lengthy scoring rubrics. We present preliminary data on an adaptation to an existing scoring system that provides quantifiable scores, can be implemented with reliability, and reduces scoring time.
Participants and Methods: Data was taken from two large-scale clinical trials focusing on EF in autistic youth. All participants completed the ROCF following standard administration guidelines. The research team reviewed commonly used scoring systems and determined that the Boston Qualitative Scoring System (BQSS) was the best fit due to its strengths in measuring EF, the process-related variables generated, and the available normative data. Initially, the BQSS full scoring system was used, which resulted in comprehensive scores but was not feasible due to the time required (approximately 1-1.5 hours per figure for research assistants to complete scoring). Then, the BQSS short form was used, which was successful at solving the timing problem, but resulted in greater subjectivity in the scores impacting the team's ability to become reliable. Independent reliability could not be calculated for this version because of the large number of discrepancies among scorers which included 2 neuropsychologists and 4 research assistants. A novel checklist was then developed that combined aspects of both scoring systems to help promote objectivity and reliability. In combination with this checklist the team created weekly check in meetings where challenging figures could be brought to discuss. Independent reliability was calculated amongst all research assistant team members (n=4) for the short form and novel checklist. Reliability was calculated based on (1) if the drawing qualified for being brought to the whole team and (2) individual scores on the checklist.
Results: Independent reliability was calculated for 10 figures scored utilizing the novel checklist by a team of 4 trained research assistants. All scorers were able to achieve 80% reliability with a high average (80-86%). Study team members reported that scoring took less time taking on average 30-45 minutes per figure.
Conclusions: Inter-rater reliability was strong on the checklist the study team created, indicating its potential as a useful adaptation to the BQSS scoring system that reduces time demands, making the tool feasible for use in large-scale clinical research studies with initially positive reliability factors. The checklist was easy to use, required little training and could be completed quickly. Future research should continue to examine the reliability of the checklist and the time it takes to complete. Additionally, the ROCF should be studied more broadly in research and examined as a potential outcome measure for large scale research studies.

Categories:
Assessment/Psychometrics/Methods (Child)
Keyword 1: autism spectrum disorder
Keyword 2: executive functions
Keyword 3: assessment
Correspondence: Rebecca Handsman Center for Autism Spectrum Disorders, Children's National Hospital rebecca.handsman@du.edu

73 Examining the Associations Between Sentence Repetition and Other Cognitive Abilities in a Clinical Sample of School-Aged Children

Rebekah E Taylor, Joseph E Casey, Ava M Flanagan
University of Windsor, Windsor, Ontario, Canada

Objective: Sentence repetition (SR) task performance is related to various cognitive abilities and not just learning and memory, as is commonly considered (e.g., Baron, 2018). Bartlett (2018) was the first to examine the associations among SR performance and other cognitive abilities within a single study, using a normative sample. Bartlett (2018) found that SR performance was predicted by language, auditory verbal working memory, processing speed, and nonverbal cognitive ability of which only language abilities and auditory verbal memory significantly added to the prediction. However, no study to date has examined the associations between SR and other cognitive abilities in a clinical sample of school-aged children. The present study sought to determine the extent to which language, working memory, nonverbal abilities (visuospatial processing and fluid reasoning), and processing speed predict children's SR in a clinical sample.
Participants and Methods: Children 6 to 14 years of age (N = 191; 65% males) were included in the present study. Participants were drawn from two separate archival data populations of children referred for neuropsychological assessment in southwestern Ontario. SR scores were obtained from

https://doi.org/10.1017/S1355617723009311 Published online by Cambridge University Press

performance on Benton's (1965) sentence repetition task. Language, working memory, fluid reasoning, visual perception, and processing speed were measured with Index scores from the Wechsler Intelligence Scale for Children (5th edition). The association of each of these cognitive domains with SR was determined by multiple linear regression. The effects of age and sex on SR were also examined (N = 226; 64% males).
Results: A multiple linear regression model including the five independent variables significantly predicted SR performance, $F(5, 185) = 30.306$, $p < .001$., *adj.* $R^2 = .435$. Only language and working memory added significantly to the prediction, $p < .05$. A mediation analysis demonstrated that processing speed indirectly predicted SR performance through working memory, b=.0241, [95% BCa [CI .0132, .0355]. A moderate positive correlation was found between age and SR performance, $r(226) = .416$, $p < .001$. Sex was unrelated to SR performance.
Conclusions: The findings from this study are consistent with other studies indicating that SR taps multiple cognitive abilities. In a large and representative clinical sample of children referred for assessment due to academic or other learning difficulties, language plays as significant a role in SR performance as does auditory verbal attention and working memory. An advantage of the present study was the use of clinically relevant summary measures of cognitive domains associated with intelligence testing.

Categories:
Assessment/Psychometrics/Methods (Child)
Keyword 1: psychometrics
Correspondence: Rebekah E. Taylor, University of Windsor, taylo136@uwindsor.ca

74 Embedded Performance Validity Utilizing the WISC-V Figure Weights Subtest

Sakina Butt[1], Austin Sellers[2], Sharon Ghazarian[3], Jennifer Katzenstein[1]
[1]Institute for Brain Protection Sciences, Johns Hopkins All Children's Hospital, St Petersburg, Florida, USA. [2]Johns Hopkins All Children's Institute for Clinical and Translational Research, Johns Hopkins All Children's Hospital, St Petersburg, Florida, USA. [3]Johns Hopkins All Children's Institute for Clinical and Translational Research, St Petersburg, Florida, USA

Objective: This study aimed to explore the possibility of using the Figure Weights subtest of the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V) as an embedded validity test (EVT).
Participants and Methods: We conducted a retrospective cohort study of patients seen in the Johns Hopkins All Children's Hospital (JHACH) Neuropsychology program between 2015 and 2019. Patient age ranged from 6-15 years (median age 11 years). All patients were administered the WISC-V as a portion of their neuropsychological evaluation. The sample included 75 patients who were generally male (63%), White (77%), English dominant (97%), and right-handed (81%).
Results: Effort determination based on RDS identified more patients as having invalid effort. Clinicians identified only 7% of patients with invalid effort; whereas, 16% of patients with invalid effort were identified using the WISC-V RDS. Although patients having valid effort generally performed better on all WISC-V subtests, no significant differences between groups were found. Over 90% of patients were able to get items 1-10 correct on the WISC-V Figure Weights subtest regardless of their effort determination. WISC-V Figure Weights item analysis showed participants in the invalid group sometimes answered more difficult questions correctly while failing easier items which supports inconsistent effort. Further statistical analyses, including discriminant validity tests, were unable to be completed given the study was underpowered due to significant disparity between effort groups.
Conclusions: This study shows support for WISC-V Figure Weights subtest items 1-10 as an embedded EVT given these items were sufficiently easy to pass regardless of whether participant gave valid or invalid effort. As this is an exploratory study, results will need to be replicated in other pediatric samples. Additionally, the discriminant ability of the WISC-V Figure Weights subtest EVT will need to be further investigated.

Categories:
Assessment/Psychometrics/Methods (Child)
Keyword 1: effort testing
Keyword 2: performance validity

https://doi.org/10.1017/S1355617723009311 Published online by Cambridge University Press