

Sternberg, and self-ordered pointing tasks strongly loaded onto the second factor, suggesting that these measures are strong indicators of working memory. The probabilistic learning task solely loaded onto the third factor, suggesting that it is an independent indicator of reinforcement learning. Finally, the effort expenditure for reward task modestly loaded onto the second but not the first and third factors, suggesting that effort is most strongly related to working memory.

Conclusions: Our aim was to examine the factor structure of 7 RDoC tasks. Results support the RDoC suggestion of independent cognitive control, working memory, and reinforcement learning. However, effort is a factorially complex construct that is not uniquely or even most strongly related to positive valance. Thus, there is reason to believe that the use of at least 6 of these tasks are appropriate measures of constructs such as working memory, reinforcement learning and cognitive control.

Categories:

Assessment/Psychometrics/Methods (Adult)

Keyword 1: cognitive control

Keyword 2: working memory

Keyword 3: effort testing

Correspondence: Emily T. Sturm, Colorado State University, emily.sturm@colostate.edu

16 Validation of the Pillbox Test in a Veteran Sample: A Replication Study.

Frances M. Bozsik¹, Timothy J. Arentsen^{1,2}, Brad L. Roper^{1,2}, Jessica M. Fett¹, Marcy C. Adler^{1,2}, Jennifer L. Jacobson¹, Charlotte Bayer¹
¹Memphis VA Medical Center, Memphis, TN, USA. ²University of Tennessee Health Science Center, Memphis, TN, USA

Objective: Assessment of medication management, an instrumental activity of daily living (IADL), is particularly important among Veterans, who are prescribed an average of 25-40 prescriptions per year (Nguyen et al., 2017). The Pillbox Test (PT) is a brief, performance-based measure that was designed as an ecologically valid measure of executive functioning (EF; Zartman, Hilsabeck, Guarnaccia, & Houtz, 2013), the cognitive domain most predictive of successful medication

schedule management (Suchy, Ziemnik, Niermeyer, & Brothers, 2020). However, a validation study by Logue, Marceaux, Balldin, and Hilsabeck (2015) found that EF predicted performance on the PT more so than processing speed (PS), but not the language, attention, visuospatial, and memory domains combined. Thus, this project sought to increase generalizability of the latter study by replicating and extending their investigation utilizing a larger set of neuropsychological tests.

Participants and Methods: Participants included 176 patients in a mixed clinical sample (5.1% female, 43.2% Black/African American, 55.7% white, $M_{age} = 70.7$ years, $SD_{age} = 9.3$, $M_{edu} = 12.6$ years, $SD_{edu} = 2.6$) who completed a comprehensive neuropsychological evaluation in a VA medical center. All participants completed the PT where they had five minutes to organize five pill bottles using a seven-day pillbox according to standardized instructions on the labels. Participants also completed some combination of 26 neuropsychological tests (i.e., participants did not complete every test as evaluations were tailored to disparate referral questions). Correlations between completed tests and number of pillbox errors were evaluated. These tests were then combined into the following six domains: language, visuospatial, working memory (WM), psychomotor/PS, memory, and EF. Hierarchical multiple regression was completed using these domains to predict pillbox errors.

Results: Spearman's correlation coefficients indicated that 25 tests had a weak to moderate relationship with PT total errors ($r_s = 0.23 - 0.51$); forward digit span was not significantly related ($r_s = 0.13$). A forced-entry multiple regression was run to predict PT total errors from the six domains. The model accounted for 29% of the variance in PT performance, $F(6, 169) = 11.56$, $p < .001$. Of the domains, psychomotor/PS made the greatest contribution, $t(169) = 2.73$, $p = .007$, followed by language, $t(169) = 2.41$, $p = .017$, and WM, $t(169) = 2.15$, $p = .033$. Visuospatial performance and EF did not make significant contributions ($p_s > .05$). Next, two hierarchical multiple regressions were run. Results indicated that EF predicted performance on the PT beyond measures of PS, $\Delta R^2 = .02$, $p = .044$, but not beyond the combination of all cognitive domains, $\Delta R^2 = .00$, $p = .863$.

Conclusions: Results of this study partially replicated the findings of Logue et al. (2015). Namely, EF predicted PT performance beyond PS, but not other cognitive domains. However,

when all predictors were entered into the same model, visuospatial performance did not significantly contribute to the prediction of pillbox errors. These results suggest that providers may benefit from investigating medication management abilities when deficits in PS, WM, and/or language are identified. Further research is needed to better understand which domains best predict PT failure.

Categories:

Assessment/Psychometrics/Methods (Adult)

Keyword 1: activities of daily living

Correspondence: Frances M. Bozsik, Memphis VA Medical Center, frances.bozsik@gmail.com

17 The Chinese Version of Craft Story Recall: A Preliminary Study on the Diagnostic Values of Mild Cognitive Impairment and Dementia.

Gelan Ying¹, Judith Neugroschl², Amy Aloysi², Dongming Cai^{2,3}, Tianxu Xia², Carolyn W Zhu^{2,3}, Xiaoyi Zeng², Jimmy Akrivos², Lingshi Liu², Yiyu Cao², Wei-Qian Wang², Mary Sano^{2,4}, Glenn E Smith¹, Clara Li²

¹University of Florida, Gainesville, FL, USA.

²Icahn School of Medicine at Mount Sinai, New York, NY, USA. ³James J. Peters VA Medical Center, Bronx, NY, USA. ⁴James J. Peters Veterans Affairs Medical Center, Bronx, NY, USA

Objective: Craft story recall test in the National Alzheimer's Coordinating Center Uniformed Data Set 3 (NACC UDS3) neuropsychological battery has been employed to assess verbal memory and assist clinical diagnosis of mild cognitive impairment (MCI) and dementia. While a Chinese version of the test was adapted, no existing literature has examined the diagnostic validity of the test in Chinese Americans. This study aimed to evaluate the predictive validity of both immediate and delayed recall.

Participants and Methods: The Chinese version of Craft Story was administered in to 78 Chinese participants per their language preference of Mandarin or Cantonese. Outcome measures were verbatim and paraphrase recall of the story immediately and after a 20-minute delay. A multiple linear regression was performed to investigate the association of each

outcome measure with age, education, gender, age when moved to the U.S., years in the U.S., and testing language. To assess its diagnostic value, cutoff standard deviation scores of -1.5 and -2.0 from the mean of the clinically cognitive normal participants were generated for MCI and dementia diagnoses, respectively. Due to the small sample size, a normative group fitting the mean age (73 years), years of education (12 years), and the majority gender (female) of the current sample were used to identify standard cut points. A receiver-operating characteristic analysis was used to compare predicted diagnosis with actual clinical diagnosis obtained through patients' overall performance and a consensus meeting by licensed clinicians.

Results: Younger age ($p < 0.05$) and being tested in Mandarin ($p < .01$) were positively associated with immediate and delayed recall. Strong positive correlations between each measure were observed (all $p < .001$), indicating a significant relationship between information encoded and retained. Among all the participants, 15 (19.2%) were diagnosed with MCI and 22 (28.2%) with dementia. For MCI diagnosis, the standard cutoff scores demonstrated adequate sensitivity (verbatim=82%, paraphrase=91%) but low specificity (verbatim=44%, paraphrase=67%) in all outcome measures. For dementia diagnosis, delayed recall showed strong sensitivity (100%) and adequate specificity (75%) in both verbatim and paraphrasing scores. Immediate recall paraphrase (sensitivity = 95%, specificity = 50%) showed a better sensitivity but lower specificity than verbatim scoring (sensitivity = 86%, specificity = 58%). The accuracy was higher in delayed recall for both MCI and dementia diagnosis. A preliminary analysis on the optimal cut points indicated higher cutoff scores to distinguish MCI and dementia from clinically cognitive normal population, and from each other (e.g., the optimal cut point for delayed verbatim in distinguishing MCI from normal is 8.0 (sensitivity=89%, specificity=73%, AUC=84.3%)).

Conclusions: Consistent with previous literature, Craft Story delayed recall served as a more accurate diagnostic tool for both MCI and dementia compared to immediate recall in older Chinese Americans. However, poor specificity might increase the chance of following false positive subjects in clinical trials. In addition, testing language appeared to impact performance on verbal memory recall of constructed information. Thus, future studies