



Psychometric Properties of Attention Measures in Young Children with Neurofibromatosis Type 1: Preliminary Findings

Sara K. Pardej¹ · Kristin M. Lee¹ · Danielle M. Glad¹ · Bonita P. Klein-Tasman¹

Received: 11 April 2022 / Revised: 6 July 2022 / Accepted: 20 July 2022 / Published online: 23 August 2022
© The Author(s), under exclusive licence to American Academy of Pediatric Neuropsychology 2022

Abstract

Children with neurofibromatosis type 1 (NF1) often demonstrate difficulties with attention and executive functioning that can be evident starting at a young age. There has been little research about which measures of attention are most suitable for use with young children with NF1. This pilot study explored several computerized measures of attention, a digits forward task, and parent report measures of attention to compare their reliability, validity, and the degree to which they capture attention difficulty in this population. Participants with NF1 ages 4 to 6 years were seen for one ($n=2$) or two ($n=18$) time points. Statistical analyses for evaluating evidence for test-retest reliability, convergent and discriminant validity, practice effects, and identification of difficulties were conducted. Each measure demonstrated relative strengths and weaknesses, and there may not be a “one size fits all” measure for use with young children with NF1. However, the Behavior Rating Inventory of Executive Function Preschool/Second Edition, Conners Early Childhood Inattention/Hyperactivity Scale, and the Conners Kiddie Continuous Performance Test Second Edition generally had the highest reliability and most evidence of validity. More specific recommendations are provided for the appropriate measure to use in clinical and research batteries.

Keywords Neurofibromatosis type 1 · Attention · Reliability · Validity · Psychometrics

Introduction

While neurofibromatosis type 1 (NF1) is associated with several medical and cognitive difficulties, one of the most apparent cognitive difficulties is attention, with the vast majority of research focusing on the school-age years. Attention deficit/hyperactivity disorder (ADHD) and attentional difficulties more broadly are prevalent across many genetic syndromes, including NF1 (Lo-Castro et al., 2010). Although 40–60% of children with NF1 meet criteria for ADHD, much higher than in the general population (Lehtonen et al., 2013; Polanczyk et al., 2014), an even larger proportion demonstrate difficulties with attention and executive function (Beaussart et al., 2018; Casnar & Klein-Tasman, 2017; Isenberg et al., 2013; North et al., 2002; Plasschaert et al., 2016). Generally, children with NF1 more frequently have difficulties relating to inattention than hyperactivity (Lidzba et al., 2012; Payne et al., 2011). While attention and executive difficulties are evident based on

the NF1 literature, there is no study to date that demonstrates the reliability or validity of attention and emerging executive measures in young children with NF1.

Attention and Executive Difficulties in Children with NF1

Attention and executive difficulties have been described as a core deficit for many children with NF1 (Templer et al., 2013) with significant difficulties with inhibition and sustained, selective, and focused attention (North et al., 2002). In addition to those children with NF1 who meet criteria for ADHD, there are children who demonstrate difficulties with visual and auditory attention, divided attention, sustained attention, shifting attention, working memory, and response inhibition (Casnar & Klein-Tasman, 2017; Isenberg et al., 2013; North et al., 2002). These findings are clinically relevant, as children with NF1 who exhibit inattentive and/or hyperactive problems tend to have lower overall intellectual functioning than children with NF1 who do not exhibit any attention difficulties (Lidzba et al., 2012). Furthermore, some findings suggest that executive function deficits may be an inherent part of NF1 and not merely due to either lower intellectual functioning or ADHD (Plasschaert et al., 2016). Further, Huijbregts (2012)

✉ Sara K. Pardej
skpardej@uwm.edu

¹ Department of Psychology, University of Wisconsin – Milwaukee, 2441 E Hartford Ave, Milwaukee, WI 53211, USA

summarized the literature including a conclusion that there are significant differences between attention features in NF1 and idiopathic ADHD at the behavioral and neuroimaging levels. Indeed, there is a lack of consensus in terms of neurological explanations for the cognitive phenotype of NF1, including attention (Baudou et al., 2020). There is some evidence of the corpus callosum volume and T2 hyperintensities in the thalamic region impacting attentional control and distractibility respectively, as well as activity in the posterior cingulate cortex and frontal pole hypoactivation being related to selective attention in NF1 (Kayl et al., 2000; Moore et al., 1996).

Working memory is consistently found to be lower in children with NF1 when compared to unaffected controls and siblings (Gilboa et al., 2014; Lehtonen et al., 2015; Payne et al., 2011, 2012; Plasschaert et al., 2016; Templer et al., 2013). A recent meta-analytic review found a moderate effect size for working memory impairment in children with NF1, as well as data to suggest that executive dysfunction worsens with age (Beaussart et al., 2018). Although the literature on young children is more limited, these difficulties appear to be evident early in life. Sangster et al. (2011) found that parent ratings of working memory on a questionnaire measure of executive functioning remained significantly lower in the NF1 group than a control group even after accounting for maternal education and individual intellectual functioning. Both parent and teacher ratings of working memory have indicated challenges in young children with NF1 (Casnar & Klein-Tasman, 2017), such that there is evidence that these difficulties are present and identifiable in multiple settings. While parent ratings of working memory are often based on items that show conceptual overlap with attention problems, there is some evidence that ratings of working memory are related to performance on some lab-based tasks (Casnar & Klein-Tasman, 2017).

The Necessity of Research About Psychometric Properties in NF1

Characterizing attention in young children with NF1 is particularly challenging because the behavioral phenotype of these children is quite diverse and there is considerable variability in functioning in young children (Mahone, 2005). Many developmental studies of attention in young children use experimental measures that do not have established psychometric properties nor do they have normative data (Mahone & Schneider, 2012). This prevents researchers from drawing clinical conclusions about the nature and severity of the difficulties experienced by young children with NF1 both in cross-sectional and longitudinal methodologies. It is important to identify measures with strong psychometric properties because attention problems in early childhood may indicate vulnerability to difficulties later on in life, such as poorer academic outcomes (Washbrook et al., 2013). The

mentioned differences between attentional profiles of children with NF1 and those with idiopathic ADHD are important to consider in the context of test development; when attention measures feature a clinical group for normative purposes, they are typically homogenous ADHD groups. Furthermore, children with NF1 are likely underrepresented in normative data, and normative data by design include mostly individuals with fewer attention difficulties. Thus, it is important to establish whether these measures demonstrate the same level of reliability and validity in a population that has more difficulties with attention and executive abilities than seen in typically developing participants and a pattern of attention difficulties that may also differ from that seen in “typical” ADHD.

Although conventional paper-pencil neuropsychological measures of attention exist that have been used in research with this population, there are new computerized tasks with more recent normative data that may be suitable for assessing young children with NF1 and are often designed with the idea of tracking change over time or with intervention. For example, challenges on the Conners Kiddie Continuous Performance Test have been shown to be an early indicator of executive difficulties in preschool-aged children (Barnard et al., 2018). Computerized measures of attention offer other advantages, including administration without a neuropsychologist present and more updated normative data. Despite these advantages, there are no comparative studies of validated attention measures to date that have investigated the performance of young children with NF1. There is a growing need for psychometric research to establish the most valid and reliable measures that can capture attention difficulties early in development in NF1 populations. This research is necessary for identification of measures most suitable for use in clinical trials research to help improve outcomes of children with NF1.

The Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) group, which is comprised of experts in the field, announced a need to identify measures to use as endpoints for clinical trials of attention in young children with NF1 (Walsh et al., 2016). The group noted that there is a gap in the literature concerning which measures of attention (including parent report measures, performance-based paper and pencil measures, and performance-based computerized measures of attention) are most appropriate for use with young children with NF1. By identifying the most promising measures for use with this population, research investigating the development and trajectory of attention and executive difficulties in children with NF1 will be more compelling. Furthermore, having reliable and valid measures will allow investigators to more accurately evaluate the effectiveness of interventions in this population. Research in this area would help to more rigorously examine the characteristics of young children who are at the highest risk of developing attention

deficits into later childhood and beyond (Mahone, 2005; Mahone & Schneider, 2012).

The Present Study

The goal of the present study is to present preliminary data to begin to identify reliable and valid measures of attention for young children with NF1. To be successful, clinical trials require accurate measurement tools that have demonstrated validity, test-retest reliability, and minimal practice effects (Walsh et al., 2016). Although there is variability in the behavioral phenotype, the literature clearly demonstrates that attention is a frequent area of concern for children with NF1 as indicated by both parents and teachers. Thus, it is important that we explore which measures of attention are most appropriate to use. The utility of several measures was examined in the present study: the NIH Toolbox Flanker, Dimensional Change Card Sort task, (Zelazo et al., 2013), List Sort Working Memory task (Tulsky et al., 2015), the Cogstate Identification Task (Cogstate, 2018), Conners Kiddie Continuous Performance Test Second Edition (Conners, 2015), Differential Abilities Scale Second Edition Digits Forward (DAS-II DF; Elliot, 2007), Behavior Rating Inventory of Executive Function (BRIEF; Gioia et al., 2003, 2015) and Conners Early Childhood Behavior Inattention/Hyperactivity Scale (Conners I/H; Conners 2009). These are neuropsychological tasks and parent report questionnaires that measure response inhibition, shifting attention, working memory, or sustained attention. The current study examined test retest reliability using both intraclass correlation coefficients and Pearson correlations. Convergent and discriminant validity were explored using Spearman correlations, due to small sample size. Measures were also compared based on how frequently they identified attention and executive difficulties in the present sample. Lastly, paired sample *t*-tests were conducted to identify any practice effects.

Methods

Participants

Participants were recruited using fliers distributed through the National Neurofibromatosis Research Registry and several Midwestern Neurofibromatosis clinics. Inclusion criteria included (a) diagnosis of NF1 by a physician, (b) having a mutation of the NF1 gene, (c) aged 4–6 years old, and (d) first and main language spoken in the home is English. The exclusion criteria included (a) not have had a major surgery or hospitalization in the past 6 months (anesthesia could impact cognitive functioning for 6 months post hospitalization), (b) deletion of the NF1 gene, and (c) not have any other genetic neurodevelopmental disorder that has a global impact on

functioning (to limit the impact of potentially confounding variables).

Twenty-two participants with NF1 were participated in the study. Eighteen children were assessed at two different time points, 8±2 weeks apart, to allow for test-retest reliability analyses. Two children were seen at one time point but did not return for a second appointment because of family circumstances (*n*=1) and COVID-19 (*n*=1). Two additional children were consented but were unable to complete the majority battery due to behavioral challenges and were therefore not included.

Thus, the present sample includes 20 children with NF1 ages 4 through 6 ($M_{age}=5.45$, $SD=0.75$). There were slightly more males (*n*=12) than females (*n*=8). Ninety-five percent of the sample was white. There were more sporadic (*n*=12) than familial (*n*=8) cases. The mean Hollingshead Index score ($M=46.15$, $SD=10.75$) suggests the average family was middle class.

Procedure

Consent documents, along with some questionnaire measures, were mailed to families prior to the first appointment. Participants were administered an age-appropriate battery consisting primarily of attention and executive measures by trained members of the study team. The battery also included a measure of cognitive functioning. There were three versions of the battery to allow for counterbalancing of the order of administration of the attention and executive functioning tasks. The Differential Ability Scales-Second Edition was administered first in each version.

The first session lasted approximately 3 hours, and the second session (8±2 weeks later) lasted about 2.5 hours. All assessments took place either at a university research lab or in a quiet conference room if the family was unable to travel. All assessments are conducted according to each measure's standardization procedures. Information about the normative samples and procedures for the computerized measures are detailed in Table 1. Parents were compensated after each appointment.

Measures

Differential Ability Scales-Second Edition (DAS-II; Elliot, 2007)

The DAS-II core was administered to determine overall cognitive functioning. This measure is frequently used in behavioral phenotyping research because it is able to characterize both strengths and weaknesses in a child's functioning (Baron et al., 2011; Bishop et al., 2011; Gillentine et al., 2017). The measure yields an overall General Cognitive Ability (GCA) standard score ($M=100$, $SD=15$). The Digits Forward task (DF) was used as a measure of attention (Casnar & Klein-

Table 1 Normative data information for the Cogstate, NIH Toolbox, K-CPT 2, and DAS-II Digits Forward (DF)

Measure	Age range (years)	4–6-year-old sample size	Sampling
Cogstate	4–99	134	Not available
NIH Toolbox	3–85	391 English, 296 Spanish	Matched U.S. Census data
K-CPT 2	4–7	320 normative, 152 clinical	Matched U.S. Census data
DAS-II DF	4:0–6:11	752	Matched U.S. Census data

K-CPT 2, Conners Kiddie Continuous Performance Test Second Edition; *DAS-II*, Differential Ability Scales-Second Edition

Tasman, 2017). The DF task yields a T-score ($M = 50$, $SD = 10$). Higher scores indicate better performance. The DAS-II demonstrates excellent reliability, validity, and standardization.

National Institute of Health (NIH) Toolbox Selected Subtests

The NIH Toolbox is an electronic battery that has a variety of measures of cognitive, emotional, sensory, motor, attention, and executive functioning. The NIH Toolbox is not designed for the purposes of clinical application to diagnostic concerns with individual patients, but instead is useful for developmental studies and clinical trials research making it a strong candidate for inclusion in clinical trials related to NF1. It has demonstrated good psychometric properties across measures in the typically developing population. All NIH toolbox measures were administered via iPad. For each NIH Toolbox measure, an age-adjusted standard score was used. On each measure, higher scores indicate better performance.

The NIH Toolbox Flanker task (Zelazo et al., 2013) required children to choose whether the middle stimulus (a fish with an arrow on it) was pointing left or right. On congruent trials, all of the stimuli were pointing in the same direction. On incongruent trials, the middle stimulus pointed in the opposite direction from the remaining stimuli. Administration included practice, which was repeated three times or until criterion is met, followed by the test. If the child was accurate on $\geq 80\%$ of trials, the final score incorporated both accuracy and reaction time. Otherwise, only the accuracy score was provided.

The NIH Toolbox Dimensional Change Card Sort (DCCS) task (Zelazo et al., 2013) required children to sort a middle stimulus either by shape or color. Sometimes the color of the middle stimulus was incongruent with the prototype of the same shape that remains at the bottom of the screen, thus requiring the child to shift between the two sets (i.e., color, shape). Administration included practice, preswitch, postswitch, and mixed blocks. If the child was accurate on $\geq 80\%$ of trials, the final score incorporated both accuracy and reaction time. Otherwise, only the accuracy score was provided.

The NIH Toolbox List Sort Working Memory (LSWM) task (Tulsky et al., 2015) is a sequencing task in which participants must remember a series of animals and/or fruit and repeat them in size order. In initial trials, they were only presented with one type of stimulus (i.e., animal, fruit). If they

were able to complete the initial trial to criterion, then they repeated various series of stimuli by first saying the fruit in size order, followed by the animals in size order. Standard scores were based on the sum of the total correct responses.

The NIH Toolbox Picture Vocabulary Test (TPVT) task (Gershon et al., 2015) is a measure of receptive vocabulary. Participants were presented with four images. The iPad played an audio recording of a word, and the participant chose which image best depicts the word. They were permitted to return to previous items and hear the word multiple times. The yielded score is a standard score.

Pediatric Version of the Cogstate Identification Task (Cogstate, 2018)

On the Identification task, participants were told to wait until each card turns over and to press “yes” if the card is red and “no” if it is black. The task was administered using an iPad. The primary outcome on the measure was \log_{10} transformed reaction time, which was converted to a T-score ($M = 50$, $SD = 10$) for analyses, with higher scores indicating more impaired performance.

Conners Kiddie Continuous Performance Test-2 (K-CPT 2; Conners, 2015)

The K-CPT 2 is a computerized measure of attention for children 4–7 years old. This measure was approximately 7.5 min and consisted of 200 scored trials. T-scores ($M = 50$, $SD = 10$) were provided for Detectability (“d”), Omissions, Commissions, Perseverations, Hit Reaction Time (HRT), Variability, Hit Reaction Block Change (HRT BC), and Hit Reaction Inter Stimulus Interval (HRT ISI). Higher scores indicated more impaired performance. Participants were instructed to press a key for every stimulus except the target stimulus. The K-CPT 2 has strong validity, reliability, and sensitivity.

Conners Early Childhood Behavior Form, Parent Version (Conners EC; Conners, 2009)

The Conners EC is a 113-item questionnaire that was administered to parents. It is a global measure of behavioral,

emotional, and developmental functioning for children 2–6 years old. Only the Inattention/Hyperactivity T-score was used in the present analyses. Higher scores indicated more impairment. The Conners EC has demonstrated good validity and reliability.

Behavior Rating Inventory of Executive Function-Preschool Edition or Second Edition (BRIEF-P; Gioia et al., 2003, 2015)

The BRIEF-P is a measure of executive function for children 2–5 years old and was administered to parents of 4- and 5-year-olds. The BRIEF-2 measures executive function in children 5–18 years old and was administered to parents of 6-year-olds. Both measures consist of 63 items. The present analyses used the Inhibit, Shift, Emotional Control, Working Memory, and Global Executive Composite (GEC) scales, as those were available across both versions and yielded T-scores. The GEC T-score consists of all of the scales. Higher scores indicated more impairment. Each version has well-established reliability and validity.

Data Analysis

The data were analyzed using IBM SPSS for Windows, version 25. Findings are interpreted using both statistical significance and effect size. Given the pilot nature of this work, uncorrected data are presented in the manuscript and are interpreted; however, corrected results are available in the tables. Corrected results use the false discovery rate (FDR; Benjamini & Hochberg, 1995) approach. A p value of $<.05$ was used to determine significance for uncorrected results and a q value of $<.05$ was used for corrected results. The following interpretations were used for Cohen's d : negligible effect = 0–.14; small effect = .15–.39; medium effect = .40–.74; large effect = .75 and above.

Results

Procedure Completion Rates

Analyses were based on the 20 participants who completed at least one assessment visit. Note that, as mentioned in the “Participants” section, two additional children were assented and began the battery but were unable to finish because of behavior challenges — it is possible that the measures were either unengaging or too difficult for their developmental level. These two children were excluded from all analyses. Table 2 summarizes the number of children in our sample who were unable to complete each specific measure or who did not pass validity indicators. Given that the validity indicators flag participants with a high number of omissions, and children with NF1 are known to have high omissions (Arnold

et al., 2018; Sangster et al., 2011), these participants were not excluded from the analyses. Given the pilot nature of the current study and the high number of participants who did not pass validity indicators (which could be due to task difficulty), these participants were included in analyses. Note that due to the small sample and preliminary nature of this work, analyses were underpowered. The present analytic plan was chosen to allow for comparison to each measure's normative data.

Individual Differences

The median DAS-II GCA of the sample was in the average range ($SS = 104$). Spearman correlations were run between each outcome measure, parent report score, and age at time 1. Age was significantly related to K-CPT 2 Variability ($\rho = .617, p = .008$), with older children performing significantly worse in comparison to same-aged peers than younger children. There were no other significant relations with age, likely at least partly due to the narrow age range. Independent samples t -tests were run to examine effects of NF etiology (spontaneous vs. inherited pathogenic variant) and sex on participant performance on each outcome score based on at both time points. Children with a familial mutation ($M_{DCCS} = 83.25, SD = 5.12$) performed significantly worse on the DCCS at time 1 than those with a sporadic mutation ($M_{DCCS} = 99.30, SD = 13.76$), $t(16) = -3.11, p = .007$. There were no significant differences based on sex.

Test-Retest Reliability

Intraclass correlation coefficients (ICC) were examined for all tasks. The ICC values from our sample are displayed in Table 3 alongside normative data when available. Using the standard cut-off of .75 (Koo & Li, 2016), Omissions, Shift, Emotional Control, Working Memory, GEC, Conners Inattention/Hyperactivity, and DAS-II DF demonstrated good test-retest reliability. The Flanker, Detectability, HRT, HRT SD, and Inhibit demonstrated moderate test-retest reliability (ICC between .5 and .75).

Pearson correlation coefficients from time 1 to time 2 were computed (note that Pearson rather than Spearman was used, despite the small sample size, to allow for comparison to the published normative data). The results are summarized in Table 3. Each measure was at least moderately correlated from time 1 to time 2, except for the DCCS and K-CPT 2 Variability scores.

Convergent and Discriminant Validity

Spearman correlations were conducted between each computerized measures' outcome scores to investigate evidence for convergent validity. The results are summarized in Table 4. For performance-based measures, DAS-II DF, Commissions

Table 2 Percentage of participants that successfully completed performance-based measures at time 1

Measure	Successful completion	Passed validity check
NIH Flanker	100%	N/A
NIH Dimensional Change Card Sort	90%	N/A
NIH List Sort Working Memory	60%	N/A
Cogstate Identification	95%	68%
Conners Kiddie Continuous Performance-2 (K-CPT 2)*	85%	64%
DAS-II Digits Forward	100%	N/A

*Most data were available for 100% of participants; however, because of participant response patterns, one participant had incomplete data

Note: *DAS-II*, Differential Ability Scales-Second Edition

and HRT BC generally demonstrated weak correlations with other measures of attention in comparison to the Identification, NIH Flanker, DCCS, LSWM, and the remaining K-CPT 2 outcome scores which showed stronger correlations with other measures of attention.

To further examine convergent validity, Spearman correlations were conducted between each computerized measure's outcome scores with the following parent-report scales: Conners Inattention/Hyperactivity, and BRIEF-P/BRIEF-2 Inhibit, Shift, Emotional Control,

Table 3 Intraclass correlation and Pearson correlation coefficients in the present sample and published literature

Measures	Sample ICC	Published ICC	<i>r</i>	Published <i>r</i>
NIH Toolbox				
Flanker	.61	.92	.67**^a	N/A
DCCS	.06	.92	.07	N/A
LSWM	.34	.77	.36	N/A
Cogstate				
Identification	.49	.79	.49*	.62
K-CPT 2				
Detectability	.61	N/A	.61*	.67
Omissions	.85	N/A	.85**^b	.62
Commissions	.49	N/A	.51*	.73
Perseverations	.43	N/A	.43	.39
HRT	.59	N/A	.62^a	.85
HRT SD	.65	N/A	.65**^a	.59
Variability	.27	N/A	.28	.21
HRT BC	.36	N/A	.37	.06
HRT ISI	.38	N/A	.38	.51
DAS-II DF	.82		.85**^b	
BRIEF				
Inhibit	.71	.90*	.74**^b	N/A
Shift	.86	.88*	.86**^b	N/A
Emotional Control	.79	.87*	.82**^b	N/A
Working Memory	.89	.85*	.89**^b	N/A
GEC	.89	.90*	.89**^b	N/A
Con I/H	.91		.92**^b	

* $p < .05$, ** $p < .01$, *** $p < .001$; ^a $q < .05$, ^b $q < .001$

Values are bolded to distinguish relations that were at least moderately correlated. *ICC*, intraclass correlation coefficient; *DCCS*, Dimensional Change Card Sort; *LSWM*, List Sort Working Memory; *K-CPT*, Conners Kiddie Continuous Performance Test Second Edition; *HRT*, Hit Reaction Time; *HRT BC*, Hit Reaction Time Block Change; *HRT ISI*, Hit Reaction Time Inter Stimulus Interval; *DAS-II DF*, Differential Ability Scales-Second Edition Digits Forward Task; *BRIEF*, Behavior Rating Inventory of Executive Function; *GEC*, Global Executive Composite; *I/H*, Inattention/Hyperactivity Scale

Table 4 Two-tailed Spearman correlations between each performance-based measure

	1 Ident.	2 Flanker	3 DCCS	4 LSWM	5 Detect.	6 Omiss.	7 Com.	8 Pers.	9 HRT	10 HRT SD	11 Var.	12 HRT BC	13 HRT ISI	14 DAS-II DF
1*	-													
2^	-.277	-												
3^	.010	.741**b	-											
4^	-.221	.450	.385	-										
5+	.342	-.313	-.456	-.474	-									
6+	.149	-.482	-.513*	-.304	.492*	-								
7+	-.002	.219	-.052	-.307	.458*	-.405	-							
8+	.336	-.245	-.398	-.318	.855**	.453	.282	-						
9+	.058	-.251	-.235	-.138	.251	.722**b	-.501*	.315	-					
10+	.375	-.466	-.345	-.292	.850**	.715**b	.104	.811**	.455	-				
11+	.365	-.583	-.281	-.275	.674**	.537*	.080	.822**	.373	.867**c	-			
12+	-.351	.172	.028	.534	-.144	-.204	.118	-.118	-.116	-.332	-.198	-		
13+	.454	-.378	-.292	-.396	.590**	.507*	.059	.721**	.371	.734**b	.695**a	-.418	-	
14	-.042	.155	.151	.578*	-.325	.155	-.544*	-.426	.240	-.139	-.299	.315	-.122	-

* $p < .05$, ** $p < .01$; ^a $q < .05$, ^b $q < .01$, ^c $q < .001$

Values are bolded to distinguish relations that were at least moderately correlated. *DCCS*, Dimensional Change Card Sort; *LSWM*, List Sort Working Memory; *K-CPT*, Conners Kiddie Continuous Performance Test Second Edition; *HRT*, Hit Reaction Time; *HRT BC*, Hit Reaction Time Block Change; *HRT ISI*, Hit Reaction Time Inter Stimulus Interval; *DAS-II DF*, Differential Ability Scales-Second Edition Digits Forward Task. **Cogstate*, ^*NIH Toolbox*, + *Conners Kiddie Continuous Performance Test Second Edition*

Working Memory, and GEC. The correlation values can be found in Table 5. The Flanker, DCCS, LSWM and every K-CPT 2 score except Commissions and HRT were at least moderately correlated ($|rho| < 0.30$) with the Inattention/Hyperactivity scale. The DAS-II DF, Detectability, Omissions, Perseverations, HRT, HRT SD, and Variability were at least moderately related to Inhibit. Detectability, Omissions, Perseverations, HRT SD, and HRT ISI were at least moderately associated with Shift. All scores except the Identification, Flanker, LSWM, Commissions, Variability, and HRT BC were at least moderately related to Emotional Control. The DAS-II DF, DCCS and Detectability were at least moderately related to Working Memory. All scores except LSWM, Commissions, and HRT BC were at least moderately related to GEC.

Spearman correlations were conducted between the TPVT and GCA with all of the performance-based measures' outcome scores to explore discriminant validity. The findings are in Table 5. Overall, most measures demonstrated weak correlations ($|rho| > 0.30$) with the TPVT. The DCCS ($rho = .606$) and K-CPT Perseverations ($rho = -.319$) were at least moderately related to the TPVT. Notably, Flanker and DCCS scores generally had higher relations with the TPVT than with parent ratings of behavior and many other performance-based measures. Many measures were highly related with GCA, including the Flanker, DCCS, and Detectability. Other measures were also moderately related ($|rho| < 0.30$) to GCA, including the DAS-II DF, Identification task, LSWM, Omissions, Perseverations, Variability, and HRT SD. Some of these measures were more highly associated with GCA than with parent-reported attention and executive difficulties.

Frequency of Difficulty Identification

Frequency of difficulties was examined using each dependent variable's standardized score (SS < 85 on measures where lower scores indicate greater difficulty or T > 60 on measures where higher scores indicate greater difficulty). Children who were not able to complete a measure were coded as having difficulty. Given that the K-CPT 2 has many dependent variables, HRT SD was chosen because it had the highest frequency of identified difficulties which also showing good reliability. McNemar's test was used to test for significant differences in identification of difficulties between dependent variables. The frequencies of at least mild difficulties on each performance-based measure can be found in Fig. 1. Significant differences emerged between the Flanker and Identification ($p = .008$), Flanker and HRT SD ($p = .004$), and Flanker and LSWM ($p = .039$), with the Flanker identifying significantly fewer difficulties in each case. Additionally, the DAS-II DF identified significantly fewer difficulties than the Identification task ($p = .031$) and HRT SD ($p = .039$). No significant differences emerged in frequency of difficulty identification between parent report outcome scores (Fig. 2).

Practice Effects

Paired samples *t*-tests were run to compare scores at time 1 and time 2 to examine practice effects. The *t*-statistics, significance, and Cohen's *d* values can be found in Table 6. Omissions ($p = .022$, $d = .349$) scores showed a significant practice effect (performance improving over time) with a small effect. Conners Inattention/Hyperactivity ($p = .031$, $d = .238$) Inattention/Hyperactivity ratings worsened over time.

Table 5 Spearman correlations of performance-based attention and executive functioning measures with parent ratings, vocabulary, and cognitive functioning

	Inhibit	Shift	Emotional control	Working memory	GEC	Inattention/hyperactivity	TPVT	GCA
Cogstate								
Identification	.220	.079	−.074	.218	.323	.249	.204	−.434
NIH Toolbox								
Flanker	−.256	−.097	−.007	−.279	−.328	−.355	.335	.608**
DCCS	−.209	−.209	−.332	−.387	−.445	−.440	.606*	.584*
LSWM	−.124	−.124	−.179	−.290	.027	−.419	.181	.339
K-CPT 2								
Detectability	.572*	.517*	.605**	.284	−.534*	.565*	−.216	−.561*
Omissions	.493*	.359	.503*	.433	.480*	.643**	−.227	−.351
Commissions	.191	.078	.089	.060	.100	−.063	−.075	−.253
Perseverations	.535*	.452	.561*	.157	.468*	.585**	−.319	−.441
HRT	.300	.115	.430	.247	.320	.235	−.184	−.096
HRT SD	.551*	.369	.484*	.137	.367	.620**	−.016	−.442
Variability	.634**	.230	.242	.025	.320	.578*	−.067	−.415
HRT BC	−.091	.068	−.147	.047	.004	−.384	−.197	.164
HRT ISI	.285	.422	.499*	.064	.330	−.574*	−.137	−.085
DAS-II DF	−.375	−.186	−.375	−.320	−.332	−.279	.144	.424

* $p < .05$, ** $p < .01$

Values are bolded to distinguish relations that were at least moderately correlated. *GEC*, Global Executive Composite; *TPVT*, Toolbox Picture Vocabulary Test; *GCA*, Global Cognitive Ability; *DCCS*, Dimensional Change Card Sort; *LSWM*, List Sort Working Memory; *K-CPT*, Conners Kiddie Continuous Performance Test Second Edition; *HRT*, Hit Reaction Time; *HRT BC*, Hit Reaction Time Block Change; *HRT ISI*, Hit Reaction Time Inter Stimulus Interval; *DAS-II DF*, Differential Ability Scales-Second Edition Digits Forward Task.

Discussion

Although it has been demonstrated that young children with NF1 have attention and executive difficulties (Casnar & Klein-Tasman, 2017; Templer et al., 2013), the psychometric

properties of the tools used to measure these domains have not been established with this population. These data are necessary given the high rate of attention problems in the NF1

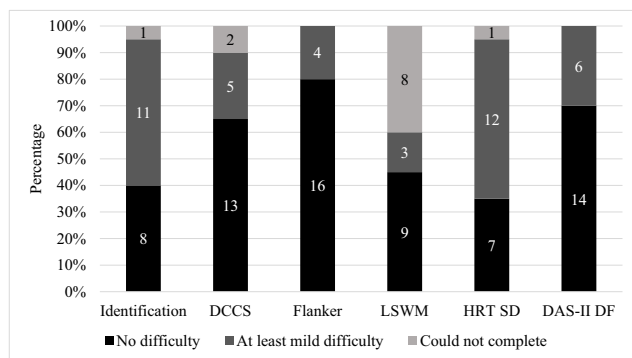


Fig. 1 Percentage and frequency of difficulty and completion across performance-based measures. Note: Difficulty was defined as either (1) performance on the task at least 1 standard deviation from the mean or (2) inability to complete the task. DCCS, Dimensional Change Card Sort; LSWM, List Sort Working Memory; K-CPT, Conners Kiddie Continuous Performance Test Second Edition; HRT, Hit Reaction Time; DAS-II DF, Differential Ability Scales-Second Edition Digits Forward Task

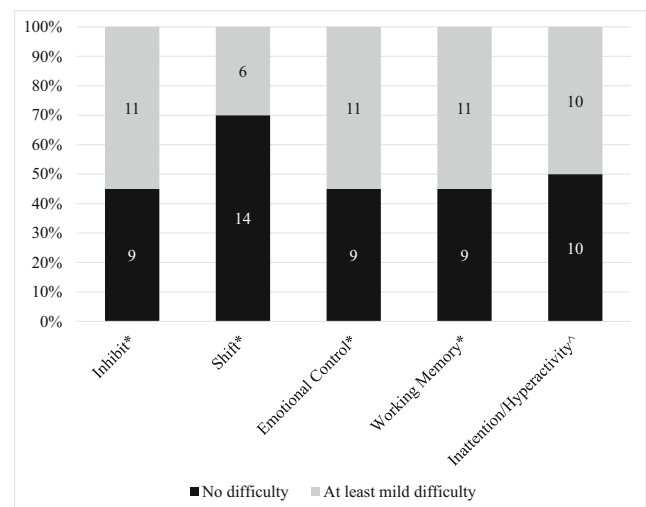


Fig. 2 Percentage and frequency of difficulty across parent-report measures. Note: Difficulty was defined as either (1) performance on the task at least 1 standard deviation from the mean. *Scale from the Behavior Rating Inventory of Executive Function Preschool or Second Edition Parent Report. ^Scale from the Conners Early Childhood Parent Report Form

Table 6 Paired samples *t*-tests between scores at time 1 and time 2

Measure	<i>n</i>	Time 1 Mean(SD)	Time 2 Mean(SD)	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Cogstate							
Identification*	17	62.17(10.22)	61.32(12.94)	0.29	16	.770	
NIH Toolbox							
Flanker [^]	18	89.67(12.85)	92.44(19.95)	−.795	17	.438	
DCCS [^]	15	91.60(10.23)	94.93(12.84)	−.813	14	.430	
LSWM [^]	11	91.91(10.72)	98.64(8.11)	−2.04	10	.068	
K-CPT 2							
Detectability*	16	62.31(8.08)	61.69(7.64)	.361	15	.723	
Omissions*	16	71.13(16.47)	65.50(15.74)	2.56	15	.022	.349
Commissions*	16	53.88(11.73)	55.56(8.61)	−.648	15	.527	
Perseverations*	16	60.69(14.85)	61.88(15.01)	−.297	15	.770	
HRT*	16	62.94(10.90)	64.13(14.49)	−.412	15	.686	
HRT SD*	16	71.25(14.81)	68.94(12.72)	.795	15	.439	
Variability*	14	65.71(13.43)	58.71(18.96)	1.31	13	.211	
HRT BC*	15	48.80(16.04)	50.02(19.57)	−.268	14	.793	
HRT ISI*	15	66.47(12.44)	66.67(14.15)	−.052	14	.959	
DAS DF*	18	45.39(9.61)	45.67(12.76)	−0.17	17	.865	
BRIEF							
Inhibit*	19	60.84(13.76)	57.21(18.51)	1.28	18	.219	
Shift*	19	51.47(10.46)	52.05(10.86)	−0.45	18	.658	
EC*	19	60.58(14.09)	57.11(11.34)	1.86	18	.079	
WM*	19	62.05(14.11)	61.42(13.05)	0.43	18	.670	
GEC*	19	62.21(13.89)	60.74(13.17)	1.01	18	.325	
Conners I/H*	19	60.72(13.05)	64.06(14.86)	−2.35	18	.031	.238

**T*-scores, [^]*standard scores*

DCCS, Dimensional Change Card Sort; LSWM, List Sort Working Memory; K-CPT, Conners Kiddie Continuous Performance Test Second Edition; HRT, Hit Reaction Time; HRT BC, Hit Reaction Time Block Change; HRT ISI, Hit Reaction Time Inter Stimulus Interval; DAS-II DF, Differential Ability Scales-Second Edition Digits Forward Task; BRIEF, Behavior Rating Inventory of Executive Function; GEC, Global Executive Composite; I/H, Inattention/Hyperactivity Scale

population (Beaussart et al., 2018; Casnar & Klein-Tasman, 2017; Plasschaert et al., 2016), differences in clinical features between children with NF1 and those with idiopathic ADHD (Huijbregts, 2012), and the call for these data to inform clinical trials and developmental research (Klein-Tasman et al., 2021; Walsh et al., 2016). In this study, we reported preliminary data on the reliability and validity of attention measures, including a performance-based task (DAS-II DF), several computerized measures (Cogstate, NIH Toolbox, and K-CPT 2), and two parent-report measures (BRIEF and Conners Inattention/Hyperactivity Scale) in a small sample of young children with NF1.

Summary of Preliminary Psychometric Properties

Cogstate

While the completion rate of the Cogstate Identification task (a measure of attention) was high, almost half of the sample

did not pass a validity integrity check. The Identification task demonstrated poor agreement and moderate consistency across time points. In terms of validity, more support was generally found for the Identification task. The Identification task had some associations with the other computerized measures of attention, but minimal relations with parent report of attention and executive function. Importantly, the task was more strongly related to general intellectual abilities than it was to parent-reported attention and executive behavioral concerns. Thus, when using this task, one must consider the effect that intellectual functioning has on performance. The Identification task did not yield significant practice effects.

NIH Toolbox

The NIH Toolbox DCCS, Flanker, and LSWM tasks were examined in the present study. The children in our sample were generally able to successfully complete the DCCS and without significant practice effects. However, performance on

the DCCS was quite different between time 1 and time 2 in terms of both agreement (ICC) and consistency (Pearson r). Clinicians and researchers should use this measure longitudinally with caution. There was considerable support for convergent validity of the DCCS as it was generally related to other computerized measures and parent-report. However, there was weak evidence of discriminant validity for this measure. The DCCS was more strongly related to general intellectual abilities and fund of vocabulary knowledge than it was with many of the attention and executive measures.

Our sample had a high completion rate for the Flanker and without significant practice effects. This measure demonstrated acceptable reliability in terms of both consistency (Pearson r) and agreement (ICC). Although the Flanker demonstrated evidence for convergent validity with other computerized measures, it had minimal relations with parent-reported attention, and the pattern of associations for the Flanker indicated that this task was highly related to general intellectual abilities, more so than to measures of attention.

Many of the young children in the present sample had difficulty with the LSWM task, as evidenced by the low completion rate. On this task, children had to first pass practice trials in which they order animals based on their size. Many children in our sample were unable to do so, and thus no data from this task were generated for almost half of the participants. The LSWM task had low agreement (ICC) and moderate consistency (Pearson r) between time 1 and time 2 scores, but these findings should be interpreted with caution, given the low completion rate. The LSWM task was related to other measures in the present study, though it was unrelated to most parent-reported attention abilities. Given that it is a working memory measure, it is not surprising that the associations were not as high as the attention measures. There was mixed evidence of discriminant validity of the LSWM, as evidenced by the low associations with vocabulary (TPVT), but not general intellectual abilities (GCA). Finally, the LSWM did not demonstrate practice effects.

Out of the NIH Toolbox measures, the Flanker demonstrated the highest agreement and consistency between scores at time 1 and time 2. In terms of validity, all of the NIH Toolbox tasks had relations with other measures of attention and, thus, have some support for convergent validity. However, both the Flanker and the DCCS had patterns of associations that were stronger with measures of intellectual and vocabulary ability than with attention or executive ability. The LSWM had stronger evidence than the Flanker and DCCS for discriminant validity. None of the NIH Toolbox tasks showed practice effects.

K-CPT 2

Similar to the Cogstate, although a large portion of our sample was able to successfully complete the K-CPT 2, about 40% of

participants did not pass the validity check. The outcome measures of the K-CPT 2 yielded a wide range of test-retest interpretations. Omissions had the highest agreement (as indicated by ICC values) between time 1 and time 2 scores and was the only score that was in the good-to-excellent range across all measures. In terms of consistency (as indicated by Pearson r), all scores except Variability demonstrated moderate-to-strong reliability. There was considerable support for convergent validity. Firstly, there were several correlations between each score and the other computerized measures. Secondly, many of the scores were also at least moderately related to most parent-reported attention and executive symptoms, with the exception of Commissions, Variability, and HRT BC. Support for discriminant validity was somewhat mixed, as Commissions, HRT SD, and Variability each had stronger correlations with measures of intellectual ability than with parent-reported attention symptoms. Analyses of practice effects indicated that overall, the K-CPT 2 yield practice effects only for Omissions. Additionally, Variability was significantly related to age at time 1, but not at time 2. This may suggest that practice does indeed play a role in Variability scores. Thus, researchers and clinicians are advised to interpret decreases in Omissions over time in children with NF1 with caution.

Overall, given its high reliability and consistent relations with other measures of attention, Omissions emerged as a strong preliminary metric of attention difficulties in young children with NF1, as long as practice effects are controlled for. Indeed, an avenue for future research is to include a control group so that it is possible compare improvements in Omissions across time points to a group of unaffected children to investigate whether the improvements are in excess of what would be expected based on practice alone. Future research should investigate whether practice effects are present at longer test-retest intervals as well.

DAS-II Digits Forward

In the present study, the DAS-II DF showed good test re-test reliability using both ICC and Pearson r values and did not demonstrate significant practice effects. The task largely showed evidence for discriminant validity, as evidenced by the very low relation with vocabulary knowledge and moderate relation with intellectual abilities. There was generally support for convergent validity of the DF task, particularly as it pertains to its relations with parent reports of attention difficulties. Interestingly, this task was moderately related to Commissions, and weakly related to Omissions. This pattern is unexpected given previous literature that showed that elevations on Omission errors are more common than on Commission errors in NF1 (Heimgärtner et al., 2019; Arnold et al., 2018). This may suggest that this DF task is less sensitive to sustained attention and more sensitive to impulsivity

difficulties in NF1, which is a less frequently reported difficulty. This may also help explain why this DF task identified fewer difficulties than the Identification task and HRT SD on the K-CPT 2; it may be tapping into a facet of attention that is less impaired than the other measures. Alternatively, the fewer evident difficulties on this task may be due to administration. While on a computerized measure, a child is tasked with maintaining their attention for an extended time, typically with minimal prompting. On a traditional digits forward task, children are interacting with a test administrator, which may be more engaging and offers more opportunities for behavioral management techniques to finish the task.

BRIEF

With the exception of Inhibit, which was in the moderate range, each BRIEF score's ICC value was in the good-to-excellent range of test re-test reliability. In terms of Pearson r , all scores demonstrated good consistency. No subscales had a significant practice effect.

Conners Inattention/Hyperactivity

The Conners Inattention/Hyperactivity scale had the highest test re-test reliability out of all of the measures in the present study. However, this scale did demonstrate a significant practice effect, with scores tending to worsen over time. Thus, we caution interpretation of changes (or lack thereof) in this scale across a similar timeframe when working with young children with NF1. Given the increase in impairment over time, improvements as a result of intervention may not be captured unless the amount of change in ratings is compared with a control group.

Implications

Given that the measures investigated demonstrated varying degrees of reliability and validity, there may not be a one-size-fits-all measure for use with this population. Clinicians and researchers must be cautious in their selection of measures and interpretation of data when using these measures with young children with NF1. When prioritizing test re-test reliability, such as in the case of longitudinal research, the performance-based indices with the highest agreement are Omissions and the Flanker and would thus be appropriate measures for use with young children with NF1. However, parent-report measures are largely more reliable, particularly the Conners Early Childhood Inattention/Hyperactivity scale.

There was generally support for validity across the measures, though Commissions was mostly unrelated to the other computerized measures and parent report measures. Importantly, many of these measures demonstrated stronger associations with cognitive functioning than other attention or

executive measures, especially the DCCS and Commissions. However, the statistical significance of these differences was not tested due to the small size of the present sample. Upon considering evidence of convergent and discriminant validity, Detectability seems to be strongly related to attention in our sample.

It is also important to consider and reflect on the high proportion of participants who were either unable to complete the tasks or did not pass validity checks. Typically, this would indicate that the performance on a task is uninterpretable; however, it may be the case that the validity check in and of itself is clinically relevant and related to the high estimates of attention deficits in this population (Hyman et al., 2005; Templer et al., 2013). Clinicians and researchers should be aware of the high rates of young children with NF1 not passing validity checks, and not necessarily discount performance when an integrity check is not met. Future research with a larger sample should investigate whether young children with NF1 who do not pass validity indicators have higher rates of attention deficits than those who do pass.

Characterization of Difficulties

There was evidence that children with NF1 are vulnerable to difficulties across many of the measures related to attention and executive functioning included here. The mean performance of the sample on Identification, Detectability, Perseverations, HRT, Variability, and HRT ISI were one standard deviation above the normative mean. This would indicate difficulty discriminating between targets and non-targets, responding slowly and inconsistently. The mean performance of our sample suggested that the participants were inattentive and lacked vigilance on the K-CPT 2. This is consistent with previous reports of the performance of young children on the First Edition of the K-CPT (Arnold et al., 2018; Sangster et al., 2011) and another continuous performance task (Heimgärtner et al., 2019). Furthermore, mean performance on Omissions and HRT SD was two standard deviations above the normative mean, further emphasizing the sample's difficulties with inattention and inconsistent performance throughout testing. Commissions, which can be an indicator of impulsivity (Halperin et al., 1991), on the other hand, was within the average range for the sample. This general profile of difficulty sustaining attention, but minimal difficulty with impulsivity is consistent with previous findings using both performance-based and parent-report measures of attention difficulties (Arnold et al., 2018; Payne et al., 2012; Sangster et al., 2011). Thus, the present performance-based findings provide further support for inattention being a central difficulty for young children with NF1.

Fewer difficulties were evident on the NIH Toolbox measures, the Shift scale, and the DAS-II DF task, with mean performances having been in the average range. The Flanker

is a measure of executive attention, which largely overlaps with executive function (Zelazo et al., 2013). Performance within the typical range would suggest that, on average, our sample demonstrated age-appropriate cognitive control. Performance on the DCCS provides further support for age-typical executive abilities, as it is thought to measure cognitive flexibility (Zelazo et al., 2013). Average performance on the DAS-II DF task is consistent with a previous study using the same task with a different sample of young children with NF1 (Casnar & Klein-Tasman, 2017). Thus, overall performance suggests that executive difficulty may be less evident in young children with NF1. Further research about the timeline of the emergence of executive challenges on assessment measures is warranted.

Although mean performance on the LSWM task was in the average range for those who completed this task, it is important to recognize that almost half of the sample was not able to complete the task because they did not pass the practice trials. In the practice trials, the participants are asked to say the animals on the screen in size order, and then practice repeating them in size order without the stimuli on the screen. If they are unable to do so, the task discontinues. Understanding size and order are fundamental math and relational vocabulary concepts. Since the rates of learning disabilities are high in the NF1 population (Hyman et al., 2005), this task may not have been developmentally appropriate for the young children in the sample. Additionally, the low rate of completion could be due to working memory being a core deficit in NF1 (Templer et al., 2013). It could be the case that the LSWM demanded too much of a working memory load for the young children in this sample, even on the practice trials. Thus, it may be the case that the children in our sample who were able to complete the task have less cognitive difficulties than those who were unable to and are hence inflating the mean performance score. In any case, the reasons for difficulty with completing the LSWM are likely heterogeneous.

Limitations and Future Directions

This study is not without limitations. Firstly, this pilot study is underpowered and limited by a small sample. However, there are currently no available psychometric data to help inform research and clinical practice when working with young children with NF1. This study also did not include a control group of unaffected children as comparison, though normative data do exist for typically developing children. Using normative data is helpful as it offers large, stratified samples to match that of the most recent census. Most notably, the testing conditions, including the length of the battery, likely vary substantially from normative data collection procedures. Thus, our sample likely had a longer study visit with many more measures than the normative sample, which could impact data in the form of fatigue. Our sample is also largely white, which

may limit the generalizability of our findings. Future research should expand upon the present study to include a more nationally representative, larger sample of children with NF1. Another avenue for future research would be to investigate the role of persistence, motivation, and effort in the completion of these tasks in young children with NF1. More generally, there is a need for additional psychometric research in a broader age range with the NF1 population. Many of the measures in the present study also provide normative data for older children and into adulthood. The reliability and validity of these measures may change with age, especially since executive dysfunction tends to worsen with age in NF1 (Beaussart et al., 2018). Furthermore, there are different measures of attention and executive abilities that are used with older children and adolescents, such as the Conners Continuous Performance Test, Third Edition (CPT-3; Conners, 2008) whose psychometric properties should be investigated in older children with NF1. Given how prevalent attention and executive difficulties are in this population, it is vital that this line of research continues to ensure the appropriate tools are being used to measure these difficulties across development in NF1.

Conclusions

Children with NF1 can exhibit difficulties with attention and executive function from a young age (Casnar & Klein-Tasman, 2017), which can differ from those seen in ADHD, yet the validity and reliability of attention measures in this population were previously unknown. These preliminary psychometric findings shed some light on which measures may be most effective at capturing these challenges. Yet, there may not be a one-size-fits-all measure of attention for use with young children with NF1, though the present analyses may lend insight into best practices for creating clinical or developmental study batteries for young children with NF1. When choosing a measure to use in a clinical or research setting, it is important to consider what the goal of the assessment is, whether to prioritize test-retest reliability and practice effects, and whether it is more important to choose a measure that has considerable support for validity. In general, the BRIEF and K-CPT 2 emerged as strong measures for use with young children with NF1, particularly because they offer a variety of scores that tended to be both reliable and demonstrated evidence of validity. However, Omissions on the K-CPT 2 and Emotional Control on the BRIEF may have practice effects, and should thus be used with caution, especially in clinical research. Additionally, our findings confirm previous work that has shown inattention to be a central concern for young children with NF1. Thus, it is particularly imperative that professionals use appropriate, reliable, and valid tools to evaluate these difficulties when assessing inattention in this population.

Acknowledgements The authors would like to thank the participants and their families for participating in this research. This work was supported by grants from NF Midwest. The authors would like to thank James Tonsgard, Dawn Siegel, Donald Basel, and Robert Listerick and the many wonderful genetic counsellors who work with them for referring participants with NF1 for this research. We also acknowledge the Child Neurodevelopment Research Lab team for their work and collaboration in the laboratory. Cogstate licensed materials were used in the present study, which were developed by and licensed from Cogstate.

Author Contribution Bonita P. Klein-Tasman and Kristin M. Lee contributed to the study conception and design. Material preparation and data collection were performed by all authors. Data analyses were performed by and the first draft of the manuscript was written by Sara K. Pardej, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was supported by a grant from NF Midwest.

Availability of Data and Material Not applicable.

Code Availability Not applicable.

Declarations

Ethics Approval This study was approved by the University of Wisconsin – Milwaukee’s institutional review board: IRB #17.326.

Consent to Participate Families were mailed consent documents prior to the testing appointment. At the testing session, families were given the opportunities to ask any questions before agreeing to participate. All children provided their written and verbal assent.

Consent for Publication All participants consented to having their data published in a journal article.

Conflict of Interest The authors declare no competing interests.

References

- Arnold, S. S., Payne, J. M., Lorenzo, J., North, K. N., & Barton, B. (2018). Preliteracy impairments in children with neurofibromatosis type 1. *Developmental Medicine and Child Neurology*, *60*(7), 703–710. <https://doi.org/10.1111/dmnc.13768>
- Barnard, H., Rao, R., Xu, Y., Froehlich, T., Epstein, J., Lanphear, B. P., & Yolton, K. (2018). Association of the Conners’ Kiddie Continuous Performance Test (K-CPT) Performance and parent-report measures of behavior and executive functioning. *Journal of Attention Disorders*, *22*(11), 1056–1065. <https://doi.org/10.1177/1087054715578271>
- Baron, I. S., Erickson, K., Ahronovich, M. D., Baker, R., & Litman, F. R. (2011). Cognitive deficit in preschoolers born late-preterm. *Early Human Development*, *87*(2), 115–119. <https://doi.org/10.1016/j.earlhumdev.2010.11.010>
- Baudou, E., Nemmi, F., Biotteau, M., Maziero, S., Peran, P., & Chaix, Y. (2020). Can the cognitive phenotype in neurofibromatosis type 1 (NF1) be explained by neuroimaging? A review. *Frontiers in Neurology*, *10*. <https://www.frontiersin.org/article/10.3389/fneur.2019.01373>
- Beaussart, M., Barbarot, S., Mauger, C., & Roy, A. (2018). Systematic review and meta-analysis of executive functions in preschool and school-age children with neurofibromatosis type 1. *Journal of the International Neuropsychological Society*, *24*(9), 977–994. <https://doi.org/10.1017/S1355617718000383>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bishop, S. L., Guthrie, W., Coffing, M., & Lord, C. (2011). Convergent validity of the Mullen scales of early learning and the differential ability scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities*, *116*(5), 331–343. <https://doi.org/10.1352/1944-7558-116.5.331>
- Casnar, C. L., & Klein-Tasman, B. P. (2017). Parent and teacher perspectives on emerging executive functioning in preschoolers with neurofibromatosis type 1: Comparison to unaffected children and lab-based measures. *Journal of Pediatric Psychology*, *42*(2), 198–207. <https://doi.org/10.1093/jpepsy/jsw042>
- Cogstate. (2018). *Cogstate pediatric and adult normative data*. Cogstate Ltd..
- Conners, C. K. (2008). *Conners (3rd Edition)*.
- Conners, C. K. (2009). *Conners early childhood: Manual* (3rd ed.). Multi-Health Systems Incorporated.
- Conners, C. K. (2015). *Conners Kiddie Continuous Performance Test 2nd Edition (Conners K-CPT 2)*. Multi-Health Systems Incorporated.
- Elliot, C. D. (2007). *Differential ability scales—Second Edition: Administration and scoring manual* (2nd ed.). Harcourt Assessment, Incorporated.
- Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., & Weintraub, S. (2015). Language measures of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society*, *20*(6), 642–651. <https://doi.org/10.1017/S1355617714000411>
- Gilboa, Y., Josman, N., Fattal-Valevski, A., Toledano-Alhadeef, H., & Rosenblum, S. (2014). Underlying mechanisms of writing difficulties among children with neurofibromatosis type 1. *Research in Developmental Disabilities*, *35*(6), 1310–1316. <https://doi.org/10.1016/j.ridd.2014.03.021>
- Gillentine, M. A., Berry, L. N., Goin-Kochel, R. P., Ali, M. A., Ge, J., Guffey, D., Rosenfeld, J. A., Hannig, V., Bader, P., Proud, M., Shinawi, M., Graham, B. H., Lin, A., Lalani, S. R., Reynolds, J., Chen, M., Grebe, T., Minard, C. G., Stankiewicz, P., & Beaudet, A. L. (2017). The cognitive and behavioral phenotypes of individuals with CHRNA7 duplications. *Journal of Autism and Developmental Disorders*, *47*(3), 549–562. <https://doi.org/10.1007/s10803-016-2961-8>
- Gioia, G. A., Espy, K. A., & Isquith, P. K. (2003). *Behavior rating inventory of executive function—Preschool version: Professional manual*. Psychological Assessment Resources, Incorporated.
- Gioia, G. A., Espy, K. A., & Isquith, P. K. (2015). *BREIF (2nd Edition)*.
- Halperin, J. M., Wolf, L., Greenblatt, E. R., & Young, G. (1991). Subtype analysis of commission errors on the continuous performance test in children. *Developmental Neuropsychology*, *7*(2), 207–217. <https://doi.org/10.1080/87565649109540488>
- Heimgärtner, M., Granström, S., Haas-Lude, K., Learch, R. A., Mautner, V.-F., & Lidzba, K. (2019). Attention deficit predicts intellectual functioning in children with neurofibromatosis type 1. *International Journal of Pediatrics*, *2019*, 1–10. <https://doi.org/10.1155/2019/9493837>
- Huijbregts, S. (2012). Cognitive-behavioral phenotype or comorbid disorder? The case of attention-deficit-hyperactivity disorder in neurofibromatosis type 1. *Developmental Medicine & Child Neurology*,

- 54(10), 873–874. <https://doi.org/10.1111/j.1469-8749.2012.04396.x>
- Hyman, S. L., Shores, A., & North, K. N. (2005). The nature and frequency of cognitive deficits in children with neurofibromatosis type 1. *Neurology*, 65(7), 1037–1044. <https://doi.org/10.1212/01.wnl.0000179303.72345.ce>
- Isenberg, J. C., Templer, A., Gao, F., Titus, J. B., & Gutmann, D. H. (2013). Attention skills in children with neurofibromatosis type 1. *Journal of Child Neurology*, 28(1), 45–49. <https://doi.org/10.1177/0883073812439435>
- Kayl, A. E., Moore, B. D., Slopis, J. M., Jackson, E. F., & Leeds, N. E. (2000). Quantitative morphology of the corpus callosum in children with neurofibromatosis and attention-deficit hyperactivity disorder. *Journal of Child Neurology*, 15(2), 90–96. <https://doi.org/10.1177/088307380001500206>
- Klein-Tasman, B. P., Lee, K., Thompson, H. L., Janusz, J., Payne, J. M., Pardej, S. K., de Blank, P., Kennedy, T., Janke, K. M., del Castillo, A., & Walsh, K. S. (2021). Recommendations for measurement of attention outcomes in preschoolers with neurofibromatosis. *Neurology*, 97(7 Supplement 1), S81–S90. <https://doi.org/10.1212/WNL.00000000000012423>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lehtonen, A., Howie, E., Trump, D., & Huson, S. M. (2013). Behaviour in children with neurofibromatosis type 1: Cognition, executive function, attention, emotion, and social competence. *Developmental Medicine and Child Neurology*, 55(2), 111–125. <https://doi.org/10.1111/j.1469-8749.2012.04399.x>
- Lehtonen, A., Garg, S., Roberts, S. A., Trump, D., Evans, D. G., Green, J., & Huson, S. M. (2015). Cognition in children with neurofibromatosis type 1: Data from a population-based study. *Developmental Medicine and Child Neurology*, 57(7), 645–651. <https://doi.org/10.1111/dmcn.12734>
- Lidzba, K., Granström, S., Lindenau, J., & Mautner, V. (2012). The adverse influence of attention-deficit disorder with or without hyperactivity on cognition in neurofibromatosis type 1. *Developmental Medicine and Child Neurology*, 54(10), 892–897. <https://doi.org/10.1111/j.1469-8749.2012.04377.x>
- Lo-Castro, A., D'Agati, E., & Curatolo, P. (2010). ADHD and genetic syndromes. *Brain & Development*, 33(6), P456–P461. <https://doi.org/10.1016/j.braindev.2010.05.011>
- Mahone, E. M. (2005). Measurement of attention and related functions in the preschool child. *Mental Retardation and Developmental Disabilities Research Reviews*, 11(3), 216–225. <https://doi.org/10.1002/mrdd.20070>
- Mahone, E. M., & Schneider, H. E. (2012). Assessment of attention in preschoolers. *Neuropsychology Review*, 22(4), 361–383. <https://doi.org/10.1007/s11065-012-9217-y>
- Moore, B. D., Slopis, J. M., Schomer, D., Jackson, E. F., & Levy, B. M. (1996). Neuropsychological significance of areas of high signal intensity on brain MRIs of children with neurofibromatosis. *Neurology*, 46(6), 1660–1668. <https://doi.org/10.1212/WNL.46.6.1660>
- North, K., Hyman, S., & Barton, B. (2002). Cognitive deficits in neurofibromatosis 1. *Journal of Child Neurology*, 17(8), 605–612. <https://doi.org/10.1177/088307380201700811>
- Payne, J. M., Hyman, S. L., Shores, E. A., & North, K. N. (2011). Assessment of executive function and attention in children with Neurofibromatosis type 1: Relationships between cognitive measures and real-world behavior. *Child Neuropsychology*, 17(4), 313–329. <https://doi.org/10.1080/09297049.2010.542746>
- Payne, J. M., Arnold, S. S., Pride, N. A., & North, K. (2012). Does attention-deficit-hyperactivity disorder exacerbate executive dysfunction in children with Neurofibromatosis type 1? *Developmental Medicine and Child Neurology*, 54(10), 898–904. <https://doi.org/10.1111/j.1469-8749.2012.04357>
- Plasschaert, E., Eylen, L. V., Descheemaeker, M. J., Noens, I., Legius, E., & Steyaert, J. (2016). Executive functioning deficits in children with neurofibromatosis type 1: The influence of intellectual and social functioning. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 171B(3), 348–362. <https://doi.org/10.1002/ajmg.b.32414>
- Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., & Rohde, L. A. (2014). ADHD prevalence estimates across three decades: An updated systematic review and meta-regression analysis. *International Journal of Epidemiology*, 43(2), 434–442. <https://doi.org/10.1093/ije/dyt261>
- Sangster, J., Shores, E. A., Watt, S., & North, K. N. (2011). The cognitive profile of preschool-aged children with neurofibromatosis type 1. *Child Neuropsychology*, 17(1), 1–16. <https://doi.org/10.1080/09297041003761993>
- Templer, A. K., Titus, J. B., & Gutmann, D. H. (2013). A neuropsychological perspective on attention problems in neurofibromatosis type 1. *Journal of Attention Disorders*, 17(6), 489–496. <https://doi.org/10.1177/1087054711433422>
- Tulsky, D. S., Carlozzi, N., Chiaravalloti, N. D., Beaumont, J. L., Kisala, P. A., Mungas, D., Conway, K., & Gershon, R. (2015). NIH Toolbox Cognition Battery (NIHTB-CB): The list sorting test to measure working memory. *Journal of the International Neuropsychological Society*, 20(6), 599–610. <https://doi.org/10.1017/S135561771400040X>
- Walsh, K. S., Janusz, J., Wolters, P. L., Martin, S., Klein-Tasman, B. P., Toledo-Tamula, M. A., Thompson, H. L., Payne, J. M., Hardy, K. K., de Blank, P., Semerjian, C., Gray, L. S., Solomon, S. E., & Ullrich, N. (2016). Neurocognitive outcomes in neurofibromatosis clinical trials: Recommendations for the domain of attention. *American Academy of Neurology*, 87(7 Supplement 1), S21–S30. <https://doi.org/10.1212/WNL.0000000000002928>
- Washbrook, E., Propper, C., & Saval, K. (2013). Pre-school hyperactivity/attention problems and educational outcomes in adolescence: Prospective longitudinal study. *The British Journal of Psychiatry*, 203(3), 265–271. <https://doi.org/10.1192/bjp.bp.112.123562>
- Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, 78(4), 16–33. <https://doi.org/10.1111/mono.12032>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.