Delay	102	.393(.215,.545)	.394 (.217,.546)	.257 (.066,.429)	.268 (.078,.439)	.196 (.002...376)
RAVLT Recognition [§]	102	.387 (.209,.540)	.346 (.163,.506)	.229 (.037,.405)	.303 (.116,.469)	.099 (-.097,.287)
CVLT-II Total	42	.446 (.165,.660)	.419 (.132,.641)	.416 (.129,.639)	.398 (.108,.626)	-.032 (-.332,.274)
CVLT-II Delay	42	.456 (.177,.667)	.363 (.067,.600)	.460 (.182,.670)	.408 (.119,.633)	.029 (-.277,.330)
CVLT-II Recognition	41	.447 (.162,.663)	.432 (.144,.652)	.373 (.074,.610)	.423 (.133,.646)	.093 (-.220,.389)
HVLT Total	96	.447 (.271,.594)	.326 (.135,.494)	.416 (.236,.569)	.469 (.297,.611)	.035 (-.166,.233)
HVLT Delay [§]	96	.381 (.196,.540)	.326 (.135,.494)	.341 (.151,.506)	.406 (.224,.560)	.003 (-.197,.203)
HVLT Recognition [§]	90	.292 (.091,.470)	.362 (.168,.529)	.182 (-.026,.374)	.317 (.118,.491)	-.033 (-.238,.175)
WMS-III LM I	65	.374 (.144,.566)	.456 (.239,.629)	.376 (.146,.567)	.326 (.090,.527)	.160 (-.087,.388)
WMS-III LM II	65	.395 (.168,.582)	.409 (.184,.593)	.332 (.096,.532)	.333 (.097,.533)	.105 (-.142,.340)
WMS-IV LM I	159	.245 (.093,.385)	.194 (.040,.339)	.310 (.163,.444)	.175 (.020,.321)	.061 (-.095,.214)
WMS-IV LM II	159	.316 (.169,.449)	.364 (.221,.491)	.354 (.210,.463)	.330 (.184,.461)	-.056 (-.209,.100)
Visual Memory						
BVMT Total	102	.371 (.191,.527)	.240 (.048,.415)	.405 (.229,.555)	.424 (.251,.571)	.002 (-.192,.196)
BVMT Delay	102	.431 (.259,.577)	.305 (.118,.471)	.458 (.290,.599)	.502 (.341,.634)	.046 (-.149,.238)
BVMT Recognition [§]	101	.207 (.013,.386)	.223 (.029,.400)	.194 (-.001,.375)	.284 (.094,.454)	-.133 (-.320,.064)
WMS-III Faces I	38	.336 (.019,.592)	.223 (-.104,.506)	.541 (.268,.733)	.455 (.159,.676)	-.035 (-.350,.287)
WMS-III Faces II	38	.311 (-.009,.573)	.159 (-.169,.455)	.526 (.249,.723)	.514 (.233,.716)	.110 (-.217,.415)
WMS-IV VR I	61	.265 (.015,.484)	.262 (.011,.482)	.177 (-.078,.410)	.190 (-.064,.421)	.372 (.133,.570)
WMS-IV VR II	61	.299 (.052,.512)	.320 (.075,.529)	.157 (-.098,.393)	.253 (.002,.474)	.290 (.042,.504)
WMS-IV VR Recognition [§]	61	.003 (-.249,.254)	.004 (-.248,2.55)	.147 (-.108,-.187)	.220 (-.033,.447)	.344 (.101,.548)
Penn Word Memory Test						

	n	Penn Word Memory Test		Penn Face Memory Test		JOLO [§] (95% CI)
		Immediate(95% CI)	Delayed(95% CI)	Immediate(95% CI)	Delayed(95% CI)	
Immediate	233	1.00	.644 (.562,.713)	.409 (.297,.510)	.412 (.300,.513)	.124 (−.004,.248)
Delayed	233	.644 (.562,.713)	1.00	.374 (.374,.479)	.408 (.295,.509)	.076 (−.053,.202)
Penn Face Memory Test						
Immediate	233	.409 (.297,.510)	.374 (.258,.479)	1.00	.736 (.671,.789)	.107 (−.021,.232)
Delayed	233	.412 (.300,.513)	.408 (.295,.509)	.736 (.671,.789)	1.00	.230 (.105,.348)

Correlations represents Pearson correlations unless otherwise noted

[§]Spearman correlation

Light gray cells = moderate effect sizes (i.e., .30-.49); Dark gray cells = large effect sizes (i.e., .50); CI = Confidence Interval; RAVLT = Rey Auditory Verbal Learning Test; CVLT-II = California Verbal Learning Test – Second Edition; HVLT = Hopkins Verbal Learning Test; WMS-III = Wechsler Memory Test – Third Edition; WMS-IV = Wechsler Memory Test – Fourth Edition; LM = Logical Memory; BVMT = Brief Visual Memory Test; VR = Visual Reproduction

Table 3.

Correlations Between PennCNP Subtests and Other Neuropsychological Measures

	n	Penn Word Memory Test		Penn Face Memory Test		JOL [§] (95% CI)
		Immediate(95% CI)	Delayed(95% CI)	Immediate(95% CI)	Delayed(95% CI)	
Visuospatial						
Judgment of Line Orientation	163	.072 (-.082,.223)	.048 (-.106,.200)	.209 (.058,.351)	.169 (.016,.314)	.661 (.565,.739)
Attention/Working Memory						
WAIS-IV Digit Span	162	.237 (.086,.377)	.259 (.110,.397)	.298 (.151,.432)	.275 (.127,.411)	.297 (.150,.431)
WAIS-IV Arithmetic	71	.137 (-.099,.358)	.192 (-.043,.407)	.157 (-.079,.376)	.235 (.002,.443)	.424 (.212,.598)
Processing Speed						
WAIS-IV Coding	145	.370 (.221,.502)	.305 (.150,.445)	.379 (.231,.510)	.387 (.240,.517)	.198 (.037,.349)
WAIS-IV Symbol Search	72	.260 (.031,.463)	.278 (.050,.478)	.386 (.170,.566)	.364 (.145,.549)	.220 (-.012,.429)
SDMT Oral	44	.349 (.059,.585)	.407 (.126,.627)	.357 (.068,.591)	.300 (.004,.548)	.226 (-.075,.490)
Language						
Boston Naming Test	227	.225 (.098,.345)	.133 (.003,.258)	.209 (.081,.330)	.249 (.123,.367)	.238 (.112,.357)
Phonemic Fluency (F-A-S)	167	.260 (.113,.396)	.288 (.143,.421)	.307 (.163,.438)	.340 (.199,.467)	.224 (.075,.363)
Semantic Fluency (Animals)	162	.329 (.185,.459)	.370 (.229,.495)	.292 (.145,.426)	.305 (.159,.438)	.144 (-.010,.291)
Executive Functioning						
Trail Making Test – Part A [§]	147	-.212 (-.361,-.052)	-.092 (-.250,.070)	-.208 (-.357,-.048)	-.230 (-.377,-.071)	-.145 (-.299,.017)
Trail Making Test – Part B [§]	146	-.160 (-.314,.002)	-.103 (-.261,.060)	-.138 (-.293,.025)	-.200 (-.351,-.039)	-.153 (-.307,.009)
WCST – Perseverative Responses [§]	150	-.132 (-.286,.028)	-.106 (-.261,.055)	-.272 (-.414,-.117)	-.250 (-.394,-.094)	-.243 (-.388,-.087)
Motor Functioning						
GPT – Dominant [§]	77	-.198 (-.404,.027)	-.253 (-.451,-.031)	-.248 (-.447,-.026)	-.248 (-.447,-.026)	.027 (-.198,.249)
GPT – Nondominant [§]	76	.020 (-.206,.244)	-.093 (-.311,.135)	.046 (-.181,.268)	-.046 (-.268,.181)	.067 (-.160,.288)
Academic Skills						
WRAT 4 – Reading	230	.215 (.089,.335)	.151 (.023,.274)	.251 (.126,.368)	.276 (.153,.391)	.325 (.205,.436)
Psychological Status						
Beck Depression Inventory-II [§]	236	-.007 (-.134,.120)	.039 (-.089,.165)	-.044 (-.170,.084)	-.008 (-.135,.119)	-.123 (-.246,.004)

	n	Penn Word Memory Test		Penn Face Memory Test		JOL [§] (95% CI)
		Immediate(95% CI)	Delayed(95% CI)	Immediate(95% CI)	Delayed(95% CI)	
Beck Anxiety Inventory [§]	225	.075 (-.056,.203)	.088 (-.043,.216)	.003 (-.127,.133)	.028 (-.103,.158)	-.162 (-.286, -.032)

Correlations represents Pearson correlations unless otherwise noted

[§]Spearman correlation

Light gray cells = moderate effect sizes (i.e.,.30-.49); Dark gray cells = large effect sizes (i.e., .50); CI = Confidence Interval; WAIS-IV= Wechsler Adult Intelligence Scale – Fourth Edition; SDMT = Symbol Digit Modalities Test; WCST = Wisconsin Card Sorting Test; GPT = Grooved Pegboard Test; WRAT 4 – Wide Range Achievement Test – Fourth Edition