Abigail Baird[1,2], Cesilia Mejia[2], Jane Booth[2], Robert Collins[2], Nicholas Wisdom[2]
[1]Texas Woman's University, Denton, TX, USA.
[2]Neurocognitive Specialty Group, Houston, TX, USA

Objective: Empirical support for inclusion of performance validity testing (PVTs) in neuropsychological assessment continues to grow (Sweet et al., 2021). However, considerable validation is still needed to understand the impact of culturally mediated factors on the reliability of current, commonly used PVTs to accurately classify effort among various cultural groups. This study sought to contribute to the literature by examining the utility of several PVTs in a non-clinical, community-dwelling sample in Kampala, Uganda.

Participants and Methods: Participants included 52 residents (25 Female, 27 Male) who were born between 1953-2003 from the Wabigalo community of central Kampala. Individuals were recruited by community leaders and volunteered to participate. All 52 participants were administered the Dot Counting Test (DCT; Boone et al., 2002), Test of Memory Malingering (TOMM; Tombaugh, 1997), and Rey 15-Item Memorization Test (Rey 15; Lezak, 1995). Twenty-five participants also completed Green's Non-Verbal Medical Symptom Validity Test (NV-MSVT; Green, 2006). Data from three participants was excluded due to suspected memory concerns. Instructions for all tests were translated into Luganda by a professional translator with experience in Luganda and were administered by Luganda-speaking individuals.

Results: Using test manual-derived cut scores, 71.4% (n = 35) participants scored in the invalid range on the DCT, 10.2% (n = 5) produced total combined scores in the invalid range on Rey 15, 6.1% (n = 3) failed TOMM Trial 2, and one participant (4.3%) exceeded cut-offs on Green's NV-MSVT.

Conclusions: In this non-clinical sample, manual cutoffs for DCT contributed to a high type-1 error rate. These findings suggest that culturally mediated factors may contribute to differences in engagement or performance on DCT. Future studies should explore these factors and continue to examine the utility of widely used tests in diverse samples.

---

77 The Shell Game Task: Pilot Data Using a Simulator-Design Study to Evaluate a Novel Attentional Performance Validity Test

Andrew M. Bryant[1,2], Kendra Pizzonia[1], Claire Alexander[1], Grace Lee[1], Olivia Revels-Strother[1], Sarah Weekman[1], Brianna Hart[1], Julie Suhr[1]
[1]Ohio University, Athens, OH, USA. [2]The Ohio State University, Columbus, OH, USA

Objective: We developed the Shell Game Task (SGT) as a novel Performance Validity Test (PVT). While most PVTs use a forced-choice paradigm with "memory" as the primary domain being assessed, the SGT is a face-valid measure of attention and working memory. We explored the accuracy of the SGT to detect noncredible performance using a simulator-design study.

Participants and Methods: Ninety-four university students were randomly assigned to either best effort (CON) (n=49) or simulating traumatic brain injury (TBI) (SIM) (n=45) conditions. Participants completed a full battery of neuropsychological tests to simulate an actual evaluation, including the Test of Memory Malingering (TOMM) and the SGT. The SGT involves three cups and a red ball shown on the screen. Participants watch as the ball is placed under one of the three cups. Cups are then shuffled. Participants are asked to track the cup that contains the ball and correctly identify its location. We created two difficulty levels (easy vs hard, 20 trials each) by changing the number of times the cups were shuffled. Participants were given feedback (correct vs incorrect) after each trial. At the conclusion of the study, participants were asked about adherence to study directions they were given.

Results: Participants with missing data (CON=1; SIM=2) or who reported non-adherence to study directions (CON=2; SIM=1) were removed from analyses. Twenty-five percent in SIM and 0% in CON failed TOMM

Trial 2 (<45) suggesting adequate manipulation of groups. Groups were not different in age, gender, ethnicity, or education (all p's>.05). There were 9 participants in each group with concussion/TBI history. TBI history was not significantly related to performance on the SGT in either group, although participants with TBI history tended to do better. Average performances on TOMM Trial 1 (36.62 vs 47.91, p<.001) and TOMM Trial 2 (37.50 vs 49.71, p<.001) were significantly lower in the SIM group. Performance on SGT was also significantly lower in the SIM group across SGT Total Correct (20.17 vs 24.65 of 40, p=.008), SGT Easy (10.60 vs 13.52 of 20, p=.002), and SGT Hard (9.57 vs 11.13 of 20, p=.068). Mixed ANOVA showed a trend towards significant group by SGT difficulty interaction ($F(1,86)=3.41$, p=.052, np2=.043). There was steeper decline in performance on SGT Hard compared to SGT Easy for CON. ROC analyses suggested adequate but not ideal sensitivity/specificity: scores <8 on SGT Easy (sensitivity=26%; false positive=11%), <7 on SGT Hard (sensitivity=26%; false positive=7%), and <15 on SGT Total (sensitivity=24%; false positive=9%).

Conclusions: These preliminary data indicate the SGT may be able to detect malingered TBI. However, additional development of this measure is necessary. Further refinement of difficulty level may improve sensitivity/specificity (e.g., CON mean performance for SGT Easy trails was 13.52, suggesting the items may be too difficult). This study was limited to an online administration due to COVID, which could have affected results; future studies should test in-person administration of the SGT. In addition, performance in clinical control groups (larger samples of individuals with mild TBI, ADHD) should be tested to better determine specificity for these preliminary cutoffs.

Categories: Forensic Neuropsychology/Malingering/Noncredible Presentations
Keyword 1: test development
Keyword 2: noncredible presentations
Keyword 3: performance validity
Correspondence: Andrew M. Bryant, The Ohio State University, andrew.bryant@osumc.edu

---

78 Utility of the D-KEFS Color Word Interference Test as a Measure of

Performance Validity in Adults Referred for a Psychoeducational Evaluation

Anthony Robinson, Eathan Breaux, Marissa Huber, Matthew Calamia
Louisiana State University, Baton Rouge, LA, USA

Objective: Previous investigations have demonstrated the clinical utility of the Delis-Kaplan Executive Function System (D-KEFS) Color Word Interference Test (CWIT) as an embedded validity indicator in mixed clinical samples and traumatic brain injury. The present study sought to cross-validate previously identified indicators and cutoffs in a sample of adults referred for psychoeducational testing.
Participants and Methods: Archival data from 267 students and community members self-referred for a psychoeducational evaluation at a university clinic in the South were analyzed. Referrals included assessment for attention-deficit hyperactivity disorder, specific learning disorder, autism spectrum disorder, or other disorders (e.g., anxiety, depression). Individuals were administered subtests of the D-KEFS including the CWIT and several standalone and embedded performance validity indicators as part of the evaluation. Criterion measures included The b Test, Victoria Symptom Validity Test, Medical Symptom Validity Test, Dot Counting Test, and Reliable Digit Span. Individuals who failed 0 criterion measures were included in the credible group (n = 164) and individuals failing 2 or more criterion measures were included in the non-credible group (n = 31). Because a subset of the sample were seeking external incentives (e.g., accommodations), individuals who failed only 1 of the criterion measures were excluded (n = 72). Indicators of interest included all test conditions examined separately, the inverted Stroop index (i.e., better performance on the interference trial than the word reading or color naming trials), inhibition and inhibition/switching composite, and sum of all conditions.
Results: Receiver Operating Characteristics (ROC) curves were significant for all four conditions (p < .001) and the inverted stroop index (p = .032). However, only conditions 2, 3 and 4 met minimal acceptable classification accuracy (AUC = .72 - 81). ROC curves with composite indicators were also significant (p < .001), with all three composite indicators meeting minimal acceptable classification