August 2020

A Psychometric Evaluation of Computerized Attention Measures in Young Children with Neurofibromatosis Type 1

Sara Katharine Pardej
University of Wisconsin-Milwaukee

A PSYCHOMETRIC EVALUATION OF COMPUTERIZED ATTENTION MEASURES IN YOUNG CHILDREN WITH NEUROFIBROMATOSIS TYPE 1

by

Sara K. Pardej

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Psychology

at

University of Wisconsin – Milwaukee

August 2020

ABSTRACT

A PSYCHOMETRIC EVALUATION OF COMPUTERIZED ATTENTION MEASURES IN YOUNG CHILDREN WITH NEUROFIBROMATOSIS TYPE 1

by

Sara Pardej

The University of Wisconsin—Milwaukee, 2020
Under the Supervision of Professor Bonita P. Klein-Tasman

Children with neurofibromatosis type 1 (NF1) often demonstrate difficulties with attention and executive functioning that can be evident even starting at a young age. Despite this consistent finding in the literature, there has been no research to determine which measures of attention are most suitable for use with children with NF1. Recently, there have been several computerized measures of attention and executive abilities that have become available to researchers and clinicians. This study explored the National Institute of Health Toolbox Flanker, Dimensional Change Card Sort, and List Sort Working Memory; Cogstate Identification; and Conners Kiddie Continuous Performance Test, second edition to conclude which are the most reliable, valid, and identify the most difficulty in this population. Participants (ages 4-6 years; *M*=5.45, *SD*=0.75) 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. The measures demonstrated a variety of patterns of strengths and weaknesses, and there may not be a "one size fits all" measure for use with young children with NF1. Specific recommendations are provided for the appropriate measure to use in clinical and research batteries.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work was supported by a grant from NFMidwest to Bonita P. Klein-Tasman. I would like to thank the participants and their families for taking the time to take part in this research. I also acknowledge the Child Neurodevelopment Research Lab team for their work and collaboration in the laboratory, especially Danielle Glad and Kristin Lee.

Introduction

While NF1 is related to 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. Neurofibromatosis type 1 (NF1) affects 1 in 3,500 births and is the most common single-gene autosomal dominant disorder (Huson & Hughes, 1994). Attention Deficit/Hyperactivity Disorder (ADHD) and attentional difficulties more broadly are prevalent across many genetic syndromes, including NF1 (Lo-Castro et al., 2010) Although 33-50% of children with NF1 meet criteria for ADHD, there are even more children with NF1 who demonstrate difficulties with attention and executive function (Casnar & Klein-Tasman, 2017; Isenberg et al., 2013; North et al., 2002; Plasschaert et al., 2016). While attention and executive difficulties are characterized in 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.

Medical and behavioral phenotype of NF1

Although the symptomatology varies among individuals with NF1, it characterized by a mutation on the NF1 gene (National Institute of Health, 2019). This gene contains the genetic code for the production of neurofibromin, a protein that acts as a tumor suppressor in both the central and peripheral nervous systems. The resulting phenotype consists of neurofibromas, or tumors, throughout the nervous system, particularly underneath the skin. Although these are typically noncancerous, individuals with NF1 are at an increased risk of developing malignancies over time. To obtain a diagnosis, one must have two or more of the following symptoms: 6 or more café au lait spots (at least 5mm), two or more neurofibromas or one plexiform neurofibroma, freckling, 2 or more Lisch nodules, optic glioma, skeletal abnormalities (scoliosis, temple of the skull, or tibia), or a first degree relative with NF1 (Friedman, 1993). Diagnoses are commonly confirmed using genetic testing. In addition to the medical symptomatology associated with NF1, there is a growing body of evidence demonstrating academic and functional impairments in older children, adolescents, and adults with NF1. These difficulties include increased rates of learning disabilities(North et al., 1994), social difficulties (Barton & North, 2004; S. C. J. Huijbregts & de Sonneville, 2011), internalizing emotional difficulties (Descheemaeker et al., 2005; Dilts et al., 1996), and particularly attention and executive problems (Casnar & Klein-Tasman, 2017; Isenberg et al., 2013; North et al., 2002; Payne et al., 2012; Plasschaert et al., 2016).

Attention and executive difficulties in children with NF1

Attention and executive difficulties have been described as a core deficit in children with NF1 (Templer et al., 2013) with significant difficulties with inhibition, sustained, selective, and focused attention (North et al., 2002). 33-50% of children with NF1 meet the Diagnostic and Statistical Manual's (DSM) criteria for Attention-Deficit/Hyperactivity Disorder (ADHD), and they are three times more likely to be diagnosed with ADHD than their unaffected siblings (Hyman et al., 2005; Templer et al., 2013). 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, given that 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 low intellectual functioning (Plasschaert et al., 2016). However, there have been no studies to date to identify a clear genetic overlap between ADHD and NF1 (S. Huijbregts, 2012).

There is debate in the field over whether ADHD is an inherent part of the phenotypic make-up of NF1, or whether it is a highly comorbid disorder. Studies have identified the inattentive-type to be more common in children with NF1 than both the hyperactive- and combined-type ADHD (Lidzba et al., 2012; Payne et al., 2011), though one study found combined-type ADHD to be most common (Hyman et al., 2005). Given that the inattentive-type seems to be more common, it is important to identify measures that can detect the possibly subtle difficulties that come along with this behavioral phenotype, especially in the preschool years.

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). These difficulties are prevalent in both parent report and lab-based measures, with 50% of one sample of children with NF1 rated by their caregivers in the clinically significant range on the working memory subscale of the Behavior Rating of Executive Function (BRIEF; Lehtonen et al., 2015)) Both parent and teacher ratings of attention have indicated problems with attention and emerging executive functioning in young children with NF1 (Casnar & Klein-Tasman, 2017), such that there is evidence that these difficulties are present and identifiable in multiple settings.

Although conventional paper-pencil neuropsychological measures of attention exist that have been used empirically in this population, there are new computerized tasks with more recent normative data that may be more suitable for assessing young children with NF1 and are often designed with the idea of tracking change over time or with intervention. Given the increase in novel assessment strategies and tools, there is a growing need for psychometric research to establish the most valid and reliable measures that are able to capture the attention difficulties early in development in NF1 populations. At present, there is a dearth of research identifying appropriate measures of attention in young children with neurofibromatosis type 1. This research is necessary for use in clinical trials research to help improve outcomes of children with NF1.

The necessity of psychometric research

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 children with NF1 (Walsh et al., 2016). This includes both parent report measures, performance-based paper and pencil measures, and as well as performance-based computerized measures of attention. The group noted that there is a gap in the literature concerning which measures of attention are most appropriate for use with young children with NF1. By identifying the most appropriate 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.

Identifying lab-based assessments of 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). However, it is necessary because attention problems in early childhood may lead to difficulties later on in life, such as poorer academic outcomes (Washbrook et al., 2013). Developmental studies tend to use experimental measures that do not have established psychometric properties nor do they have normative data (Mahone & Schneider, 2012). This prevents researchers from drawing valid conclusions about the true nature of the difficulties experienced by young children with NF1 both in cross-sectional and longitudinal methodologies. Computerized measures of attention and executive function can help identify these risks. 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, more updated normative data, and ease of administration. Despite these advantages, there are no studies to date that have demonstrated the psychometric properties of normed attention measures for use in young children with NF1. Research in this area would help to better determine 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 identify reliable and valid measures of attention for young children with NF1. These findings will identify which measures are appropriate to use as endpoints for clinical trials, including those studies which will characterize the developmental trajectory of NF1 and those researching treatments. To be successful, clinical trials require accurate measurement tools that have demonstrated validity, test-retest reliability, and minimal practice effects. 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 imperative that we demonstrate which measures of attention are most appropriate to use.

The utility of several computerized measures was examined: 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), and the Conners Kiddie Continuous Performance Test second edition (C. Keith Conners, 2015) tasks. These are neuropsychological tasks that measure response inhibition, shifting attention, working memory, or sustained attention. They are relatively easy to administer and have been used in several studies of attention in children with NF1 (Bluschke et al., 2017; Plasschaert et al., 2016). The current study examines whether performance on these tasks is reliable over time and consistent with other measures of attention, both with parent measures of attention and executive function and with each other. The proposed study will also investigate whether these measures differentially detect attention difficulties and 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) have 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 consented to participate. 18 children were assessed at two different time points, 8±2 weeks apart, in order to allow for test-retest reliability analyses. 2 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). 2 additional children were consented, however were unable to complete the battery due to behavioral challenges.

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), though this was not statistically significant ($\chi^2(1) = 0.80$). 95% percent of the sample was white. In terms of NF Classification, there were more sporadic (*n*=12) than familial (*n*=8) cases, but this was not statistically significant ($\chi^2(1) = 0.80$). The mean Hollingshead Index score (M=46.15(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 function to control for cognitive ability in the proposed analyses. There were 3 versions of the battery in order to allow for counterbalancing 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 lasted about 2.5 hours. All assessments took place either at the University of Wisconsin-Milwaukee's Child Neurodevelopment Laboratory or in a quiet conference room if the family was unable to drive to Milwaukee. All assessments are conducted according to each measure's standardization procedures. Parents were compensated with a $20 electronic gift card after the first appointment and a $30 electronic gift card after the second appointment. Children chose an age-appropriate children's book at the end of each appointment. Families also received a summary of their child's performance on the developmental measures.

Measures

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

The DAS-II core was administered to determine overall intellectual functioning. The measure yields an overall General Cognitive Ability (GCA) standard score (M=100, SD=15). The DAS-II demonstrates excellent reliability, validity, and standardization. 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).

Information about the normative samples and procedures for the computerized measures are detailed in Table 1.

National Institute of Health Toolbox selected subtests.

The NIH Toolbox is an electronic battery that has various measures of cognitive, emotional, sensory, motor, attention, and executive functioning. 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.

This version of the Flanker task (Zelazo et al., 2013) requires children to choose whether the middle stimulus (a fish with an arrow on it) is pointing left or right. On congruent trials, all of the stimuli are pointing in the same direction. On incongruent trials, the middle stimulus points in the opposite direction from the remaining stimuli. Administration includes practice, which is repeated three times or until criterion is met, followed by the test. If the child was accurate on ≥

80% of trials, the final score incorporates both accuracy and reaction time. Otherwise, only the accuracy score is provided.

The Dimensional Change Card Sort (DCCS) task (Zelazo et al., 2013) requires children to sort a middle stimulus either by shape or color. Sometimes the color of the middle stimulus is 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 includes practice, preswitch, postswitch, and mixed blocks. If the child was accurate on ≥ 80% of trials, the final score incorporates both accuracy and reaction time. Otherwise, only the accuracy score is provided.

The 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 are only presented with one type of stimulus (i.e., animal, fruit). If they are able to complete the initial trial to criterion, then they must repeat various series of stimuli by first saying the fruit in size order, followed by the animals in size order. Standard scores are based on the sum of the total correct responses.

The Toolbox Picture Vocabulary Test (TPVT) task (Gershon et al., 2015) is a measure of receptive vocabulary. In this task, participants are presented with four images. The iPad plays an audio recording of a word, and the participant must choose which image best depicts the word. They are 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 are 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 is $\log_{10}$ transformed reaction time, which was converted to a T-score (M=50, SD=10) for analyses.

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 is approximately 7.5 minutes and consists of 200 scored trials. T-scores (M=50, SD=10) are provided for Response Style ("C"), Detectability ("d'"), Omissions, Commissions, Perseverations, Hit Reaction Time (HRT), Variability, Hit Reaction Block Change (HRT BC), and Hit Reaction Inter Stimulus Interval (HRT ISI). Participants are 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 Short Form, Parent and Teacher versions (Conners EC; Conners, 2009).

The Conners EC is a 49-item questionnaire that was administered to caregivers. 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. 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 will use the Inhibit, Shift, Working Memory, and Global Executive Composite (GEC) scales, as those are available across both versions and yield T-scores. The GEC T-score is obtained by summing all of the scales. Each version has well-established reliability and validity.

Research Aims and Analytic Strategy

Research Aim 1

To investigate the reliability of computerized measures of attention for use with young children with NF1.

Research Aim 1 Analytic Strategy

To assess test-retest reliability, an intraclass correlation coefficients (ICC) were conducted. The ICC assesses the agreement between the two scores (Time 1 and Time 2). The ICC was assessed using standard cut-offs (Koo & Li, 2016). Thus, an ICC greater than .75 will demonstrate good to excellent test-retest reliability. Additionally, Pearson correlation coefficients from time 1 to time 2 were computed. Pearson correlations are generally used to assess how consistent scores are. Although Pearson correlations are generally not recommended for assessing test-retest reliability (Weir, 2005), since both the Cogstate and K-CPT 2 assessed reliability using Pearson *r*, it was calculated for comparison. Correlation coefficients that were at least moderately correlated ($r > 0.30$) provided evidence for test-retest reliability. It was expected that the computerized measures of attention would demonstrate excellent test-retest reliability in young children with NF1 (*Hypothesis 1*).

Research Aim 2

To examine the evidence for the validity of computerized measures of attention for use with young children with NF1.

Research Aim 2 Analytic Strategy

To test for evidence of validity, Spearman correlations were conducted between each computerized measures' outcome scores. If these scores were at least moderately correlated ($|r|>0.30$), this provided further support for the validity of the tasks for use with children with NF1. Furthermore, Spearman correlations were conducted between each computerized measure's outcome scores with the Conners Inattention/Hyperactivity scale, and BRIEF-P/BRIEF-2 Inhibit, Shift, Emotional Control, Working Memory, and GEC scales. The Inhibit scale reflects a child's ability to resist acting on impulse. Thus, this scale may be related to the computerized measures of attention that rely on impulse-control (i.e., K-CPT 2, NIH Flanker, NIH DCCS, Cogstate). The Shift scale measures a child's capacity to move from one aspect of a problem to another, thus it may be related to the NIH Flanker and NIH DCCS. The Working Memory scale assesses a child's ability to hold information in mind to finish a task. All of the computerized measures rely on paying attention and keeping the overall goal in mind, thus they may all be related to this scale. If these scores were at least moderately correlated ($|r|>0.30$), this provided support for the validity of each computerized measure for use with children with NF1 Lastly, Spearman correlations were conducted between the TPVT and DAS-II GCA with all of the computerized measures' outcome scores to investigate discriminant validity. Weak correlations ($|rho|<0.30$) provided evidence for discriminant validity. It was expected that computerized measures of attention would be at least moderately related to parent reports of attention difficulties (*Hypothesis 2a*). computerized measures of attention would be at least moderately related to each other (*Hypothesis 2b*), and computerized measures of attention would be weakly related to the TPVT and GCA (*Hypothesis 2c*).

Research Aim 3

To investigate the ability of performance on computerized tasks to identify attention difficulties in children with NF1.

Research Aim 3 Analytic Strategy

First, frequency of difficulties was examined using each dependent variable's standardized score (SS<85 or T>60). Given that the K-CPT 2 has many dependent variables, the variable that had the highest frequency of identification of difficulties was used for analyses. Children who were not able to complete a measure were coded as having difficulty. McNemar's tests was used to test for significant differences in identification of difficulties between dependent variables. It is expected that some dependent variables on these measures would identify more children with difficulties than would others (*Hypothesis 3*).

Research Aim 4

To explore the practice effects of computerized measures of attention for use with young children with NF1.

Research Aim 4 Analytic Strategy

To test for practice effects, paired samples t-tests were run to compare scores at time 1 and time 2. If time 1 and time 2 scores were not significantly different, then the measures did not demonstrate practice effects. It was expected that the computerized measures of attention would not demonstrate any practice effects (*Hypothesis 4*).

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 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; these participants were not excluded from the analyses even though they did not pass the validity indicators.

Data Analysis

The data were analyzed using IBM SPSS for Windows, version 25. Findings are interpreted using both statistical significance and effect size. A *p* value of <.05 was used to determine significance. 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.

Individual Differences

Spearman correlations were run between each outcome measure, parent report score, and age at each time point. Age at Time 1 was significantly related to K-CPT 2 Variability (*rho*=.617, *p*=.008), with older children performing significantly worse than younger children on both measures. No scores at Time 2 were significantly related to age at Time 2.

Independent samples t-tests were run to examine effects of NF classification 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 effects of NF Classification at Time 2. No significant differences emerged based on sex at either Time 1 or at Time 2. There were no significant differences between ratings on parent report measures and any demographic variables.

Test-retest Reliability of Computerized Measures of Attention (Research Aim 1)

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 demonstrated good test-retest reliability. The Flanker, Detectability, HRT, and HRT SD demonstrated moderate test-retest reliability.

Pearson correlation coefficients from time 1 to time 2 were computed for all computerized measures (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 4. Each computerized 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 of Computerized Measures of Attention (Research Aim 2)

To investigate evidence for convergent validity, Spearman correlations were conducted between each computerized measures' outcome scores. The results are summarized in Table 5. Generally, Commissions and HRT BC demonstrated weak correlations with other measures of attention in compared to the Identification, NIH Flanker, DCCS, LSWM, and the remaining K-CPT 2 outcome scores which showed stronger correlations with other measures of attention. After a False Discovery Rate (FDR) correction, the following relations remained significant: HRTSD and Variability ($q<.001$), Perseverations and Variability ($q=.001$), Perseverations and HRT SD ($q<.001$), Variability and HRT ISI ($q=.015$), Omissions and HRTSD ($q=.005$), HRTSD and HRT ISI ($q=.005$), Flanker and DCCS ($q=.005$), Omissions and HRT ($q=.005$), Perseverations and HRT ISI ($q=.005$).

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, Working Memory, and GEC. The correlation values can be found in Table 6. 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. 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. The 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. No significant correlations survived FDR correction.

Spearman correlations were conducted between the TPVT and GCA with all of the computerized measures' outcome scores to explore discriminant validity. The findings are in Table 6. 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. Many measures were highly and significantly related with GCA, including the Flanker, DCCS, and Detectability. Other measures were also moderately related (|*rho*|<0.30) to GCA, including the Identification task, LSWM, Omissions, Perseverations, Variability, and HRT SD. Generally, these measures were more highly associated with GCA than parent-reported attention and executive difficulties.

Frequency of Difficulty Identification in Computerized Measures of Attention (Research Aim 3)

The frequencies of at least mild difficulties on each measure can be found in Figure 1. HRT SD had the highest number of identified difficulties on the K-CPT 2 (*n*=13) and will therefore be used as the dependent variable in the following analyses. 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. After FDR correction, HRT SD (*q*=.04) and the Identification task (*q*=.04) still identified significantly more difficulties than the Flanker.

Practice Effects of Computerized Measures of Attention (Research Aim 4)

To test for practice effects, paired samples t-tests were run to compare scores at time 1 and time 2. The t-statistics, significance, and Cohen's d values can be found in Table 7. Only Omissions were significantly different (*p*=.022), with a small effect (*d*=.349).

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 established with this population. In this study, we investigated the reliability and validity of the Cogstate, NIH Toolbox, and K-CPT 2 in a sample of young children with NF1. Each measure had its own pattern of psychometric strengths and weaknesses; thus, researchers and clinicians should consider the goals of their assessment or study when choosing one of these measures for use with young children with NF1.

Evaluation of psychometric properties

Cogstate

In this study, we used the Identification task, which is a measure of attention. In our sample, children were generally able to complete the task, however, 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 related to general intellectual abilities than is 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 DCCS, Flanker, and LSWM were all 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 and consistency. 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, as well and parent-report. However, there was weak evidence of discriminant validity for of this measure. The DCCS was more highly related to general intellectual abilities and fund of vocabulary knowledge than it was with many attention and executive measures.

Our sample had a high completion rate for the Flanker and completed the task without significant practice effects. This measure demonstrated acceptable reliability in terms of both consistency and agreement. 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 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 some support for discriminant validity of the LSWM, as evidenced by the low associations with vocabulary, but not general intellectual abilities. 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 complete the K-CPT 2, about 40% of them 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. 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 also investigate whether practice effects are present at longer test-retest intervals as well.

Implications

Given that the various 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 children with NF1. When prioritizing test re-test reliability, such as in the case of longitudinal research, the indices with the highest agreement are Omissions and the Flanker and would thus be appropriate measures for use with young children with NF1.

There was generally support for validity across the measures, though Commissions was largely unrelated to the other computerized measures and parent report measures. Importantly, many of these measures demonstrated stronger associations with intellectual 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 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 findings provide further support for inattention being a central difficulty for young children with NF1.

Fewer difficulties were evident on the NIH Toolbox measures, with mean performance 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).

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 heterogenous.

Limitations and Future Directions

This study is not without limitations. Firstly, this study is underpowered and limited by a small sample. 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. However, there are also limitations. Most notably, the testing conditions, including the length of the battery, likely varies 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. Additionally, more psychometric research must be completed 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) that should be investigated. 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

There may not be a one-size-fits-all measure of attention for use with 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, and whether to prioritize test-retest reliability and practice effects, or whether it is more important to choose a measure that has considerable support for validity. In general, the K-CPT 2 emerged as a strong measure for use with young children with NF1, particularly because it offers a variety of scores that tended to be both reliable and demonstrated evidence of validity. However, Omissions 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 especially imperative that professionals use appropriate, reliable, and valid tools to evaluate these difficulties when assessing inattention in this population.

Table 1. *Normative data information for the Cogstate, NIH Toolbox, and K-CPT 2*

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

Table 2.

Percentage of participants that successfully completed study attention 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%

**Most data were available for 95% of participants, however because of participant response patterns, only 85% had complete data*

Table 3.

Intraclass correlation coefficient values in the present sample and published literature

Measure	Sample ICC	Published ICC
NIH Toolbox		
Flanker	.61	.92
DCCS	.06	.92
LSWM	.34	.77
Cogstate		
Identification	.49	.79
K-CPT 2		
Detectability	.61	N/A
Omissions	.85	N/A
Commissions	.49	N/A
Perseverations	.43	N/A
HRT	.59	N/A
HRT SD	.65	N/A
Variability	.27	N/A
HRT BC	.36	N/A
HRT ISI	.38	N/A

Table 4.

Pearson correlation coefficients for K-CPT 2 test-retest reliability

Measures	*r*	Published *r*
NIH Toolbox		
Flanker	**.67****	N/A
DCCS	.07	N/A
LSWM	**.36**	N/A
Cogstate		
Identification	**.49***	.62
K CPT-2		
Detectability	**.61***	.67
Omissions	**.85****	.62
Commissions	**.51***	.73
Perseverations	**.43**	.39
HRT	**.62***	.85
HRT SD	**.65****	.59
Variability	.28	.21
HRT BC	**.37**	.06
HRT ISI	**.38**	.51

Table 5.

2-tailed Spearman correlations between each outcome 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
1*	-												
2^	-.277	-											
3^	.010	**.741****	-										
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****	**-.501***	**.315**	-				
10^{+}	**.375**	**-.466***	**-.345**	-.292	**.850****	**.715****	.104	**.811****	**.455**	-			
11^{+}	**.365**	**-.583***	-.281	-.275	**.674****	**.537***	.080	**.822****	**.373**	**.867****	-		
12^{+}	**-.351**	.172	.028	**.534**	-.144	-.204	.118	-.118	-.116	**-.332**	-.198	-	
13^{+}	**.454**	**-.378**	-.292	**-.396**	**.590****	**.507***	.059	**.721****	**.371**	**.734****	**.695****	**-.418**	-

**Cogstate, ^NIH Toolbox, $^{+}$K-CPT*

Table 6.

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

Table 7.

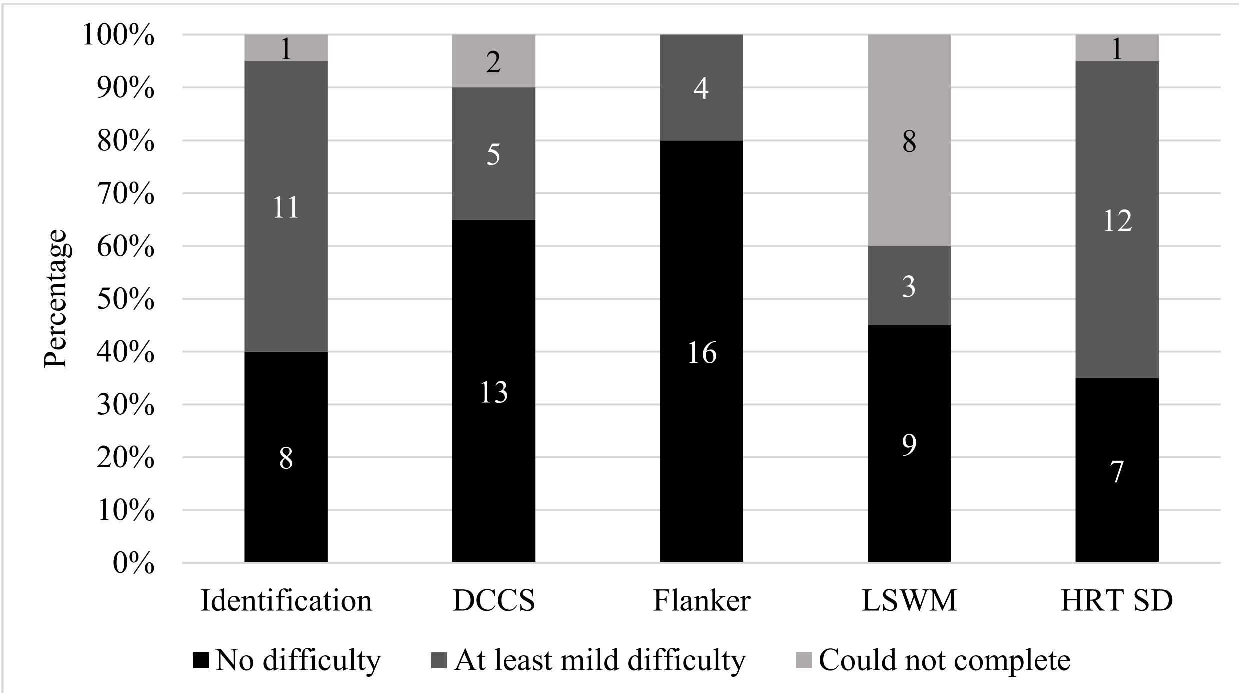
T-tests between scores at time 1 (T1) and time 2 (T2)

Measure	N	Mean(SD) T1	Mean(SD) T2	t	df	p	d
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	

**T-scores, ^standard scores*

Figure 1.

Percentage of difficulty and completion across measures



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 & Child Neurology*, *60*(7), 703–710. https://doi.org/10.1111/dmcn.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

Barton, B., & North, K. (2004). Social skills of children with neurofibromatosis type 1. *Developmental Medicine & Child Neurology*, *46*, 553–563.

Beaussart, M., Barbarot, S., Mauger, C., & Roy, A. (2018). Systematic review and meta-analysis of executive functions in preschool and school-age children with neruofibromatosis type 1. *Journal of the International Neuropsychological Society*, *24*(9), 977–994. https://doi.org/10.1017/S1355617718000383

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

Bluschke, A., Hagen, M. von der, Papenhagen, K., Roessner, V., & Beste, C. (2017). Conflict processing in juvenile patients with neurofibromatosis type 1 (NF1) and healthy controls—Two pathways to success. *NeuroImage: Clinical*, *14*, 499–505. https://doi.org/10.1016/j.nicl.2017.02.014

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. Keith. (2015). *Conners Kiddie Continuous Performance Test 2nd Edition (Conners K-CPT 2)*. Multi-Health Systems Incorporated.

Descheemaeker, M. J., Ghesquière, P., Symons, H., Fryns, J. P., & Leguis, E. (2005). Behavioral, academic and neuropsychological profile of normally gifted Neurofibromatosis type 1 children. *Journal of Intellectual Disability Research*, *49*(Pt 1), 33–46. https://doi.org/10.1111/j.1365-2788.2005.00660

Dilts, C. V., Carey, J. C., Kircher, J. C., Hoffman, R. O., Creel, D., Ward, K., Clark, E., & Leonard, C. O. (1996). Children and adolescents with Neurofibromatosis 1: A behavioral phenotype. *Journal of Developmental and Behavioral Pediatrics*, *17*(4), 229–239. https://doi.org/10.1097/00004703-199608000-00004

Elliot, C. D. (2007). *Differential ability scales—Second Edition: Administration and scoring manual* (2nd ed.). Harcourt Assessment, Incorporated.

Friedman, J. M. (1993). Neurofibromatosis 1. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. Bean, K. Stephens, & A. Amemiya (Eds.), *GeneReviews®*. University of Washington, Seattle. http://www.ncbi.nlm.nih.gov/books/NBK1109/

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-Alhadef, 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., Leark, 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

Huijbregts, S. C. J., & de Sonneville, L. M. J. (2011). Does Cognitive Impairment Explain Behavioral and Social Problems of Children with Neurofibromatosis Type 1? *Behavior Genetics*, *41*(3), 430–436. https://doi.org/10.1007/s10519-010-9430-5

Huson, S. M., & Hughes, R. A. C. (1994). *The Neurofibromatoses: A pathogenetic and clinical overview*. Chapman & Hall Medical.

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. *Jounal of Child Neurology*, *28*(1), 45–49. https://doi.org/10.1177/0883073812439435

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., 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 & 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 & 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

North, K., Hyman, S., & Barton, B. (2002). Cognitive deficits in Neurofibromatosis 1. *Jounal of Child Neurology*, *17*(8), 605–612. https://doi.org/10.1177/088307380201700811

North, K., Joy, P., Yuille, D., Cocks, N., Mobbs, E., Hutchins, P., McHugh, K., & Silva, M. de. (1994). Specific learning disability in children with neurofibromatosis type 1: Significance of MRI abnormalities. *Neurology*, *44*(5), 878–883. https://doi.org/10.1212/WNL.44.5.878

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 & Child Neurology*, *54*(10), 898–904. https://doi.org/10.1111/j.1469-8749.2012.04357

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

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

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., Blank, P. de, 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

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. [Review]. *Journal of Strength*, *19*(1), 231–240.

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