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

accuracy (AUC = .71- .80). At the previously identified cutoff of age corrected scale score of 6 for all four conditions, specificity was high (.88 - .91), with varying sensitivity (.23 - .45). At the previously identified cutoff of .75 for the inverted stroop index, specificity was high (.87) while sensitivity was low (.19). Composite indicators yielded high specificity (.88 - .99) at previously established cutoffs with sensitivity varying from low to moderate (.19 - .48). Increasing the cutoffs (i.e., requiring higher age corrected scale score to pass) for composite indicators increased sensitivity while still maintaining high specificity. For example, increasing the total score cutoff from 18 to 28 resulted in moderate sensitivity (.26 vs .52) with specificity of .91.

Conclusions: While a cutoff of 6 resulted in high specificity for most conditions, the sum of all four conditions exhibited the strongest classification accuracy and appears to be the most robust indicator which is consistent with previous research (Eglit et al., 2019). However, a cutoff of 28 as opposed to 18 may be most appropriate for psychoeducational samples. Overall, the results suggest that the D-KEFS CWIT can function as a measure of performance validity in addition to a measure of processing speed/executive functioning.

Categories: Forensic Neuropsychology/Malingering/Noncredible Presentations
Keyword 1: performance validity
Keyword 2: validity (performance or symptom)
Keyword 3: noncredible presentations
Correspondence: Anthony Robinson, Louisiana State University, arob181@lsu.edu

---

79 The Clinical Utility of Reliable Digit Span in Predicting Effort on the TOMM in Youth with a History of a Concussion

Ashlynn Steinbaugh, Rachel Murley, Bradley R Forbes, Daniel Baldini, Talamahe'a A Tupou, Julius Flowers, Rayna B Hirst
Palo Alto University, Palo Alto, California, USA

Objective: Performance validity testing (PVT) is important in neuropsychological evaluations to ensure accurate interpretation of performance. While research shows children pass PVTs with adult cut-offs, PVTs are more commonly used with adults (Lippa, 2018). The Test of Memory Malingering (TOMM), a standalone PVT, is commonly used with adults and children (DeRight & Carone, 2015). The Reliable Digit Span (RDS), an embedded PVT derived from the Digit Span subtest of the Wechsler Intelligence Scales (Wechsler Intelligence Scale for Children-4th Edition, WISC-IV; Wechsler, 2003), is less commonly used with children (DeRight & Carone, 2015). RDS cut-off scores are associated with an increased rate of false positives in children, indicating mixed results regarding the clinical utility in pediatric populations (Welsh et al., 2012). Research shows that youth with a history of concussion (HOC) may demonstrate suboptimal effort for many reasons (e.g., external incentives, boredom, pressure), thus highlighting the need to investigate the utility of PVTs in this population (Araujo et al., 2014; DeRight & Carone, 2015). The present study aimed to examine the clinical utility of RDS in detecting poor effort on the TOMM in youth athletes with a HOC.

Participants and Methods: Participants included 174 youth athletes aged 8 to 18 (20.1% female; 42.5% people of color (POC)) who completed a baseline neuropsychological battery that included the TOMM and WISC-IV Digit Span. Of the sample, 29 youth athletes reported a HOC (13.8% female; 37.9 POC). RDS was calculated for each Digit Span administration, and sensitivity (SN) and specificity (SP) were calculated for RDS when invalid performance was operationalized by a more stringent cut-off score of <49 to increase the SN of the TOMM Trial 1 (Stenclik et al., 2013). Receiver operator characteristics (ROC) curve analysis determined whether RDS performance accurately predicted participants' performance on the TOMM.

Results: The ROC curve analysis resulted in an area under the curve (AUC) of just 0.427 for RDS. A cut-off score of <7 (as suggested by Kirkwood et al. (2011)) for RDS results in 100% SN, 8.3% SP, 5% positive predictive validity (PPV), and 95% negative predictive validity (NPV). However, a cut-off score of <9 for RDS results in 75% SN, 15% SP, 25% PPV, and 75% NPV.

Conclusions: Little research shows the utility of different PVTs predicting children's performance on other PVTs, despite evidence that children with a HOC are vulnerable to variable or insufficient effort (Araujo et al., 2014; DeRight & Carone, 2015). In a sample of 29 youth athletes with a HOC, RDS predicted TOMM performance