August 2020

Stability and Predictive Value of Intellectual Functioning in Neurofibromatosis Type 1 Beginning in the Preschool Years

Gregor Nathanael Pau Schwarz
University of Wisconsin-Milwaukee

STABILITY AND PREDICTIVE VALUE OF INTELLECTUAL FUNCTIONING IN NEUROFIBROMATOSIS TYPE 1 BEGINNING IN THE PRESCHOOL YEARS

by

Gregor Nathanael Schwarz

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Psychology

at

University of Wisconsin-Milwaukee

August 2020

ABSTRACT

STABILITY AND PREDICTIVE VALUE OF INTELLECTUAL FUNCTIONING IN NEUROFIBROMATOSIS TYPE 1 BEGINNING IN THE PRESCHOOL YEARS

by

Gregor Nathanael Schwarz

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

Neurofibromatosis type 1 (NF1) is a rare genetic disorder that affects multiple aspects of cognitive functioning, including intellectual functioning, attention, and executive functioning. The predictive value of intellectual functioning (IF) in the preschool years for IF in the school-age and early adolescent years has not been examined in youth with NF1. In this study, the reliability and predictive value of preschool IF for school-age IF were examined using both bivariate correlation and more complex linear mixed models. The participants were 55 youth with NF1 from ages 3 to 13 years. Intellectual functioning was measured with the Differential Ability Scales 2nd edition (DAS-II), an individually administered comprehensive measure of IF. Results indicate that both general IF and verbal functioning can be reliably measured in the preschool years in NF1 and that they hold predictive value for school-age functioning in NF1. In the bivariate correlation models, general IF in the early preschool years appeared to be a particularly strong predictor of school-age IF. Linear mixed models appeared to enhance the prediction of verbal functioning, with models including socioeconomic status (SES) and attention-deficit/hyperactivity disorder (ADHD) symptoms as predictors of IF. Nonverbal IF was generally unreliable in the preschool years and had limited predictive value, particularly once confounding variables and extreme cases were removed in the linear mixed model analysis. In

addition, based on the linear models, youth appear to increase in general and verbal IF relative to their peers during the preschool years. However, this trend does not appear to continue in the school-age years, and General Conceptual Ability (GCA), Verbal and Nonverbal scores as a group decrease into the school-age years, likely associated with increased conceptual demands in this period of development. Nuanced understanding of the predictive value of IF in the preschool years in NF1 may be helpful in the assessment of early risk predictors and treatment planning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I am incredibly grateful for the extraordinarily skillful guidance of my primary faculty advisor and dissertation committee chair Dr. Bonita Klein-Tasman. She helped me shape my research questions into a focused and achievable set of questions and analyses. She tirelessly helped me stay focused on my goal of completing the dissertation. She gently prompted me to think about the big picture and not get lost in the data-analytic weeds. She provided steady encouragement and honest, productive feedback in an extremely prompt fashion and made herself available to assist and guide beyond what would be expected. She generously corrected my grammar as a writer and speaker of English as a second language and the unfortunate, Germanic tendency to write in excessively long sentences. Additionally, she made it possible for me to write in my own voice. It was an incredible honor to work with someone with such a brilliant intellect and extensive knowledge about the topics of interest for this dissertation.

I want to thank the funding partners who made this research possible including NF Midwest, NF MidAtlantic, NF Northeast, University of Chicago CTSA grant (UL1 RR024999) and the University of Wisconsin – Milwaukee Research Growth Initiative.

My wife was for all intents and purposes a single parent while conducting this dissertation and took care of our four loud and active children 24/7 with incredible skill. In the midst of the Covid-19 pandemic she also was the sole instructor to our children while they were out of school for months and I needed time to finish my dissertation. Her encouragement was vital to the completion of the project.

I am incredibly grateful to Dr. Cristan Farmer from NIH for all her consulting on linear mixed models without expecting anything in return. There were moments when I doubted that I had a real grasp on how these models work and worried that I lacked the ability to estimate them with any skill. Dr. Farmer provided me with the critical guidance needed to gain conceptual and methodological clarity. Dr. Farmer took the time to read my entire dissertation proposal and answered endless questions. Once, she even availed herself for a same day "emergency" consultation.

There are numerous individuals at the UWM psychology department who helped me produce this work. First, I am very grateful for the guidance, thought provoking comments and challenging questions by my dissertation committee members including Drs. Krista Lisdahl, Kristin Smith, Hobart Davies and David Osmon along with my dissertation chair Dr. Bonita Klein-Tasman. There are many fellow graduate students that were helpful during the process including Kristin Lee, Danielle Glad, Brianna Yund, Cristina Casnar, Natalie Brei, Faye van der Fluit and Kelly Janke and numerous numerous research assistants. Particularly my fellow graduate students in the Child Development Research lab provided feedback on drafts of the paper, practice presentations, and helped with extensive data checking and data collection. I thank Kristin Lee for a particularly thorough early round of edits to all my tables and graphs resulting in much improved clarity, precision and consistency. Sarah Hamilton helped polish my writing by editing with extraordinary care to thoroughly understand what I was trying to express, impressive grasp for inconsistencies in at times very technical portions of the paper resulting in editing that was exquisite in its precision and exceptional attention to detail.

I must also thank the families of children with NF1 who brought in their children, filled out numerous questionnaires and for the hard work for often several hours by our participants with NF1 who were often just in the preschool age.

I am thankful for my fellow interns who provided me with steady encouragement, opportunities to practice my defense presentation. A good friend of mine also provided me with a

quiet writing sanctuary in the midst of the COVID19 pandemic. This space afforded me with critical quiet time and focus to actually write my dissertation when my own home was rarely quiet, and my university office was unavailable.

Finally, I want to thank my four children (Milo 11, Homer 7, Hazel 5, Birdie 1.5) who exhibited extraordinary patience with me as I spent countless evenings and weekends on my dissertation while on full-time pre-doctoral internship. In particular, I want to thank my oldest son. I have been pursuing a doctoral degree for his entire childhood and it has kept me away from home and family life much more than most parents.

Stability and Predictive Value of Intellectual Functioning in Neurofibromatosis Type 1 Beginning in the Preschool Years

Neurofibromatosis type 1 (NF1) is a rare genetic disorder that affects a variety of aspects of cognition, including intellectual functioning (IF). Among individuals with NF1, there is robust evidence of a moderate reduction in general-level IF in comparison to same-aged peers, although IF is in the average range on a group level. In school-age children with NF1, cognitive difficulties are associated with academic difficulties. Evidence from a moderate number of cross-sectional studies and a small number of longitudinal studies suggests that the level of IF difficulty experienced by youth with NF1 is independent of age and is similar in the preschool, school-age, and early adolescent years. The predictive value of IF in the preschool years has not been examined in relationship to later IF. A detailed understanding of the variability in the IF trajectories of children with NF1 and an examination of the predictive value of preschool IF for later cognitive difficulties is critical to developing a more nuanced understanding of the early indicators of later cognitive difficulties in individuals with NF1. This more nuanced understanding of the cognitive difficulties of youth with NF1 would potentially assist in both the assessment of early risk predictors for later IF difficulties and early intervention planning for youth with NF1.

Stability of IF in the Preschool Age and Prediction of Later IF in Typical Development

Early Preschool Intellectual Development as a Predictor of Later Preschool IF

The measured IF of younger preschool-age children predicts IF in later preschool with moderate effectiveness. There is specific evidence that some stability exists in the development of general intellectual, verbal, and nonverbal abilities between the ages of three and six years, as is indicated by moderate to large-sized cross-age correlations (.40s–.80s) of standard scores (Baker et al., 1958; Bayley, 1949; Hindley & Owen, 1978; Honzik et al., 1948; Scheider et al., 1999; Schneider & Bullock, 2010; Wilson, 1974). These moderate-sized correlations suggest that the prior year's IF accounts for 15–65% of the current year's IF. There is some evidence that a higher initial IF negatively predicts an increase in IF (Hindley & Owen, 1978). However, the high mean IQ of the sample (in the 120s) may not accurately reflect the same relationship within the broader population (Hindley & Owen, 1978). To more confidently conclude that a higher initial IF relates to reduced slopes in developmental trajectories, this finding needs to be replicated in a sample that is more representative of the general population.

Preschool Intellectual Development in Predicting IF in the School-age and Early Adolescent Years

There is a large body of research indicating that both verbal and nonverbal intelligence in the preschool years are significant predictors of intelligence in the school-age and early adolescent years. Moderate correlations between the predicted and measured values of school-age and early adolescent IF (.30–.50s) suggest that roughly 10–25% of variability in later intelligence can be accounted for by an individual's intelligence level at age 3 (Bayley, 1949; Crockett et al., 1975; Gardner & Clark, 1992; Hindley & Owen, 1978; Sameroff et al., 1993; Schneider & Bullock, 2010). The similar effect size of prediction within the preschool years compared to prediction into adolescence is indicative of somewhat increased stability of IF in the pre-adolescent and adolescent years. The possible increased stability in IF during these years may be due to the increased effects of heritability with increasing age (Davis et al., 2009; Ferrari & Sternberg, 1998). In addition, McCall and Owen (1973) found that the two clusters whose IQ performances decreased over the preschool period showed significantly lower IQs in late adolescence when compared to the two clusters whose IQ performances increased over the preschool period. As is common for studies of intellectual development, most samples have a mean in the high average to superior range, and therefore, results may not be representative of patterns in the general population. Overall, there is robust evidence that preschool intelligence levels predict intelligence levels in the school-age and adolescent years with generally moderate effectiveness and that developmental trajectories in preschool may be relevant to intellectual development in adolescence.

Evidence of Stability and Variability in Intellectual Development in Preschool-Age Children

In contrast to many long-held assumptions about the highly stable development of intelligence, and particularly those assumptions of the early 20th century (Baker et al., 1958; Bayley, 1949), multiple studies using clusters or cluster-like analyses have presented evidence of significant instability in the development of intelligence during a child's preschool years (Hindley & Owen, 1978; McCall et al., 1973). According to both Hindley and Owen (1978) and McCall, Appelbaum, and Hogarty (1973), about half of typically developing children demonstrate stable development of IF throughout the preschool period, and about one quarter of the same population makes sizeable improvements (e.g., 13+ Standard Score points, SS). The final quarter of children show decrements in IF when compared with their peers (e.g. ~10 SS, Hindley & Owen, 1978; McCall et al., 1973). Using factor analysis, McCall et al. (1973) identified five clusters of cognitive development from toddlerhood to late adolescence. Of these five clusters, one cluster showed relatively stable scores throughout development, two clusters showed significant increases, and two clusters showed significant decreases over the preschool period. The evidence from both the proportional and cluster-based analyses suggests that both significant stability and instability occur in the development of intelligence over the preschool period.

NF1: General Description and Review of Cognitive Development in NF1

Description of NF1 and General Cognitive Phenotype in NF1

NF1 is a genetic disorder caused by the mutation of a single gene on chromosome 17 (17q11.2). Occurring in about 1 in 3000 live births, NF1 manifests with physical features including café-au-lait spots, optic glioma, skin freckling, lich nodules, and cutaneous neurofibromas. Cognitive features include mildly lowered intellectual functioning, increased rates of attention-deficit/hyperactivity disorder (ADHD), executive function (EF) difficulties, and academic problems (Acosta et al., 2006; K. North et al., 1995). Neurofibromin, the product of the NF1 gene, is generally considered a tumor suppressor gene. The relations between abnormal neurofibromin and the non-tumor-related features of NF1 have yet to be clearly delineated (North, 2000). A model of cognitive deficits in NF1 by North (2000) proposes that the mutated NF1 gene produces abnormal neurofibromin, which causes aberrant cell growth and differentiation in the central nervous system, particularly during embryonic development. As a consequence, aberrant gliosis and myelination are hypothesized to manifest as T2 signal hyperintensities on magnetic resonance imaging (MRI) T2 weighed images. This abnormal myelination is assumed to disrupt higher cognitive processing in NF1. T2 hyperintensities are found in 30–70% of individuals with NF1 and are located in several areas, including the basal ganglia, the cerebellum, the brain stem, the thalamus, and the subcortical white matter (North, 2000; Payne et al., 2014). The evidence regarding the relation between T2 hyperintensities (presence, number, location) and IF (as early as preschool age) is somewhat mixed (North, 2000).

One of the most robust findings regarding IF in children with NF1 is that both older children (ages 8 and older) and preschoolers (ages 3–6) have lower levels of IF in comparison to same-aged peers (e.g., IQ ~90) at a group level of about 10 SS points below average (Coutinho et al., 2016; Cutting et al., 2002; Erdoğan-Bakar et al., 2009; Klein-Tasman et al., 2014; Lidzba et al., 2012; Lorenzo et al., 2013; Lorenzo et al., 2015; Nupan et al., 2017; Sangster et al., 2010). On a group level, youth with NF1 have similar variability in IF, which is indicated by similar standard deviations when compared to the general population (Cutting et al., 2002; Klein-Tasman et al., 2014; Lorenzo et al., 2013; Lorenzo et al., 2015; Nupan et al., 2017). Another robust finding is that, within the different commonly measured domains of intelligence, there is an absence of any strong and consistent pattern of relative strengths and weaknesses in children with NF1 (Klein-Tasman et al., 2014; Nupan et al., 2017).

Children with NF1 also show frequent attention problems. Around 40% of children with NF1 meet the diagnostic criteria for ADHD, and a large additional proportion of children experience attention symptoms below the diagnostic threshold for ADHD (Acosta et al., 2006). Youth with NF1 also show more EF difficulties than their typically developing peers (Nupan et al., 2017). Given the range of cognitive difficulties previously described, it is unsurprising that many children with NF1 struggle academically, frequently meeting the diagnostic criteria for specific learning disabilities (Hyman et al., 2006; North et al., 1995). Because NF1 is characterized by a mild global lowering of IF, elevated rates of ADHD, EF difficulties, and academic difficulties, it is important to consider what specific difficulties, including general intellectual phenotype and trajectories over time, may be used to effectively identify predictors for later IF difficulties. Consideration of the role of ADHD and EF difficulties in trajectories of

IF is an important topic that is beyond the scope of the current study, though it warrants further investigation.

Development of IF in NF1 - Clues from Cross-sectional Designs

Cross-sectional studies of preschoolers, school-age children, and adolescents with NF1 suggest that youth with NF1 seem to neither fall further behind nor catch up to their typically developing peers. Current evidence based on cross-sectional data suggests that, for older children with NF1, the slope of overall intelligence development is similar to that of typically developing children. This assertion is supported by the lack of a statistically significant correlation between age and Wechsler Intelligence Scale for Children (WISC) Full Scale IQ (FSIQ) scores in what is, for the NF1 literature, a relatively large sample of about 100 children with NF1 (Hyman et al., 2005). More detailed investigations of second- or third-order intellectual factors correlated with age have not been conducted.

Similar to older children, the somewhat lower IF in preschoolers with NF1 appears to be independent of age, as is indicated by the lack of a statistically significant correlation between age and overall IQ score (Klein-Tasman et al., 2014). Additionally, no correlations between age, domain (i.e., verbal, spatial, nonverbal), or individual subtest standardized scores (including short-term memory) were found (Klein-Tasman et al., 2014). However, this study's sample of preschoolers with NF1 was small (N=37) and did not have sufficient power to detect small- to moderate-effect sizes (r's <.38; (Klein-Tasman et al., 2014).

Slopes of IF Development in Older Children with NF1 - Findings from Longitudinal Designs

There is very limited evidence from longitudinal studies that is consistent with the cross-sectional findings that indicate that the slope of intellectual development in older children with NF1 may be similar to that of their typically developing peers in several domains of IF. To the knowledge of this author, only two longitudinal studies of intellectual development utilize a sample of older children with NF1.

Using growth curve modeling with a mixed effects model, Cutting et al. (2002) found that vocabulary, visual-spatial, and fluid reasoning in youth with NF1 developed with similar slopes to those of their unaffected siblings. This study, however, was significantly underpowered (overall N=19, N=12 for 2–5 time points), and p-values (.2–.3) suggest that group-level slopes may actually differ significantly between youth with NF1 and typically developing children. Another caveat is that the study involves the use of three different IQ test versions (i.e., WISC-R, WISC-III, and WAIS-R). The authors attempted to correct for this statistically; however, when considering the high probability of age-to-test correlation, especially between the youth and adult versions, the interpretation of these results becomes more challenging.

In an 18 year follow-up study, a sample of 8- to 16-year-old children with NF1 (N NF1=18, control=5) demonstrated an apparently significantly improved FSIQ at a non-specific point in early adulthood which fell between the study's 8 year follow-up at Time point 2 (T2) and Time point 3 (T3)'s 18 year follow-up (Payne et al., 2014). While the authors reported no additional statistically significant differences at T2 between individuals with NF1 and control participants, this analysis did not have sufficient power to detect moderate or even large effect sizes. Therefore, although the longitudinal evidence is largely consistent with the cross-sectional evidence that indicates that older children with NF1 may show similar slopes of intellectual development to those of their typically developing peers, sample sizes in this study were too small to detect small to moderate effects and to allow for more confident conclusions.

Development of IF in Preschoolers with NF1 – Findings from Longitudinal Designs

Longitudinal study of intelligence development in preschoolers with NF1 is extremely limited and primarily indicates an urgent need to investigate this topic further. Wessel et al. (2012) conducted the only study that investigated cognitive development in NF1 across the preschool years. This longitudinal study of 43 infants, preschoolers, and school-age children with NF1 (all under 9 years of age) used the developmental screener *The Parents' Evaluation of Developmental Status* (Brothers et al., 2008), which very briefly screens for expressive and receptive language as well as gross motor, math, self-help, fine motor, and social-emotional skills based either on a single prompt to the child or parental report, using a single question to the parent; questions are tailored to the child's chronological age. Participants were categorized as "delayed" if they fell below the 17^{th} percentile on an overall index score of development based on this brief screener. At follow-ups of unspecified distances, participants exhibited a general trend towards an increasing number of delayed areas with increases in age (from 0 to 8 years). The large age range in early childhood and the varying follow-up intervals make these results difficult to interpret. Furthermore, the measure used relies on a single item per domain and does not have the strong psychometric properties characteristic of standardized measures of intellectual functioning.

Limitations of Prior Research and Extensions of Prior Research

As of 2020, the stability and predictive value of preschool IF in relation to later IF in the school-age and early adolescent years has not been investigated in youth with NF1. In cross-sectional samples, the existing literature on IF in NF1 has focused on group-level average comparisons with typically developing children. The small number of longitudinal studies of IF in NF1 have not investigated the stability and predictive value of preschool IF in relationship to later IF in children, and there is only one longitudinal study of IF in NF1 that included longitudinal data on preschoolers. However, the study included children from infancy into the early school years without specification of the follow-up interval, without specific examination of preschool IF, and using only a developmental screener to measure IF. Therefore, it remains unclear how stable and predictive preschool IF is for later IF in late preschool, the school-age years, and early adolescence in NF1-affected youth.

Conclusion and Rationale for the Current Study

In summary, although there is relatively consistent evidence for the general scope and degree of IF difficulties in NF1, research into the stability and predictive value of preschool IF in NF1 is generally lacking. Evidence from typical intellectual development suggests variability in the stability of intelligence during the preschool years and in trajectories of development from the preschool years into both the school-age and early adolescent years. However, sample and measurement characteristics of the available studies on typical intellectual development limit the generalizability of these findings in reference to the overall population such that findings from most existing studies may not accurately reflect the stability and variability of the IF trajectories of the general population. More specifically, many investigations included samples consisting primarily of children with above-average to superior intellectual abilities and from families with a relatively high socioeconomic status (SES). Additionally, longitudinal studies that included assessments over multiple years frequently used either revisions of a single IQ test or tests from different developers (Bayley, 1949; Hindley & Owen, 1978; McCall et al., 1973), which, in the absence of infrequently used structural equation modeling, creates barriers to precise interpretation of study results. Furthermore, very few longitudinal studies of typical intellectual development have investigated predictors and outcomes of the trajectories of intellectual development. These limitations in prior research of typical intellectual development increase the difficulty of forming precise hypotheses regarding the development of IF in NF1.

This study is the first longitudinally designed study with children with NF1 that tracks intellectual development from preschool into adolescence. Furthermore, this is the first study with children with NF1 to test the ways in which preschool IF is predictive of later IF in late preschool, the school-age years, and early adolescence. This study is also the first to examine NF1 IF longitudinally using the same examiner-administered IQ test over the entire study period from preschool to adolescence, rather than utilizing parental reports of cognitive functioning as was done in earlier studies.

Brief Study Description

This study will describe the reliability and predictive value of IF in the preschool years (overall IQ, verbal IQ, and nonverbal IQ) in relation to IF in the school-age and adolescent years in NF1-affected youth. The study will examine year-to-year reliability as well as the reliability of IF when comparing the IF of early and late preschool-age participants to their IF at between 9 and 13 years of age. The predictive value of early preschool IF will be examined by determining whether IF at one or more time points in the early preschool-age group predicts IF in the late preschool-age (e.g., 6 years). The predictive value of preschool IF in reference to IF in the 9- to 13-year-old age range will also be evaluated using predictions based on either a single time point or trajectories. The study will use both traditional correlational models and multiple regression models to predict IF. In addition, more complex statistical models, including linear mixed model growth curve models (LMMGCs), will be used to predict later IF.

Methods

Participants

As illustrated in Table 1, 69 children with NF1 participated in 1 or more visits, and 55 children with NF1 participated in 2 or more visits. Of the group that participated in 2 or more visits, 47 children attended 2 visits by age 8. Of all study participants, 27 children have follow-up data in school-age or adolescence (9–13 years). Table 2 describes the age of each participant at each visit.

Recruitment

Children with NF1 were primarily recruited from the Neurofibromatosis Clinic at the Children's Hospital of Wisconsin (CHW) Genetics Center and the Neurofibromatosis Program at the University of Chicago Medical Center (U of C). Additionally, other children with NF1 from the Chicago and Milwaukee areas that were not followed by the CHW or U of C were included; these individuals generally heard about the study through NF Midwest, which is one of the organizations that provided research funding. When families visited the CHW and U of C clinics, they were offered general information about the study through a study representative.

Screening and Diagnosis

Participants with brain tumors and seizures were excluded from the study. Diagnosis was made by the above-mentioned collaborating clinics and also by review of case records.

Materials

The measures selected are appropriate for preschoolers, school-age children, and adolescents and were selected to provide information about participants' cognitive and psychosocial functioning.

Differential Ability Scales, 2nd Edition

The Differential Ability Scales, 2nd edition (DAS-II) is a comprehensive, individually administered measure of cognitive abilities (Elliott, 2007). It has an excellent floor and is designed to examine cognitive abilities in a wide range of child populations, including children with intellectual disabilities and learning disabilities and problems. The DAS-II is normed for children between the ages of 2.5 and 17 years. The test has excellent internal consistency, strong test-retest reliability, and high correlation to other measures of cognitive abilities. A large, nationally representative sample was used to provide adequate age norms. The DAS-II was chosen for this study because it has excellent psychometric properties and appears sensitive in its ability to detect individual patterns of strengths and weaknesses. Administration time varies with age and ability and lasts an average of about one hour.

The DAS-II factor structure and its underlying tasks vary somewhat by age. The test uses three different age ranges to measure general cognitive abilities. The “lower level” includes ages 2.5–3.4 years, the “upper level” includes ages 3.5–6.9 years, and the “school-age level” includes ages 7–17.9 years. The General Cognitive Ability (GCA) composite is intended to reflect general cognitive functioning and includes verbal and nonverbal reasoning factors for the lower level participants (2.5–3.4 years) and verbal, nonverbal, and spatial reasoning factors for all other children (ages 3.5+). Most tasks for the second-order factors change, particularly from the upper-level to the school-age level, with the notable exceptions of the nonverbal Matrices task and the nonverbal/later spatial reasoning Pattern Construction Task. Beginning at age five years and continuing through the school-age assessments, there are additional tasks that load onto working memory and processing speed composites. Of note is the fact that, unlike the Wechsler tests

(WISC, WIPPSI), the “overall” IQ for the DAS-II never includes working memory and processing speed.

Conners Parent Rating Scales-Revised Short Form

The Conners Parent Rating Scales-Revised Short Form (Conners, Conners, 1997) is a commonly used measure of ADHD symptoms in children ages 13–17. The Conners norm-referenced t-scores used in the current study include the Cognitive Problems/Inattention Index and Hyperactivity Index. The Conners was included as a dimensional measure of attention and hyperactivity/impulsivity problems that may relate to IF performance in youth with NF1.

Hollingshead Index

The Hollingshead Index is a measure used to estimate socioeconomic status (Hollingshead, 1975). It includes the education and occupation(s) of the head(s) of the household, as well as their employment status. Of note, occupation has a somewhat higher weight than education (5x vs. 3x). The index has adequate internal consistency and strong cross-sectional convergent validity based on 1970 census data and is widely used in psychosocial research. For the purposes of this study, partners living in the same household without being married were treated equally to married partners in the calculation of the index.

Procedure

Session Procedure and Assessment Instruments

This study uses data collected from a larger longitudinal study that included an extensive battery of neuropsychological tests for the children with NF1, parent interviews, and questionnaires for both parents and teachers. The children’s direct assessments usually lasted three hours. The Differential Abilities Scale II (DAS-II), whose scores were used for this investigation, was always administered first to maximize the chances that scores reflected the best abilities of the child.

Statistical analysis, testing for normality, and assessment of outliers were conducted with R version 3.6.1 and R version 4.0.2. Potential univariate outliers for independent t-tests were identified as values with extreme z-scores (+-3.29) as recommended by Tabachnick and Fidell (2013b). Potential bivariate outliers for bivariate correlations were identified as having studentized residuals higher than +-2.

Study Aims, Hypotheses, and Analytic Strategy

Unless otherwise noted, all analyses were conducted with norm-referenced scores. When possible, R^2 values were reported as the effect size with 95% confidence intervals. Potential study drop-out effects were examined by conducting t-tests comparing children with a single visit to children with two, three, or four visits on general IF (DAS-II GCA Score), Conners Inattentive and Hyperactive symptom scores, and Hollingshead SES index scores. LMMGCs have been successfully fitted on smaller or similar sized longitudinal samples (Cutting et al., 2002; Lorenzo et al., 2015; Sansavini et al., 2014).

Aim 1: Describe the Reliability of General IF, Verbal IF, and Nonverbal IF and Evaluate the Degree of Change in IF Over the Preschool Period and Into Pre-adolescence and Adolescence in NF1.

To describe general reliability, year-to-year Pearson correlations were conducted for general IF, verbal IF, and nonverbal IF for participants between three and eight years of age (Table 4). It was expected that almost all year-to-year correlations would be statistically significantly positive correlations since there is a large range of correlations (.40–.80) reported in existing literature discussing typically developing children. Since the data from a more representative, typically developing sample suggests lower cross-age correlations, most correlations were expected to fall in the .40–.60 range. R^2 values were also reported. Power (see Table 2) is estimated to be adequate primarily for effect sizes in the large range (r's .49–.75) but low for medium effect sizes.

A reliable change proportion table was prepared to describe shifts of performance in the sample. Reliable Change Index (RCI) scores are a measure of the statistical stability of the scores of an individual on the same measure on repeated occasions (Jacobson & Truax, 1991). First, an RCI was computed for both preschool-age children and older children based on DAS-II Standardization sample test-retest reliability data (based on a 1–9 week test-retest interval) using means, standard deviations, and test-retest correlations for general IF, verbal IF, and nonverbal IF, based on the approach by Jacobson and Truax (1991). The test-retest correlations from the DAS-II were used to create RCI scores. These scores captured instability expected in the general population over brief test-retest intervals, reflecting primarily fluctuations in state and not necessarily long-term performance shifts (data for longer term stability are not available). The standard deviations of the relevant scores in the study's sample were used to capture the score variability in the youth with NF1. Hence, RCI scores beyond the cut-off of 1.96 indicated that changes on an individual level occurred were unlikely to be due to chance and reflected a change beyond the day-to-day variation in performance expected in the general population and considering the variability in IF in NF1. The proportions of the sample whose performance shifts were greater than the RCI cut-off value were reported using several time points across the preschool age and an additional time point in the school-age and early adolescent age categories, which were collapsed due to the presence of a single data collection point within the age range.

Comparisons were conducted between the 3- to 4-year and 6-year scores, the 3- to 4-year and 9- to 13-year scores, and the 6- and 9- to 13-year scores.

Aim 2: Identify Patterns in Early Preschool IF as Predictors for Late Preschool IF Difficulties in NF1

Aim 2a: Identify Whether Early Preschool-age IF at a Single Time Point is Predictive of IF at the End of the Preschool Period. It was expected that general IF at ages three to four years would predict general IF at age six years, and similarly that verbal and nonverbal IF at ages three to four years would predict verbal and nonverbal IF at age six years. Bivariate Pearson correlations were conducted to determine whether IF (general, verbal, and nonverbal) from ages 3.0 years to 4.9 years significantly predicted IF at age 6 years. R^2 values were reported as an effect size of the predictor. Power (see Table 2) was estimated to be adequate primarily for effect sizes in the large effect size range but low for medium effect sizes.

Aim 2b: Examine Whether Trajectories of Change in Early Preschool (Before Age 6) Predict IF in Late Preschool (at Age 6). The intention was to use LMMGCs (Fitzmaurice et al., 2011) to test how predictive individual linear growth curves, as well as an overall average growth curve, are in predicting intellectual functioning late in preschool (at age 6). In the interest of parsimonious modeling and because of the number of observations per child, models using the entire 3–13 age range were estimated, and the plan was to set the intercept at age 6. The correlation between the random intercept and the random slope (age) effect represents the degree to which the individual growth trajectories predict late preschool IF. If a linear slope over the 3–13 age range represents the data adequately (as evidenced by linear Lowess curves), then the age 3–13 linear slopes could be used to predict IF at age 6. If the slope is not linear and preschool and school-age IF appear to have different slopes (as evidenced by non-linear Lowess curves indicating two different slopes), growth slopes can be separated into two separate splines (e.g., 3–6 years, 6–13 years). It was expected that the individual growth trajectories (individual slopes) of the preschool years would significantly predict IF in late preschool (general IF, verbal IF, nonverbal IF).

Specification of Linear Mixed Model Growth Curves (LMMGC). To begin, Lowes smoothing curves were fitted on general IF, verbal IF, and nonverbal IF data to confirm that group-level trajectories were linear and if that was not the case, an adjusted model that more adequately represented the group-level growth curve was attempted. Individual participants' age at testing was used to anchor observations in order to account for the age of participants at the time of their entry into the study (Tabachnick & Fidell, 2013a). The LMMGC only uses available information to build the growth curve estimates for each individual case and can flexibly account for missing data, as linear slopes are estimated individually and anchored at the time of testing (West et al., 2014a). A "bottom up" strategy to model building was used starting with the simplest "unconditional means" model and progressively adding complexity to the model, including age effects and covariates (see Table 6). A bottom up strategy was employed to systematically build the linear mixed model and if modeling challenges arose, to identify the source of the model specification problem. For a brief introduction to linear mixed models, see Curran, Obeidat and Losardo's (2010) excellent primer. For a more detailed introduction to multi-level based growth modeling please refer to West, Welch and Galecki (2014a) and Singer and Willet (2003).

Covariate Selection and Specification in the LMMGCs. Several covariates were included because they are theoretically linked to the performance of intellectual functioning and were also linked in the attrition analyses to data not missing at random. First, sex was included as a time-invariant fixed effect because there was a tendency for boys to return for additional visits. SES is widely known to be related to IF and verbal functioning and vocabulary in particular, and SES was included as a time-varying covariate (values can change at each visit) and was measured with the Hollingshead Index. Parent-reported ADHD symptoms were measured as time-varying covariate by Conners Parent Rating Scales-Revised Short Form Cognitive Problems/Inattention Index Scores and Hyperactivity Index Scores (T-scores). Medication status was entered as a time-varying variable. Medication status was assessed at each visit and, for the purposes of this study, “ADHD medication status” was defined as a participant who was at the time of the appointment prescribed and taking a medication that likely affected ADHD symptoms (typically stimulants, occasionally SSRIs, and once Risperdal, an atypical antipsychotic medication). To center variables (in order to create meaningful intercept and avoid modeling problems), an adjusted “grand mean centered” approach was used wherein all continuous covariates (i.e., Conners Cognitive Problems/Inattention, Conners Hyperactivity, SES) were centered by using the general population mean for norm-referenced scores (Conners) and sample mean for non-norm referenced scores (Hollingshead SES), meaning that the intercept, as well as the fixed effects' beta values, reflected initial status and change in status based on population average ADHD symptoms and sample average SES.

Specification of DAS-II Related Covariates in LMMGCs. The composition of tasks mapping onto the “Nonverbal” DAS-II index changes at age 3 years and 6 months (Copying becomes a spatial task, Matrices become a new/nonverbal task). Because Age was centered at age 11 (in the middle of the follow up age range of 9–13 years), “2 GCA Indices” was coded as 1=GCA with 2 indices and 0=GCA with 3 indices so that at the intercept (age 11), the average child was assumed to have a nonverbal index in the context of GCA based on 3 factors. This covariate was only included in Nonverbal models, as Verbal models were not affected by this shift in tasks between the “lower level” to “upper level” DAS-II Early Years form battery. Therefore, any group-level shifts in performance that occurred on the Nonverbal Index scores from an early 3-year-old visit (<3-years, 6 months) to a later visit that appear to be related to change in the tasks included in the index were accounted for with this time-varying covariate. Similarly, many tasks in the Verbal, Nonverbal, and Spatial Indices on the DAS-II between the DAS-II Early Years and DAS-II School Years forms change; only “Pattern Construction” and “Matrices” stay as tasks across the preschool and school-age batteries, otherwise new tasks with qualitative differences are included (e.g., Naming in Early Years battery vs. Word Definitions in School-Age form) that reflect the qualitative shifts in cognition expected across IF development from the preschool period into the school-age period, reflecting an increase in more conceptual and abstract thinking. The variable “DAS Early Form” was coded as “0” for any child age 9 or older that completed the school-age form so that the intercept of the models reflects an 11-year-old participant with performance on the school-age form. This “DAS Early Form” variable was included in models with GCA, Verbal, and Nonverbal scores as a time-varying covariate. The “DAS Early Form” variable was included in the “final” (M4a) models with the goal of optimal representation of the shape of the growth curve if there is a non-continuous (step-like) shift in scores associated with the change to the school-age battery.

Estimators and Final Sample Selection for LMMGCs. For estimation, maximum likelihood (ML) estimation was generally used (as long as assumptions are met, see section on assessment of assumptions below) because it provides efficient model estimates and allows for model comparison of nested models with varying numbers of fixed effects, and restricted maximum likelihood (REML) estimation was also used to improve the robustness of fixed effect estimates (Singer & Willet, 2003; West et al., 2014a). REML fixed effect estimates were generally extremely similar to ML fixed effect estimates. To be concise, only ML estimates were reported, as they were used to compare nested models (e.g., models of same dependent variable where the more complex model can be created by the addition of parameters/covariates without omission of other variables/parameters). All analyses for linear mixed models were computed using the lme4 package for R with the lmer function. ML estimation involves an iterative process to estimate parameters with the smallest degree of residual variance; for explanation of how ML and REML estimation works, see West (2014a). To compare nested models with log-likelihood tests, each compared model needs to contain the exact same set of observations, and the estimators used (ML) require that, for any observation, all variables included in the model are complete. Therefore, for the growth models, only participants with complete data (i.e., DAS-II GCA/Verbal/Nonverbal scores, age, sex, Conners Cognitive Problems/Inattention Index, Conner's Hyperactivity Index and SES) for each observation (visit) were included. In a few cases, updated SES information was missing and was input based on the SES information from prior visits. Due to the limited number of observations per participant (2–3 on average), leaving participants with only one visit in the sample created modeling problems with random effects, as random effects (of random intercept and slopes) were also created for participants with only one visit. Therefore, to limit the number of different samples used and compared, the sample used for the LMMGC models included only participants with two or more visits (one visit in the preschool years and a second visit either in preschool or the follow-up school-age period). This resulted in a sample of 50 participants with complete data for at least two visits.

Assumptions Testing of LMMGCs Using ML and REML Estimators. Assumptions were made about the true individual change trajectories; however, only observed sample properties can be tested and evaluated with regard to the assumptions. The following strategies of assumptions testing were adapted from both guidelines by Singer and Willet (2003) and West et. al. (2014b). First, the functional form of the individual growth curves (at "Level 1") was assumed to be linear. Lowess curves were plotted for GCA, Verbal, and Nonverbal standard scores to test whether the form of the growth curve was linear at a group aggregate. Next, Ordinary Least Squares (OLS) fitted growth trajectories and data points of dependent variables, as well as empirical growth plots (scatter plots of data points), were plotted for each participant and examined for linearity. Of note, Singer and Willet (p128, 2003) note that, for three waves of data, it is difficult to "declare" any "curvilinear" functional form of the growth curves and that generally four or more observations per participant/case is required for curvilinear growth shapes. Next, linearity at the "Level 2" equations (the group-level covariates) was assumed. For dichotomous covariates, a linear model is de-facto acceptable (p126, Singer & Willet, 2003) and scatter plots with continuous covariates should reflect linear relationships. All continuous variables were time-varying covariates were already included in the Level 1 equations and therefore did not need to be tested for linear Level 2 relationships with the criterion variables. Linearity assumptions were tested before the linear mixed model was run. Normality and homoscedasticity related assumptions were tested after the model was estimated. Residuals were assumed to be normally distributed at both Level 1 and Level 2 of the linear mixed model. Q-Q plots of raw residuals (Level 1 and Level 2) involved plotting residual values against their normal probability scores and any significant departure from the line suggested departure from normality. Next, extreme values and normality were examined by plotting standardized residuals by participant ID (for models without covariates) and through bivariate plots between standardized residuals and each predictor variable (for models with covariates). Singer and

Willet (2003) suggested that, as rule of thumb, if standardized residuals are normally distributed, approximately 95% of residuals will fall within +-2SD of the center (e.g., only 5% greater than 2). Linear mixed models estimated with ML and REML also assume homoscedasticity for Level 1 and Level 2 residuals. Level 1 residuals were plotted against fitted values, and residual variability was assumed to be approximately equal at every level of the fitted value. Singer and Willett (2003) noted that, in small samples (e.g., in their example N=82), it is difficult to make definitive conclusions about homoscedasticity. Similarly, for models with covariates, Level 2 residuals (the random effects) are plotted against fitted values in "fitted vs residual" plots. As recommended (Singer & Willet, 2003; West et al., 2014b), only the "final" models (Table 6) were selected for assumptions testing as assumptions testing in linear mixed models is extensive.

Model Comparison and Effect Sizes of Linear Mixed Models. All nested models were compared with log-likelihood tests as well as with -2 log likelihood (deviance) values, AIC and BIC values. AIC values reflect the -2 log likelihood deviance and additionally account for model complexity (number of parameters estimated in the model). BIC values additionally reflect a penalty for small sample size. A "marginal" R^2 was also computed to allow for comparison between models during the model building process as well as between models with different dependent variables. Of note, see Lorah (2018) for a discussion of effect size measures for multilevel models, including controversies about computing R^2 or variance explained estimates, as variance can be explained at multiple levels in multi-level models and there is some concern as to whether variability of random effects truly reflects variability accounted for/predicted. Marginal pseudo R values allowed for comparison between non-nested models (e.g., models with different dependent variables) (see Tables 7–9 for model fit and Table 11 for model comparison). The marginal R^2 reflects the additional total variance explained by a multivariate model compared to a simpler nested model. R^2 was computed by R^2= 1- ((Level 1 + Level 2 variances of full model)/(Level 1 and Level 2 variances of empty model/simpler model)) as suggested by Lorah (2018). This R^2 was labeled "Marginal R^2" here because it is always in reference to another model. The intra-class correlation coefficient of the unconditional means models reflects an "Initial R^2", the amount of variability accounted for by the unconditional means model. This reflects the degree to which parallel trajectories of the same slope without covariates and only varying in intercept account for variability in scores over time. At a technical level, the ICC reflects the proportion of variance accounted for by the Level 2 cluster membership (i.e., the participant). The Initial R^2 of "empty" linear mixed models (with random intercept as the only parameter as in Model 1, "Initial R^2") is similar to the R^2 of correlations of two time points; a bivariate correlation between T1 and T2 basically reflects to what degree the slope of the trajectories from T1 to T2 are the same. In contrast, the Initial R^2 of linear mixed models refers to the variable number of data points per participant (2–6), and the bivariate correlation reflects the trajectories of only two time points. Of note, the Initial R^2 does not account for variable time intervals between visits; however, the models with Age as a time-varying covariate account for age at each visit and therefore account for varying time intervals between visits.

Rights and Serba (2019) proposed a framework of defining R^2 measures in multilevel models that provides both general total variance accounted for R^2 as well as flexible partitioning of the variance into the various components of linear mixed models. Their R^2 formulations were used to describe the total effect size of the final (M4a) models $R^{2}{}_{t}^{(fvm)}$ and the relative contributions of both the random intercept (flat trajectories) $R^{2}{}_{t}^{(m)}$ and fixed effects (age effect and other covariates) $R^{2}{}_{t}^{(f)}$ and the total variance accounted for by the final models. This allowed also for the informal comparison between the final models of different dependent variables and comparison with effect size of bivariate correlations. This was conducted with caution as there are complex considerations related to comparing R^2 of correlations and linear mixed effects models.

Aim 3: Identify Patterns in Preschool IF that Serve as Predictors for IF Difficulties Between Ages 9 Years and 13 Years

Aim 3a: Test the Hypothesis That Preschool IF at a Single Time Point Predicts IF Between Ages 9 Years and 13 Years. To test whether IF (general, verbal, and nonverbal) at ages 3.0–4.9 years, as well as at ages 5–6 years, significantly predicted IF at ages 9–13 years, bivariate correlations were conducted. R^2 values were reported as the effect size of the predictor. Power (see Table 2) was estimated to be adequate primarily for effect sizes in the large range but low for medium effect sizes.

Aim 3b: Test whether preschool IF trajectories predict IF between ages 9 years and 13 years. It was expected that the individual growth trajectories (individual slopes) of preschool-age children would significantly predict IF between ages 9 years and 13 years. LMMGCs were used to model individual linear growth curves as well as an overall (average) growth curve for the given model. Three separate models were computed for general IF, verbal IF, and nonverbal IF. Individual participants' age at testing was used to anchor observations in order to account for the variable ages of participants at the time of their entry into the study. R^2 values were reported. For some children, it was possible that a significant difference existed when comparing the slope of growth in preschool to the slope of growth from late preschool into ages 9–13 years. In an exploratory analysis, a splined slope LMMGC was used to estimate a model with two separate slopes, one from ages 3 to 6 years and one from ages 6 to 13 years. Fit of the prior models was compared to this more complex splined slope model based on R^2, AIC, and BIC values. See Aim 2b notes on linear mixed model specification and model comparison as well as assumptions testing.

Results

R 3.6.1 and R 4.0.2 was used for the analyses. A p-level of .05 was considered statistically significant. Effect sizes for mean level differences (Cohen's D) were interpreted as .3=small effect, .5=medium effect and .8=large effect; for correlations, .1=small effect, .3=medium effect, and .5=large effect; for R^2, .01=small effect, .09=medium effect, and .25 large effect. Assumptions for normality were fulfilled for all correlational analyses. Outliers for bivariate correlations (identified as ones with extreme studentized residuals) were omitted from the calculation if omitting the extreme values led to a change in the significance level of the p-value. The number of omitted participants is indicated in each given table.

Sample Attrition

Attrition was analyzed to test whether missing data (participants not returning or returning fewer times) was random (see Table 3). Participants who participated in one preschool visit (between ages 3 and 8) and returned for a school-age visit were compared with those who did not. The groups showed equivalent SES, level of parent reported attention problems and nonverbal functioning, and relatively equivalent GCA (4.3 SS difference, p=.180). However, those who returned showed lower parent reported hyperactivity symptom levels and higher verbal functioning. Within the preschool visit years, participants who returned for more visits had higher SES and fewer hyperactivity symptoms and were somewhat more likely to be male (statistical trend). Therefore, all the above-mentioned variables were included as covariates in the linear mixed models.

Aim 1: Description of the Reliability of General IF, Verbal IF, and Nonverbal IF and Evaluation of the Degree of Change in IF Over the Preschool Period and Into Pre-adolescence and Adolescence in NF1

Reliability

To describe general reliability, year-to-year Pearson correlations were conducted for general IF, verbal IF, and nonverbal IF for participants between three and eight years of age (see Table 3a). As expected, year-to-year correlations fell primarily in the .40–.60 range. For the GCA and Verbal standard scores, all year-to-year correlations were statistically significant. In contrast, only the Age 5 to Age 6 correlation for nonverbal standard scores was statistically significant. For GCA scores, the prior year typically predicted about 40% of the variability in scores the next year with possible increases in accounted variance to around 70% between Age 7 and Age 8 from the late preschool age (Age 5 to Age 6, see Table 4).

Reliable Change

To describe the frequency of reliable change, RCI scores were computed (see Methods section for computational details) for GCA SS, Verbal SS, and Nonverbal SS scores, and proportions of participants falling into reliably increased, reliably decreased or unchanged are reported in Figure 2. Of note, the cut-off values of a standard score difference considered to be reliable change varied significantly between different types of scores and was generally higher for Nonverbal scores than Verbal SS or GCA SS scores, reflecting the expected variability in these scores. The samples of the different time spans (ages 3–4 to 6, 6 to 9–13, 3–4 to 9–13) share some, but not all, participants. To maximize sample size, all participants in a given follow-up time span were included. The pattern of stability and change is relatively similar for GCA SS and Verbal SS scores over the preschool period. With close to 40% of participants showing a significant increase in scores over the preschool period. Some participants showed a significant decrease from age 6 into the school-age years for GCA (25%) and Nonverbal (20%) SS scores.

Aim 2: Identify Patterns in Early Preschool IF as Predictors for Late Preschool IF Difficulties in NF1

Aim 2a: Identify Whether Early Preschool-age IF at a Single Time Point is Predictive of IF at the End of the Preschool Period.

As expected, GCA, Verbal and Nonverbal scores at ages 3–4 statistically significantly predicted scores at Age 6 with moderate to large effect (Table 5), with early preschool performance accounting for 18–32% of variance of Age 6 scores.

Aim 2B and 3B: Are Slopes of IF in Early Preschool Age Predictive of IF at the End of the Preschool Period and Are Slopes of IF in Early Preschool Predictive of IF in the School-age Years?

Lowess smoothed curves of all observations of all participants were examined for GCA SS, Verbal SS, and Nonverbal SS scores with a smoothing span of 1/3, 5 “robustifying” iterations, and delta=.01 * range of x using R (see Figures 3–5). The Lowess curves suggested an increase in scores of almost 8 points during the preschool age and a subsequent similarly sized reduction in scores, albeit with a more moderate slope. Individual trajectories based on observations of each participant were also examined separately and supported two splines (Ages 3–6, 6–13), suggesting along with the Lowess curves that two splines may meet the linearity criterion at the Level 1 equations. In the step-up model selection process, initially an Unconditional Means (UM) model (M1) was estimated that includes no fixed effects and only one random effect (the intercept) (see Table 6 and Figure 1). The UM model reflects a model in which each participant has a flat (slope=0) growth curve - essentially, the mean score across different observations for a given participant. The UM model can be used to calculate the intraclass coefficient (ICC), which is a measure of the degree of correlation of subsequent observations of the same participant on average for the sample; however, it does not account for varying time between intervals. The ICC also reflects the proportion of variance accounted for by the random effects in the model. Using the REML estimation method, the model failed to converge. The UM GCA model did converge, however, with the ML estimator (see Figure 1) with an ICC of .59, indicating that about 59% of the variance in participants' scores is accounted for by a one flat trajectory (that varies only by level of intercept not by degree of slope).

Next, an unconditional growth (UG) model (M2) was estimated utilizing a fixed effect of age (the group-level slope of an age effect) and two additional random effects: 1) an age random effect allowing for individual linear growth slopes (which can be different from 0) and 2) a covariance random effect that allows the age random effect (slope) and the random intercept (individual intercept) to be related, which allows for a trajectory to predict the intercept. Age was initially centered at Age 11 (median of Ages 9–13) to reflect growth slopes predicting the outcome of Age 11. The UG model using the ML estimator had the convergence problem of a singularity of the correlation between the random intercept and the random age effect was 1. Of note, while there was sufficient variability in the intercepts of growth curves (SD=14.66), the slopes (age random affect) only had a standard deviation of 0.76, indicating that slopes tended to vary by less than 1 SS point in GCA. Given the estimation problems with random slope effects with the unconditional growth model (M2), the age random effect was removed and M2a was estimated.

Next, M3a, a model with a growth spline (Ages 3–6, 6–13), was fit according to the Lowess curve, suggesting that a growth spline may provide a more optimal fit of the true growth curve. For M4a, covariates (as explained earlier in the model specification portion of the Methods section) were added to estimate growth while also controlling for likely factors influencing intellectual functioning performance, such as ADHD symptoms, ADHD medication status, SES, as well as factors that may influence IF performance, such as change of forms/battery of tests from the DAS-II "Early Years" Form to "School Age" Form and the DAS-II factor structure shift from 2 to 3 factors with the addition of the Spatial factor for ages >3:5. For both M3a and M4a, models were also run with the age random effects, but, similarly to M2, ML and REML estimated models had convergence problems of "singularities" in both the correlation of random effects and the size of random effects. Given that participants primarily had three observations, and at times only two, the linear mixed model adjusts for a low number of observations per participant by estimating the individual growth curve as being relatively similar to the group/population growth curve, which reduces the size of the age random effects. Similar problems were observed when building the Verbal and Nonverbal models with regard to singularities for the models involving age random effects.

Assumptions of the final model (M4a) were examined for each outcome variable (GCA, Verbal, Nonverbal) regarding homogeneity and normality of residuals and possible influential values. Residuals for GCA and Verbal scores were generally homogenous and relatively normally distributed at both Level 1 and Level 2 of the respective final models. In contrast, Level 1 residuals (Pearson standardized based on the standard deviation of the outcome variable) based on Levine's test indicated significant heterogeneity ($p=.043$). A square root transformation of the criterion (Nonverbal SS) resulted in trend-level heterogeneity based on Levine's test, and log transformation resulted in homogenous Level 1 residuals for Nonverbal scores in the final (M4a) model. Given that the focus of this study related to the predictive value of the models, emphasis was placed on examining potential influential values relevant to the variance of the intercept and Level 1 residuals. For the GCA and Verbal scores, 4 participants for each were identified (out of 50) with abnormal covariance ratios (based on the R HLMdiag function; for internal criteria, see Loy & Hofmann 2014) and removing these participants from the models tended to reduce the intra-class-correlation coefficients of the empty models and increase marginal R^2 covariates (see Table 11). For the Nonverbal SS, the standardized residuals, Cook's D values for Level 1 and 2, covtrace, and covariance ratios indicated a larger number of potentially influential values. Between 6 and 11 potentially influential values were identified. The M4a models with different numbers and groupings of influential values removed varied in accounting for 19–40% of total variance in scores. The final M4a model for the Nonverbal scores (see Table 12) reflects a compromise of removing six influential participants, which reflects the most conservative results while also removing relatively few influential participants. Of note, several of these participants reflect very large swings in SS scores (at times 30–50 SS points), which were verified by multiple graduate researchers and appear to reflect genuine severe fluctuations in performance from year to year or over longer time spans, rather than scoring errors. Table 12 illustrates the impact of removing varying amounts/types of influential values from the Nonverbal M4a models (spline with all covariates).

Nested models were compared with log-likelihood tests (see Table 11). Given that spline models fit significantly better than models with just a single slope, the final estimated models included a spline. The final models (M4a) for GCA, Verbal, and Nonverbal SS as criterion variables also included the covariates identified on a theoretical basis. The addition of the covariates statistically significantly improved only the Verbal Linear Mixed Model but not the GCA and Nonverbal models. Figure 9 illustrates that the unconditional means model accounted for 59% of the variability in GCA SS, 66% of the variability in Verbal SS and only 18% of the variability in Nonverbal SS. The addition of the age spline, which allowed the general group/population slope during the Preschool years to be different from the slope during the School-age years, accounted for an additional 2–4% of variance of IF performance over time (see Table 11). Inclusion of the covariates accounted for an additional 5–10% of variance in the GCA and Nonverbal models and 16% of variance in the Verbal Model.

In the final model (M4a), the random intercept (the flat and uniform trajectory portion of the model) accounted for 53-56% of the total variance in the GCA SS and Verbal SS models and only 11.9% of the total variance in the Nonverbal SS model (see Figure 8). All fixed effects combined (i.e., age and all other covariates) accounted for 7-10% of the total variance in GCA and Nonverbal SS models and 17% in Verbal SS models. Finally, Figure 9 illustrates amount of total variance accounted for by the entire model for the final linear mixed models (M4a) in the context of variability accounted for by bivariate correlations from Aim 2a and 3a. For both GCA and Nonverbal SS, effect sizes are relatively similar between bivariate correlations and linear mixed models, especially for the bivariate correlations between Ages 3–4 and Ages 9–13. For verbal scores, about 43-55% of additional total variability in scores over time can be accounted for by the linear mixed model that includes more than 2 time points for most participants and also includes covariates (sex, SES, inattention and hyperactivity symptoms, DAS-II form).

Table 9 illustrates the shape of the population/group growth curves and the effects of various fixed effects. Of note, IF tended to increase (2.5–3.6 SS points per year for GCA and Verbal) during the preschool years and was relatively flat during the school-age years. In the

Verbal model, higher SES was associated with higher IF (and at trend level in the GCA model). The nonverbal M4a model with all participants indicated that youth with NF1 scored almost 9 points higher in the Nonverbal model when taking ADHD medications (an effect that was statistically significant). However, this effect disappeared when 6 influential participants were removed (see Table 10); hence, for a few children who displayed a large shift in scores, medication was associated with their Nonverbal score shifts. Of note, only a few observations included children who took ADHD medications, and a larger sample, especially with older children, is needed to replicate an ADHD medication effect. Also, youth with NF1 scored, on average, 7 SS points lower on Verbal scores once they moved to the School Age Form at Age 9 ($p=0.051$). Over time, parent-reported attention symptoms tended to be associated with lower IF for both GCA and Verbal scores, with trend level findings. The level of IF was not different by sex.

Table 13 illustrates that scores decreased significantly for GCA, Verbal, and Nonverbal SS scores into the school-age years. M4a models indicated that scores tended to drop about 7 Standard Score points for the Verbal model once youth completed the School-age battery, and there appeared to be a similar direction of effect, though not statistically significant, for the GCA and Nonverbal models. To test whether there is a general reduction in scores over time in the school-age years regardless of the shape (e.g., primarily one drop in scores associated with the more conceptual battery versus continuous decrease in scores), M5a models were computed. They indicated that GCA, Verbal, and Nonverbal SS scores all statistically significantly declined into the school-age years.

Aim 3: Identify Patterns in Preschool IF That Serve as Predictors for IF Difficulties Between Ages 9 Years and 13 Years.

Aim 3a: Test the Hypothesis That Preschool IF at a Single Time Point Predicts IF Between Ages 9 Years and 13 Years (Without Any Covariates)

GCA and Verbal SS early in the preschool period (ages 3–4.9 years) predicted school-age respective scores with medium to large effect and there was a trend for Nonverbal SS in early preschool period to predict school-age Nonverbal SS. Late preschool IF predicted school-age IF in GCA SS and Verbal SS but not in Nonverbal SS. Nonverbal SS at ages 5–6 accounted for only 7% variability in Nonverbal SS at ages 9–13.

Discussion

In the current study, the stability and predictive value of preschool intellectual functioning in NF1 for intellectual functioning in late preschool and into the school age years were examined. Results indicate that general IF and verbal functioning can be reliably measured in NF1 in the preschool years and that these measures hold some predictive value for IF in the school-age years. In contrast, particularly in the preschool years, measures of nonverbal functioning are not reliable in youth with NF1. Both bivariate correlational approaches and more complex linear mixed models were used. For Verbal scores, variability in scores was better accounted for in complex models than in simple (bivariate) models. For GCA scores, bivariate and linear mixed models both accounted similarly well for variability in scores. For Nonverbal scores, both simple models and the complex models specified in this study accounted relatively poorly for variability in scores over time from the preschool years into the school-age years in NF1. A portion of youth with NF1 appear to increase general IF performance and verbal performance within the preschool years relative to their same-aged typically developing peers (as measured by changes in standard scores), and there is evidence for a decrease in performance (relative to same-aged peers) into the school-age years based on the linear models. While there is variability in the level of functioning of youth with NF1, the trajectories of youth with NF1 appear to be uniform in slope, with children of different abilities (though same level of covariates, e.g., same level of SES) following generally the same developmental trajectory. These findings are discussed in more detail below.

Reliability

Consistent with expectations, the prior year statistically significantly predicted the next year's IF for both the general IF and verbal domains. However, nonverbal IF was unrelated from year to year throughout the preschool age, except for ages 5–6. A contrasting pattern in effect size was observed, with verbal reliability the highest early in preschool and early in the school-age years. Given the lower test-retest reliability (1–9 weeks) of the nonverbal tasks on the DAS-II in the standardization sample, it was expected that reliability would be lower on nonverbal tasks and that about 40% of variability may not be accounted for due to day-to-day fluctuations in nonverbal task performance. However, youth with NF1 showed additional score variability not accounted for by day-to-day fluctuations seen in the general population. A control group of typically developing participants (preferably matched for SES and mean IQ) with test-retest at one year intervals, as well as data on NF1 short test-rest reliability (1–9 weeks), could provide context to potential sources of unaccounted variability, allowing researchers to discern measurement unreliability from shifts in the underlying construct. Ultimately, structural equation modeling approaches to measurement would be optimal, though generally prohibitive in sample size requirements for samples in NF1, which tend to be less than 100 and frequently are less than 50. Given that many youth with NF1 have significant attention difficulties, it makes sense that nonverbal IF performance, which may be more vulnerable to difficulties in sustaining attention, is less consistent over time than verbal IF performance. When compared to the representative sample of Schneider et al. (1999) of typically developing children, the preschool participants with NF1 in this study demonstrated similar stability in the development of their general IF across preschool. The reliability of verbal performance in this study's NF1 sample seems to be similar or higher than the Schneider et. al. (1999; 2010) sample .

Stability - Reliable Change Over the Preschool and into the School-age Period in IF

Relative to their same-aged peers, a sizeable proportion of youth with NF1 make significant gains in verbal functioning and in overall intellectual functioning scores over the preschool period. In contrast, few children show statistically significant gains or losses in nonverbal functioning over the preschool years. Of note, the cut-off for a statistically significant gain or decrease in nonverbal performance for age 3 was 20 SS points, which is a rather large criterion; this emerges from the test-retest reliability of nonverbal scores on the DAS-II from the norming sample, together with the variability in Nonverbal SS seen in the current sample (which was larger than for Verbal or GCA SS in our sample). Shifts in nonverbal performance as large as 20 SS points can be expected in NF1 over the preschool or school-age periods, such that changes at the individual level are difficult to interpret.

In the school-age years, Verbal scores for the vast majority of the sample stabilized; hence, the youth with NF1 did not shows significant changes in verbal functioning into the school-age years at an individual level. Few children made gains or losses in Nonverbal scores into the school-age years. On Nonverbal functioning and on the overall measure of intellectual functioning, the GCA, some youth with NF1 demonstrated decreases into the school-age period.

The similarity of patterns of gains and losses between GCA and Nonverbal scores may be driven by the fact that the GCA scores include nonverbal functioning task performance as a component.

Patterns of IF Early in the Preschool Age That Predict Late Preschool or School-age Difficulties in NF1

Early preschool IF predicted 18–34% of late IF, which is at the lower end of the effect size range of the prior literature (15-65%, Hindley & Owen, 1978; Schneider et al., 1999), though similar to the results by Schneider's nationally representative sample in Germany (13-20%). Children with NF1 may show even greater score instability than typically developing children; this would be an important question for future research. The predictive value of IF for the school-age years was highly variable. Depending on the aspect of IF being measured (GCA, Verbal, Nonverbal) and age at the time of first measurement, between 7% and 56% of school-age performance was predicted by preschool IF performance. In the Schneider et. al. (1999), sample preschool IF predicted 10–20% of school-age IF, and older studies tended to indicate that preschool functioning did not account for more than about 25% of school-age functioning (Hindley & Owen, 1978; Honzik et al., 1948). Given that GCA at ages 3–4 accounted for an impressive 56% of variability in school-age general IF (a finding that needs to be replicated with a larger sample), general IF measured with GCA may be a particularly promising predictor of future IF in NF1, particularly if only one preschool measurement point is available.

Linear mixed models were estimated in an attempt to assess the degree to which group level and individual growth trajectories of IF can predict school-age IF in NF1. Models that allowed for individual slopes to vary (random effects for age) showed convergence problems, apparently due to minimal variability of estimated individual growth curves, indicating that in the current sample, trajectories were estimated to be primarily unitary in NF1 across the preschool and school-age years. Since individually varying slopes were not estimated in the final models, it was not possible to address whether there are differential trajectory patterns in early childhood that might predict school-age outcomes. However, the uniform trajectories predicted a sizeable amount of variability in school-age GCA and Verbal SS over time, though in the not Nonverbal SS. Spline and covariates generally accounted for little additional predictive value except for Verbal SS; for Verbal SS, inclusion of covariates substantially increased the amount of variance for which the growth trajectories accounted. In sum, although there may be some methodological issues obscuring variability in individual trajectories of IF in NF1 as measured by standard scores, the current data do seem to suggest that youth with NF1 tend to have relatively uniform trajectories relative to each other and may vary more in initial level of IF than in individual trajectory.

To provide some context for the respective predictive value, some tentative comparison of the predictive value of the simple and complex models is warranted. When comparing the full linear mixed models to the bivariate correlations, both GCA and Nonverbal score prediction did not seem to meaningfully benefit from the more complex models. In contrast, in the Verbal final model, the linear mixed effects model accounted for a much larger proportion of the variability in scores over time, with an impressive 73% of variance accounted for as compared to 18–30% variability accounted for in bivariate correlations. Hence, particularly for verbal functioning, it may be helpful to have more than one time point to improve prediction accuracy of verbal functioning in the school-age years. As mentioned earlier, while general IF appears to be reasonably reliable and of good predictive value (bivariate and linear mixed models), it may not automatically be a preferred predictor of later IF, at least in the NF1 population, because it relies in part on highly unreliable Nonverbal scores and the predictive value of general IF may be driven by the relatively reliable Verbal scores. More power (larger sample size) is needed to effectively compare the predictive values of different variables of IF (general intellectual vs. verbal functioning) as the precision of the parameter estimates (e.g., correlation coefficients) of the current study was insufficient to effectively test this.

The existing literature on the development of IF in NF1 during the preschool years is extremely limited. Studies have primarily measured IF cross-sectionally (Klein-Tasman et al., 2014; Lorenzo et al., 2013; Sangster et al., 2010). Findings indicate that mean scores fall around 90 SS, with the cross-sectional variability of scores similar to that of typically developing children and differences between different domains yet not reliably identified (Klein-Tasman et al., 2014; Nupan et al., 2017; Sangster et al., 2010). Data regarding the reliability and predictive value of preschool IF in NF1 were sparse. In the only longitudinal study including preschoolers, which was based on a developmental screener with only a single question/task per developmental domain, there was no significant change in proportions of children with delayed or non-delayed classification for receptive and expressive language (Wessel et al., 2012) within ages 0–8-years. The current study used a well-validated, comprehensive, individually administered measure of IF instead of a developmental screener and found that a sizeable portion of the sample (~40%) made significant verbal gains over the preschool years. On a group level, general IF and verbal functioning increased over the preschool years and general IF, verbal, and nonverbal IF decreased into the school-age years. It is important to consider that the nature of the tasks on the DAS-II changes from the preschool period to the school-age period, consistent with the kind of skills expected with increasing development. More specifically, within the verbal domain, tasks moved from identification and naming of objects to more conceptual understanding of words, with similar increases in levels of abstraction and conceptual relationships expected within the nonverbal domain. It may be that traditional interventions in the preschool age years are effective in helping preschoolers with NF1 catch up to some to their peers in the kinds of intellectual functioning-related tasks that are administered during those years, and that remediation of challenges with more conceptual understanding is more difficult in youth with NF1.

Preschool IF in the nonverbal domain was generally not reliable, and early preschool nonverbal functioning did not predict late preschool nonverbal functioning. It is possible that high rates of attention problems in NF1 may contribute to this lack of reliability and predictive value of early preschool nonverbal functioning. Measurement issues related to the changing factor structure of domain scores on the DAS-II may also be a contributing factor. Of note, about 15 (ages 3.0–3.5) of the 150 total included observations included Nonverbal scores that were also partially based on performance of the Pattern Construction task, which relies on visuo-spatial performance, one of the hallmarks of the cognitive phenotype of NF1 (North, 2000; Nupan et al., 2017). However, the residuals in the Nonverbal linear mixed model were similarly variable for those observations as compared to the older children, indicating that performance on the Pattern Construction task under the Nonverbal domain was unlikely to be the primary cause of unreliable Nonverbal scores. Youth with NF1 have demonstrated difficulties with simple and complex motor tasks (Nupan et al., 2017), difficulties with processing of visuo-spatial stimuli (some of which based on EEG evidence may be related to ineffective attentional allocation, Ribeiro et al., 2014), and very prevalent general attentional difficulties. General attention difficulties, fine-motor difficulties, and difficulties processing visual stimuli may interact in youth with NF1, resulting in unstable nonverbal task performance in NF1.

Similarly, there is very little literature on the development of IF in NF1 in the school-age years and no published literature about the predictive value of preschool functioning for school-age IF. Cross-sectional studies involving samples that span both preschool age and school-age children and lack of significant correlations of IF with age suggested a flat growth trajectory for children with NF1, including not falling further behind their peers nor gaining ground in their IF (Hyman et al., 2005; Klein-Tasman et al., 2014) and a longitudinal study with a small sample (N=12, Cutting et al., 2002) seemed to confirm the hypothesis that the slope of IF in NF1 is similar to that of typically developing children. Although Cutting et al. (2002) collected some longitudinal data on the IF of 5–20-year-olds with NF1, they reported no information regarding the predictive value of late preschool IF and reliability of school-age IF in NF1. The current study provides data that general IF in the preschool age - at a single point and regardless of when measured - is a robust predictor of general IF in the school-age years in NF1. Furthermore, particularly for the prediction of verbal functioning, inclusion of other potential predictors (e.g., SES, ADHD symptomatology, type of verbal task/level of abstraction) may be helpful in youth with NF1. In contrast, Nonverbal scores, particularly late in the preschool years, seem to be minimally relevant to nonverbal functioning in the school-age years based on the current findings. Of note, the current study was able to demonstrate that there is an increase in IF relative to peers during the preschool age years, yet youth with NF1 appear to plateau or perhaps even fall somewhat behind over time into the school-age years. One hypothesis based on the current findings is that, as conceptual demands increase, youth with NF1 appear to be unable to keep pace with their typically developing peers. Overall, this study provides the first evidence, with a moderately-sized sample, about the reliability and predictive value of IF in NF1 and the shape of the overall growth trajectory of IF in NF1 across early and middle childhood.

Clinical Implications

The findings of the current study provide important information for the clinician trying to capture IF for a child with NF1 and determine a prognosis. Children with NF1 tend to increase their IF scores on average by about 2–3 Standard Score points over the preschool years (~6–9 SS over the preschool period). Hence, preschoolers appear to be able to make some moderate gains on verbal and nonverbal tasks that require minimal deeper conceptual understanding. Once in the school-age years, youth with NF1 struggle with more abstract understandings of concepts (e.g., being able to articulate the core features of a word, understanding/articulating understanding of abstract terms) somewhat more than would be expected based on their prior verbal functioning. Scores in early preschool age may actually be a better indicator of anticipated level of difficulty in the school-age years than scores in the late preschool age, particularly for Verbal scores. The verbal domain may likely be a relatively effective predictor of subsequent verbal functioning and, given the generally unreliable nonverbal functioning for preschoolers with NF1, nonverbal functioning appears to be limited in relevance to current functioning. Given that the complex model involving multiple measurement points was more useful in accounting for variability in verbal functioning, it may be important for children with NF1 to be assessed multiple times over the preschool period (particularly for Verbal scores). Tracking of IF over time allows for the identification of trajectories that reflect stable, improving, or decreasing IF standard scores. However, repeated testing may also produce practice effects that may mask a child with NF1's increased challenges keeping up with peers in the preschool age or may overestimate gains.

Limitations and Future Directions

In the current study, a number of methodological and statistical factors made it difficult to effectively detect variability in individual growth trajectories and the extent to which such trajectories would predict school-age IF. In the linear mixed models, the correlation between the intercept ("outcome" at Age 11) and the individual participant's growth trajectory tended to be close to 1 or considered 1, indicating that generally the slope was "too perfect" of a predictor for the outcome at Age 11. This finding has several causes, some of which have to do with variables related to data collection and data analysis; others may have to do with the nature of the development of IF in NF1. Linear mixed growth curve models tend to be conservative when it comes to the estimation of individual growth curves. Individual slopes for individual growth curves are estimated not only by a participant's observations, but also by the average slope of the group. Further, many participants only had two observations, and few had three or four observations. The linear growth models are also conservative and, when few observations are available for a participant, the model assumes that the growth curve is relatively similar to the mean growth curve (Singer & Willet, 2003). In addition, only cases with three or more observations contribute to the variance estimation (e.g., estimation of age random effect) and a sizable portion of the sample only had two visits, such that the sample contributing to the variance estimation is small. Most participants had only one visit in the school age years, such that longitudinal research within the school age years is still needed. Additionally, participants with more observations tended to have fewer ADHD symptoms, which therefore may have decreased variability in the longitudinal sample by including fewer individuals with greater ADHD symptoms. Random effects can be more difficult to estimate effectively in smaller samples. Finally, standard scores are designed to be more stable and reduce some of the variability of performance measured by raw scores.

Even the individually estimated lines of best fit (based on ordinary least squared estimation of simple linear regression, not linear mixed models) indicate fairly similar trajectories are present across participants. Of note, Figure 6 displays GCA grouped (by participant) Spaghetti plots and Figure 7 displays lines of best fit for individual participants (individually fitted linear models) without adjustment to a mean trajectory (as in linear mixed models where the individual trajectory can be estimated as a combination of mean trajectory of the group and individual trajectory based individual data points). Hence, while the relatively low number of observations per case may have somewhat obscured the detection of differences in individual growth trajectories of IF in NF1, overall growth trajectories may be similar across youth with NF1 (holding covariates constant). An important focus of future research in IF development in NF1 should be accounting for more of the still relatively high variability of performance of each child with NF1 around the relatively uniform (group) growth trajectory.

Given that nonverbal functioning was not reliably measured, future research needs to identify effective predictors of fluctuations in IF performance in NF1, especially for nonverbal performance. Nonverbal scores showed an especially low intra-class correlation in the linear mixed model compared to the GCA and Verbal scores. This was graphically apparent with a lot of individual data point variability around the linear trend of best fit of the individual growth curves. Of note, a key will be that these predictors are time-varying covariates that are measured at every time point. Our parent-reported attention and hyperactive symptom reports were not each uniquely helpful in contributing to the growth curve models. The current study is a first attempt at modeling IF development in NF1 from the preschool into the school-age years with emphasis on the predictive value of preschool IF for school-age IF in NF1, and a detailed analysis of effective covariates is beyond the aims of this study. There are multiple ways of measuring ADHD symptoms and including them in linear mixed models (e.g., different dependent measures from questionnaires, performance-based measures), which may more effectively capture both day-to-day fluctuations in performance and long-term fluctuations in performance.

The current study did not include assessment of effort, task persistence or frustration during the assessment process. Measurement of effort has been acknowledged as important for neuropsychological assessment in children (DeRight 2014; Carone, 2015). Children with ADHD in particular struggle with task persistence (Foley et al., 2008; Scime & Norvilitis, 2006). In children with ADHD, level of incentive has improved task persistence (Dovis et al., 2012). Anecdotally, during the assessments of the current study, some children exhibited variable effort and limited task persistence and given the high prevalence of ADHD in NF1, assessment of effort, task persistence and frustration may be important covariates to assess intellectual functioning with optimal validity and account for fluctuation in intellectual functioning performance.

Substantial attrition and a small sample size were significant limitations of the current study. Addressing these limitations will be critical to increasing the sensitivity of detecting true IF variability in NF1. More observations per participant are key (e.g., mostly 3, preferably 4+) so that the linear mixed model can be more sensitive to the true variability in slopes that exists between youth with NF1. A moderately sized initial sample combined with large overall attrition (~50%) resulted in few cases with four or more observations. Future research in NF1 needs to assess and proactively problem solve barriers that cause high levels of attrition in longitudinal samples with NF1 (e.g., financial incentive size, length of battery, contact information likely consistent over time). Furthermore, non-random missing data and drop out resulted in a "higher functioning" longitudinal sample (e.g., higher IF, fewer ADHD symptoms) from more privileged

SES backgrounds, which may have also reduced variability in trajectories due to a more homogenous sample than the general NF1 population.

Finally, randomized clinical trials (RCT's) for medications aimed at treating NF1 cognitive symptoms or potentially normalizing NF1 brain development will require reliable measures of IF. The current study suggests that general IF and verbal IF may be moderately reliable measures and may be acceptable to be included in an RCT as independent measures but that nonverbal IF (at least as measured by DAS-II) may not be sensitive enough to detect changes/differences between groups and that it will be important to improve concurrent and future prediction of nonverbal performance in youth with NF1. Raw or other absolute performance scores (e.g., ability scores on the DAS-II), which may be more sensitive to change, have yet to be tested regarding their reliability properties in NF1. Further, the reliability of subtest-level performance of IF tasks needs to be examined to elucidate whether subtest level performance has sufficient sensitivity and specificity to be included in randomized clinical trials as an outcome measure. The current study also did not account for possible practice effects, a concern that Chelune et al. (1993) raised regarding repeatedly assessing neuropsychological functioning; the effect of repeated testing warrants further investigation, but can also be addressed by including a control group.

In sum, challenges for future research are to a) increase the number of observations for the average participant (preferably 4+ observations for improved individual slope variability detection and so individual non-linear trends can be detected); b) increase sample size (to allow for more complex non-linear modeling); c) include covariates that may explain variability in scores, especially nonverbal scores (e.g., attentional state on day of assessment, etc.); d) include a typically developing comparison group (siblings or non-siblings) with multiple benefits, including increased sample size for increased power, some improvement in random effect estimation, and more precise parameter estimates and increased power for comparison of effect sizes; e) use scores that increase variability (e.g., raw scores, DAS Ability scores), which may improve the likelihood of more sensitive detection of variation in individual growth trajectories across individuals. f) delve deeper into the IF development period between ages 5–7 in NF1, as there appears to be a shift in patterns of reliability and potentially increased inter-individual variability in slopes; h) test hypotheses about what underlies the increase in IF in preschool and change in slope in the school-age years (e.g., increased treatment of ADHD symptoms, change in level of conceptual demands); i) investigate why some participants increase and others stay stable over the preschool period (e.g., by using classifications from the RCI).

Now that some of the predictive value of preschool IF for school-age IF in NF1 is demonstrated, it is critical to investigate whether preschool IF predicts later difficulties in academic functioning (e.g., learning disabilities or subthreshold LD learning problems) and adaptive functioning. Many youth with NF1 experience academic problems, and IF difficulties can be a risk factor for academic problems (North, 2000; Nupan et al., 2017). The prevalence of substantial ADHD symptoms in youth with NF1 (Acosta et al., 2006) may make IF assessment less reliable in NF1 and reduce the predictive value of preschool IF for later academic problems. Additionally, clinicians and parents need to know about risk factors in the preschool age that may be effective targets for early behavioral and medication-based intervention to mitigate the risk for later academic problems. Research addressing risk factors such as ADHD symptoms and phonological awareness, in addition to IF, could provide this critical early intervention relevant information.

Summary

The current study investigated the stability and predictive value of intellectual functioning starting in the preschool years in NF1 into the school-age years. The results of the current study support the idea that general IF and verbal IF can be relatively reliably measured across the preschool period and that preschool general and verbal IF hold predictive value for school-age IF. In contrast, performance on nonverbal tasks was inconsistent across the preschool period in youth with NF1 and had very limited predictive value for school-age IF. Significant portions of the current sample made moderate gains during the preschool years, and performance dropped in the school-age years almost half of a standard deviation; increased conceptual demands arising in the elementary school years may present particular challenges for some youth with NF1. A large amount of variation of Verbal scores was accounted for by the general group trajectories and other covariates (i.e., SES, ADHD symptoms, sex, change in DAS form, ADHD medication status) indicating that for Verbal scores, the context of relevant covariates is important. While the relatively small number of observations per participant and attrition patterns may have resulted in underestimates of individual variability between individual growth trajectories, based on the data from the current study, growth trajectories of IF in youth with NF1 are relatively uniform and variability occurred mainly around these general group trajectories. In sum, general IF and verbal IF can be measured reliably in NF1 in the preschool years and, depending on timing and number of measurement points, general IF and verbal IF very early in the preschool years can be strong predictors of school-age IF. In contrast, nonverbal IF seems unreliable in NF1 and more efficient predictors/covariates need to be identified in order to make meaningful predictions of school-age nonverbal IF based on preschool IF. While it can be expected that children with NF1 will make gains in general and verbal IF over the preschool years, those gains appear to be lost over the school-age years and IF measured late in preschool may somewhat overestimate IF in the later school-age years. General IF and verbal IF difficulties in preschool are a promising risk factor candidate for academic difficulties in the school-age years in NF1. Nuanced understanding of the predictive value of IF in the preschool years may be helpful in the assessment of early risk predictors and treatment planning.

Table 1.

Summary of Participant Visits by Age

	Number of Participants at Visits by Age						N per Age Ranges		
Visit number	3	4	5	6	7	8	3–6 years	3–8 years	9–13 years follow up
Visit 1	21	13	7	6	5	3	47	55	
Visit 2		17	15	6	4	5	39	48	5
Visit 3			11	10	4	2	21	27	7
Visit 4				7	1	5	7	13	8
Visit 5					2	1	0	3	7
Visit 6									2
Total # of visits by age/range	21	30	33	29	16	16	113	145	27

Table 2

Sample Size and Power Calculations for Correlational Models

Aim 1a						Aim 2a	Aim 3a	
	Cross-age correlations					Cross-age-range correlations		
Correlation age pairs	3,4	4,5	5,6	6,7	7,8	3–4,6	3–4,9–13	5–6,9–13
Observed n	17	23	22	9	11	18	17	24
Power for r=.3	33%	41%	40%	20%	24%	34%	33%	42%
Power for r= .5	68%	81%	79%	41%	50%	71%	68%	82%
r for 80% power	0.56	0.49	0.51	0.75	0.69	0.55	0.57	0.49

Table 3

Attrition - Mean Differences of Participants That Remained and Those That Dropped out at 9–13-year-old Follow-up and Correlations with Number of Preschool Visits

Variable	With and without 9–13-year-old follow up			Number of early years form visits	
	Mean Difference	*p*	Cohen's d	*r*	*p*
Female	r=.03	.63	-	-.17†	.082
SES	2.04	.489	0.187	.32*	.016
Inattention T	-3.74	.257	-0.29	-.08[b]	.520[b]
Hyperactivity T	-10.85 [a] **	.0001a	1.00	-.27[b]*	.038[b]
GCA SS	4.3	.180	0.33	-.04	.752
Verbal SS	7.70*	.023	0.58	.05	.730
Nonverbal SS	-1.32 [c]	.651 [c]	0.11	.03	.87

Note. T = T-score (M=50, SD=10), GCA SS = General Conceptual Composite Standard Score from DAS-II, Verbal SS = Verbal Cluster Standard Score from DAS-II, Nonverbal SS = Nonverbal Cluster Standard Score from DAS-II. Two-sided p-values with significance level indicated.

[a] One participant removed with high Conners Hyperactivity scale T-scores and high residuals.

[b] Conners Inattention Problems T-scores and Conners Hyperactivity T-scores were square root transformed for residuals to be normally distributed.

[c] One participant removed with greater than 130 Nonverbal SS score and high residual

NS p>.1, †.10<p<.05, *p<.05, ** p<.01, *** p <.001

Table 4

Year-to-Year Bivariate Correlations Ages 3–8 (Aim 1a)

Domain	Correlations (CI's) by Age Pairs				
	3,4 (N [a]=17)	4,5 (N [a]=23)	5,6 (N [a]=22)	6,7 (N [a]=9)	7,8 (N [a]=11)
GCA SS	.63** (.22–.85)	.63*** (.30–.82)	.46* (.05–.74)	.68† (-.04–.94)	.84*** (.49–.95)
Verbal SS	.82*** (.56–.93)	.69*** [b] (.39–.86)	.51*[b] (.09–.77)	.67† (-.06–.93)	.88*** (.60–.97)
Nonverbal SS	.14 (-.36–.58)	.21[c] (-.23–.58)	.57** [d] (.17–.81)	-.30 [e] (-.85–.58)	.43 (-.23–.82)

Note. Pearson bivariate correlations, r values with 95% confidence intervals in parentheses and 2-sided p-values with significance level indicated, Two-sided 95% confidence intervals.

[a] Number of complete paired observations, some correlation pairs had lower N's if participants were removed due to being considered influential data point because of extreme studentized residuals (statistically significant Bonferroni corrected studentized residual)

[b] One participant with Verbal SS <70 was removed from correlation due to extreme residual, with influential value Age 4–5 r=.82***, Age 5–6 r=.70***

[c] One participant with 50 SS increase was removed, with outlier, r=-.01, p=.95

[d] Two participants with 40 SS shifts in scores showed extreme residuals and were removed from correlation, with influential values Age 5–6 r=.-0.04 ns

[e] Two participants with extreme residuals were removed from correlation, with influential values Age 6–7 r=.02 ns

NS p>.1, †.10<p<.05, *p<.05, ** p<.01, *** p <.001

Table 5

Bivariate Correlations to Predict Late Preschool-Age and School-Age Intellectual Functioning (Aim 2a & 3a)

Parameter	Bivariate correlations 3–4.9 to 6 (N[b]=18)	3–4.9 to 9–13 (N[b]=17)	5–6.9 to 9–13 (N[b]=24)
GCA	.58*[c] (.14–.83)	.75***[e] (.42–91)	.61**[e] (.26–.82)
Verbal	.47†[c] (-.01–.78)	.55* (.10–.82)	.43*[c] (.02–.71)
Nonverbal	.42† (-.07–.75)	.43†[e] (-.07–.76)	.27[f] (-.16–.62)

[a] Pearson bivariate correlations, r values with 95% confidence intervals in parentheses and 2-sided p-values with significance level indicated

[b] Number of complete paired observations, some correlation pairs had lower N's if participants were removed due to being considered influential data point because of extreme studentized residuals (statistically significant Bonferroni corrected studentized residual)

[c] One participant with SS <60was removed from correlation due to extreme residual, with influential value for 3–4 to 6 GCA r=.76**, 3–4 to 6 Verbal r=.78***, 3–4 to 9–13 Verbal r=.56**

[d] One participant with SS >120 was removed from correlation due to extreme residual, with influential value r=.57*

[e] One participant with SS >140 was removed from correlation due to extreme residual, with influential value 3–4 to 9–13 GCA r=.67**, 3–4 to 9–13 Nonverbal r=.73**, 5–6 to 9–13 GCA r=.68***

[f] Two participants with SS >130 were removed from correlation due to extreme residual, with influential value r=.38†

NS p>.1, †.10<p<.05, *p<.05, ** p<.01, *** p <.001

Table 6

Model Specification of Linear Mixed Model Growth Curves of Compared Models

Term	Unconditional means M1	Unconditional growth M2	Unconditional growth w/o age random effect M2a	Unconditional growth with spline w/o age random effects M3a	Conditional growth model with spline w/o age random effects M4a	Conditional growth model with spline w/o age random effects and DAS Early Form M5a
			Fixed effects			
Intercept	x	x	x	x	x	x
Age_11 – centered		x				
Age 3–6				x	x	x
Age 6+ (spline)				x	x	x
Gender (TIC)					x	x
SES (TVC)					x	x
Conners Inattentive (TVC)					x	x
Conners Hyperactive (TVC)					x	x
ADHD med status (TVC)					x	x
2 GCA indices (TVC)a – for NV models					x	x
DAS Early Form					x	
			Random effects			
Intercept	x	x	x	x	x	x
Age_c11		x				
			Covariance Parameters for D Matrix Child (i)			
Variance of intercepts	x	x		x	x	x
Variance of age effects		x				
Covariance of intercepts, age effects		x				
Dependent variable			GCA, Verbal, Nonverbal SS			
Covariance Parameters for R Matrix: Time level: Residual variance	x	x	x	x	x	x

Table 7

Model Fit and Standard Deviation of Random Intercept and Level 1 Residuals for GCA Models Without Age Random Effect

Fit Indices	Unconditional means model (M1)	Unconditional growth w/o age random effect (M2a)	Unconditional growth spline w/o age random effects (M3a)	Conditional growth spline w/o age random effects (M4a)
.-2 log likelihood (deviance)	1035.2	1034.9	1023.0	1015.9
AIC	1041.2	1042.9	1033.0	1037.9
BIC	1050.0	1054.7	1047.8	1070.4
Residuals				
Random Intercept SD (Level 2)	8.512	8.522	8.569	7.860
Residual SD (Level 1)	7.106	7.095	6.691	6.667

Note. Akaike Information Criterion (AIC) adjusts for model complexity. Bayesian Information Criterion (BIC) adjusts for model complexity and sample size relative to model complexity. Nonverbal SS was log transformed because Level 1 residuals were initially not homogeneous, reflects removal of 6 participants due to influential values (extreme Relative Covariance, Covtrace, Cook's D values), with influential values, ADHD med status p=.04 (+8.67SS). N=4 participants were removed for Verbal SS, with all participants: Conners Inattentive NS. N=4 participants were removed from GCA SS model, with all participants: SES NS, Conners Inattentive NS.

Table 8

Model Fit and Standard Deviation of Random Intercept and Level 1 Residuals for Verbal Models Without Age Random Effects

Fit Indices	Unconditional means model (M1)	Unconditional growth w/o age random effect (M2a)	Unconditional growth spline w/o age random effects (M3a)	Conditional growth spline w/o age random effects (M4a)
.-2 log likelihood (deviance)	1062.1	1061.9	1044.6	1030.0
AIC	1068.1	1069.9	1054.6	1052.0
BIC	1077.0	1081.8	1069.4	1084.6
Residuals				
Random Intercept SD (Level 2)	10.294	10.267	10.520	9.341
Residual SD (Level 1)	7.302	7.303	6.654	6.491

Note. Akaike Information Criterion adjusts for model complexity (AIC). Bayesian Information Criterion (AIC) adjusts for model complexity and sample size relative to model complexity (AIC). Nonverbal SS was log transformed because Level 1 residuals were initially not homogeneous, reflects removal of 6 participants due to influential values (extreme Relative Covariance, Covtrace, Cook's D values), with influential values, ADHD med status p=.04 (+8.67SS). N=4 participants were removed for Verbal SS, with all participants: Conners Inattentive NS. N=4 participants were removed from GCA SS model, with all participants: SES NS, Conners Inattentive NS.

Table 9

Model Fit and Standard Deviations of Random Intercept and Level 1 Residuals for Nonverbal Models Without Age Random Effects

Fit Indices	Unconditional means model (M1)	Unconditional growth w/o age random effect (M2a)	Unconditional growth spline w/o age random effects (M3a)	Conditional growth spline w/o age random effects (M4a)
.-2 log likelihood (deviance)	-182.5	-183.1	-187.8	-191.9
AIC [a]	-175.5	-175.1	-177.8	-167.9
BIC [b]	-167.8	-163.5	-163.3	-133.2
Residuals				
Random Intercept SD (Level 2)	0.053	0.052	0.051	0.043
Residual SD (Level 1)	0.113	0.113	0.111	0.111

Note. Akaike Information Criterion adjusts for model complexity (AIC). Bayesian Information Criterion (AIC) adjusts for model complexity and sample size relative to model complexity (AIC). Nonverbal SS was log transformed because Level 1 residuals were initially not homogeneous, reflects removal of 6 participants due to influential values (extreme Relative Covariance, Covtrace, Cook's D values), with influential values, ADHD med status p=.04 (+8.67SS). N=4 participants were removed for Verbal SS, with all participants: Conners Inattentive NS. N=4 participants were removed from GCA SS model, with all participants: SES NS, Conners Inattentive NS.

Table 10

Fixed Effects Estimates for Conditional Spline Growth Models (M4a)

		Dependent Variable					
		GCA SS		Verbal SS		Nonverbal SS (log)[a]	
Parameter	Covariate type	β	*p*	β	*p*	β	*p*
Intercept		104.82		119.19		4.54	
Age 3–6 spline	TVC	2.48	0.002	3.61	<.001	0.01	.453
Age 6 spline	TVC	-0.20	.807	.46	.568	-0.01	.292
Gender	TIC	.40	.882	-2.02	.510	0.01	.769
SES	TVC	0.19	.072	.35	.003	<.001	.901
Conners Inattentive	TVC	-0.13	.095	-0.14	.074	<-0.01	.358
Conners Hyperactive	TVC	0.10	.277	.07	.467	<.01	.103
ADHD med status	TVC	2.05	.49	.99	.737	0.04	.456
2 GCA Factors	TVC					0.05	.912
DAS Early Form	TVC	4.27	.235	7.00	.051	0.05	.469

Note. With p-values of Type II Wald Chi-Square Tests for fixed Effects, TIC= Time Invariant Covariate, value is constant across time, TVC= Time Varying Covariate, value can change over time, measured at every visit, p-values in parentheses, 2 GCA Factors Variable coded as "1" if observation was during age 3.0–3.5 and DAS GCA had only verbal and nonverbal factors as compared to Age 3:6+ when DAS GCA includes verbal, nonverbal and spatial composites, DAS Early Form Variable coded "1" when DAS Early Form was used (ages 3–8) as compared to the DAS School Age form, Empty cells: for the Intercept p-values were not reported by lme4 function for fixed effect intercepts and "Covariate type" does not apply, "2 GCA Factors" variable was not included in GCA and Verbal models because for Verbal tasks there is no change at Age 3:6 and only one task changes for GCA at age 3:6; N=4 participants were removed for Verbal SS, with all participants: Conners Inattentive NS, N=4 participants were removed from GCA SS model, with all participants: SES NS, Conners Inattentive NS, for Nonverbal model reflects removal of 6 participants due to influential values (extreme Relative Covariance, Covtrace, Cook's D values), with influential values, ADHD med status p=.04 (+8.67SS).

[a] Nonverbal SS was log transformed because Level 1 residuals were initially not homogeneous.

Table 11

Model Comparison of Final Linear Mixed Effect Models (Without Age Random Effects With ML [e] Estimation by Log-Likelihood Tests, M4a)

Models Compared (Nested vs. Reference)	chi square (marginal)	*p*	R2 (marginal)	R2 (marginal, with outliers)
		GCA		
UM vs UG w/o age random effects	0.240	0.624	.000	0.000
UM vs UG spline w/o age random effects	12.162	0.002	.039	.014
UM vs CG spline w/o age random effects	19.277	0.013	.136	.074
UG Spline w/o age random effects vs CG spline w/o age random effects	7.115	0.310	.101	.060
		Verbal		
UM vs UG w/o age random effects	0.170	.680	.003	.003
UM vs UG spline w/o age random effects	17.553	0.0001	.027	.007
UM vs CG spline w/o age random effects	32.069	<.0001	.188	.146
UG Spline w/o age random effects vs CG spline w/o age random effects	14.516	0.024	.165	.139
		Nonverbal		
UM vs UG w/o age random effects	0.541	0.462	.008	.003
UM vs UG spline w/o age random effects	5.23	0.073	.040	.028
UM vs CG spline w/o age random effects	9.401	0.401	.084	.098
UG Spline w/o age random effects vs CG spline w/o age random effects	4.164	0.761	.046	.072

Note. ML=Maximum Likelihood Estimation, UM = Unconditional Means Model (M1), UG = Unconditional Growth Model (M2a), UG Spline w/o age random effects = Unconditional Growth Model without age random effects (M3a), CG Spline w/o age random effects = Conditional Growth Model without age random effects (M4a), GCA: without influential values N=4 based on assumptions testing (including Cook's D, Covtrace, RVC beyond threshold values with focus on random effects and residuals), Verbal SS: without influential values – N=4 based on assumptions testing (including Cook's D, Covtrace, RVC beyond threshold values with focus on random effects and L1 residuals), Nonverbal SS: without influential values – N=6 based on assumptions testing (including Cook's D, Covtrace, RVC beyond threshold values with focus on random effects and residuals) and log transformation of criterion variable NV_SS.

Table 12

Model Comparison for Nonverbal Linear Mixed Effect Models (M4a) –With Varying Numbers of Participants Removed From Initial Sample of N=50

N removed/Type of R2	With all outliers	Group 1 removed [a]	Group 2 removed [b]	Group 3 removed [c]
Number participants omitted	0	6	11	9
Unconditional Means Model ICC	33.0%	18.3%	36.5%	28.4%
Spline Additional R^2	2.9%	3.8%	4.5%	3.8%
Covariates Additional R^2	6.7%	4.6%	9.8%	10.7%
$R^2_t{}^{(fvm)}$	35.0%	18.6%	40.0%	32.3%

Note. ICC is the intra-class correlation coefficient and reflects the proportion of variance that the random intercept in the Unconditional Means Model accounts for;$R^2_t{}^{(fvm)}$ reflects the proportion of total variance accounted for by the full M4a Nonverbal model.

[a] Group 1: Influential RVC values

[b] Group 2: Influential RVC values, also 1 extreme L2 Cook's D value, 3 Level 1 residual extreme values

[c] Group 3, Influential RVC values, most influential L1 and L2 Cook's D values

Table 13

Selected Fixed Effects Estimates for Conditional Spline Growth Models (M4a) and M5a models

Parameter	Dependent Variable: GCA SS		Verbal SS		Nonverbal SS (log) [f]	
	M4a	M5a	M4a	M5a	M4a	M5a
Age 3–6 spline	2.48 (0.002)	2.53(.001)	3.61 (<.001)	3.71(<.001)	0.01 (.292)	0.01 (.441)
Age 6 spline	-0.20(.807)	-1.05(.01)	.46 (.568)	-0.92(.020)	-0.01 (.769)	-0.01 (.045)
DAS Early Form[e]	4.27 (.235)		7.00 (.051)		0.05 (.469)	

Note. P-values in parentheses, 2 GCA Factors Variable coded as "1" if observation was during age 3.0–3.5 and DAS GCA had only verbal and nonverbal factors as compared to Age 3:6+ when DAS GCA includes verbal, nonverbal and spatial composites, DAS Early Form Variable coded "1" when DAS Early Form was used (ages 3–8) as compared to the DAS School Age form; Verbal SS: 4 participants were removed for Verbal SS, with all participants: Conners Inattentive NS; GCA SS=4 participants were removed from GCA SS model, with all participants: SES NS, Conners Inattentive NS; Nonverbal SS: reflects removal of 6 participants due to influential values (extreme Relative Covariance, Covtrace, Cook's D values), with influential values, ADHD med status p=.04 (+8.67SS).

[f] Nonverbal SS was log transformed because level1 residuals were initially not homogeneous,

Figure 1

Linear Mixed Model Selection and Related Hypotheses

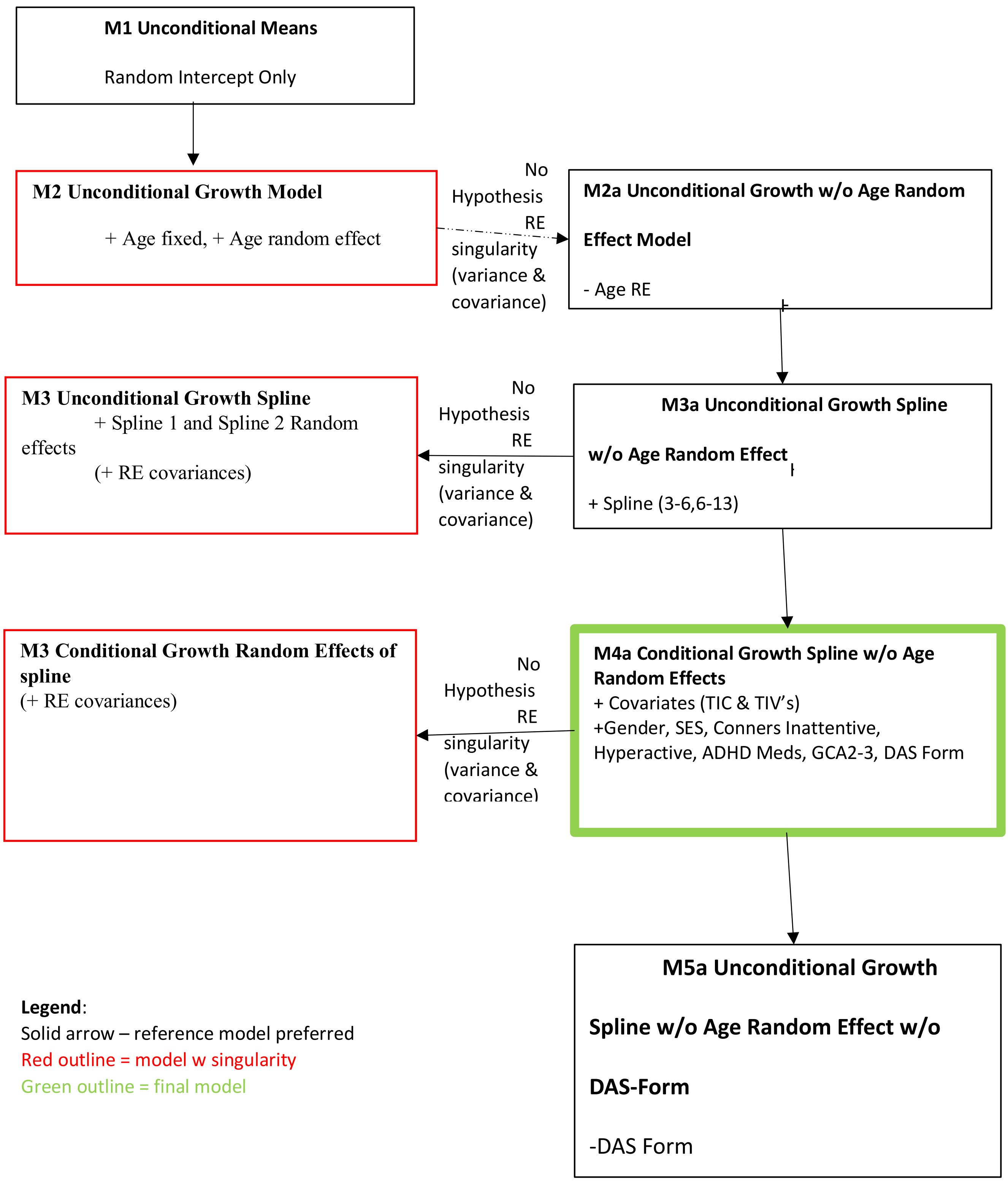
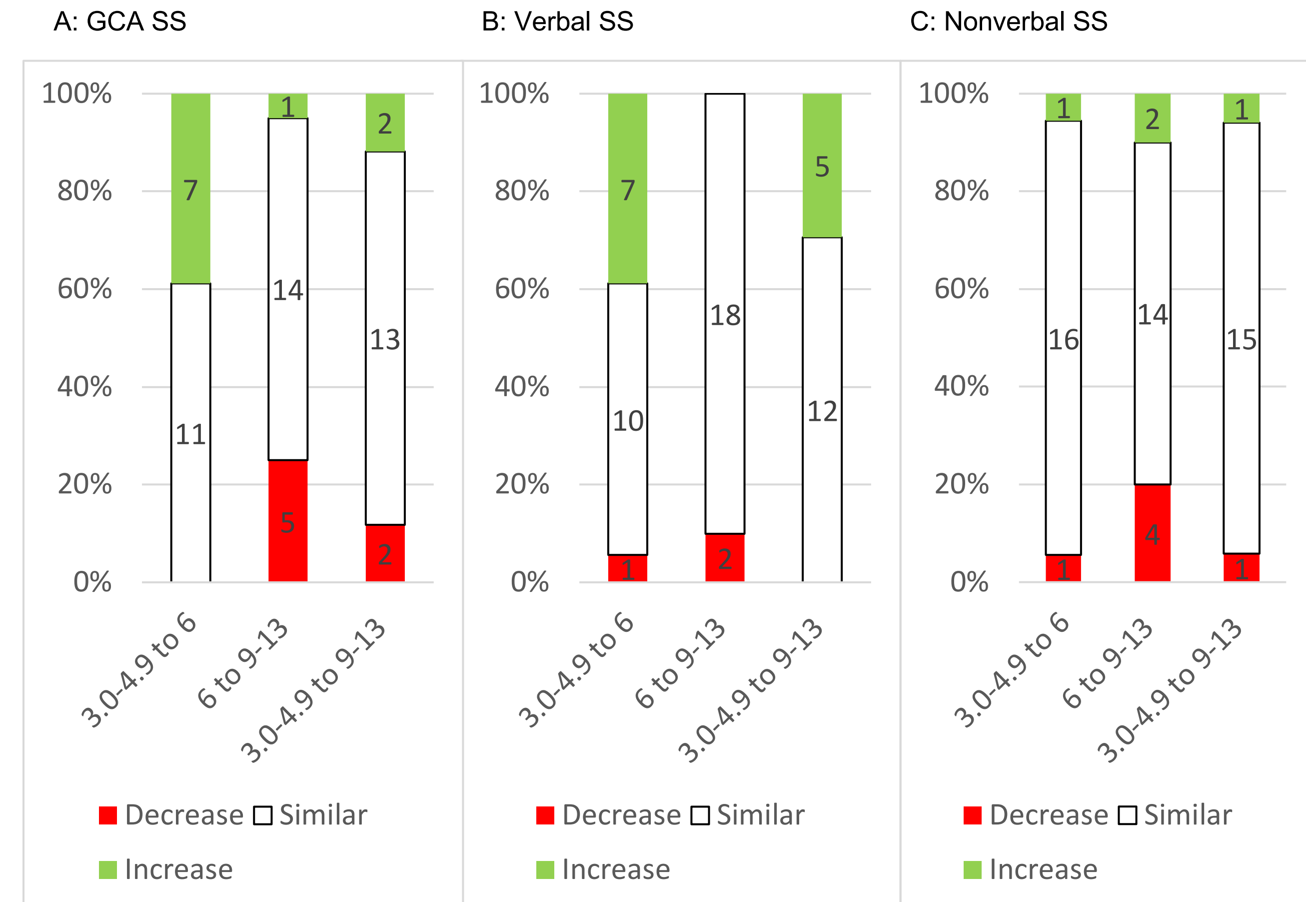


Figure 2

Proportion of Participants with Increased, Decreased, or Stable for GCA SS, Verbal and Nonverbal SS Based on Reliable Change Index Scores

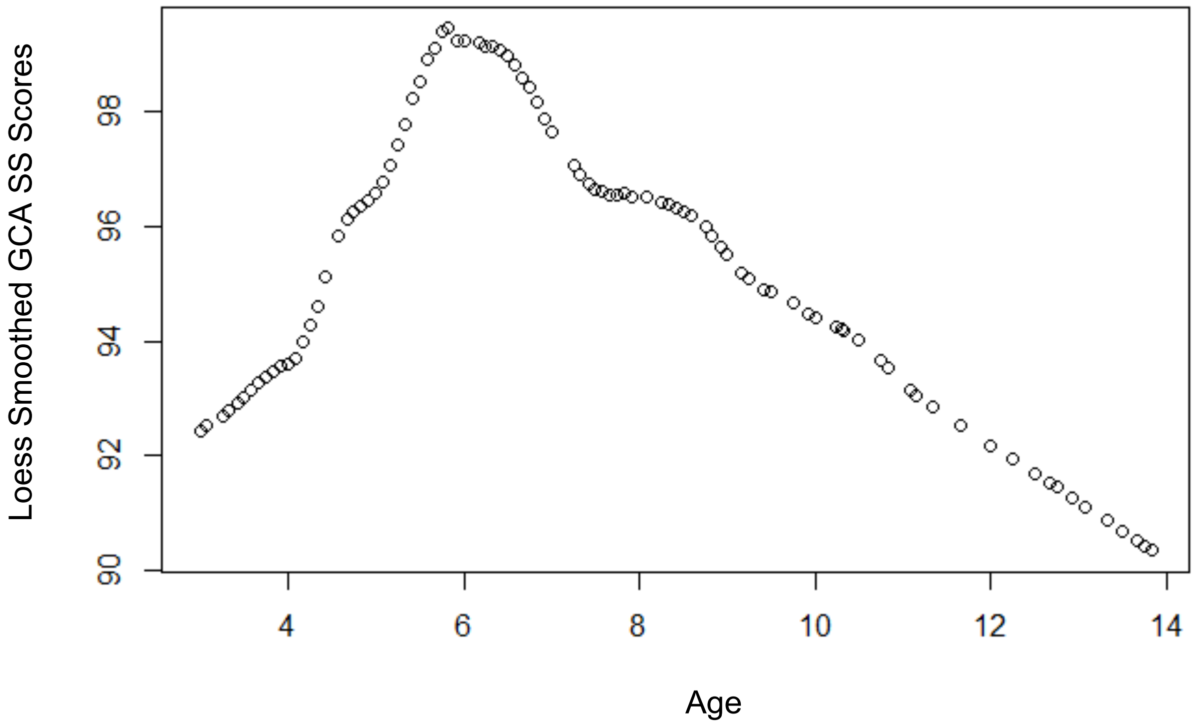


	GCA SS		Verbal SS		Nonverbal SS	
Age	3	6	3	6	3	6
r	.91	.91	.90	.87	.73	.77
SD	11.21	12.63	11.91	16.60	14.11	12.79
RCI cut off	9.33	10.50	10.44	16.58	20.32	17.01

Note. A= GCA SS RCI classification graph, B= Verbal SS RCI classification graph, C=Nonverbal SS RCI classification graph. “r” is the DAS-II norming sample test-retest reliability, “SD” is the standard deviation of the current study sample at the given age, “RCI cut off” is the two-sided 95% cut off based on DAS-II norming sample test-retest r and the current study’s SD for the given variable and given age.

Figure 3

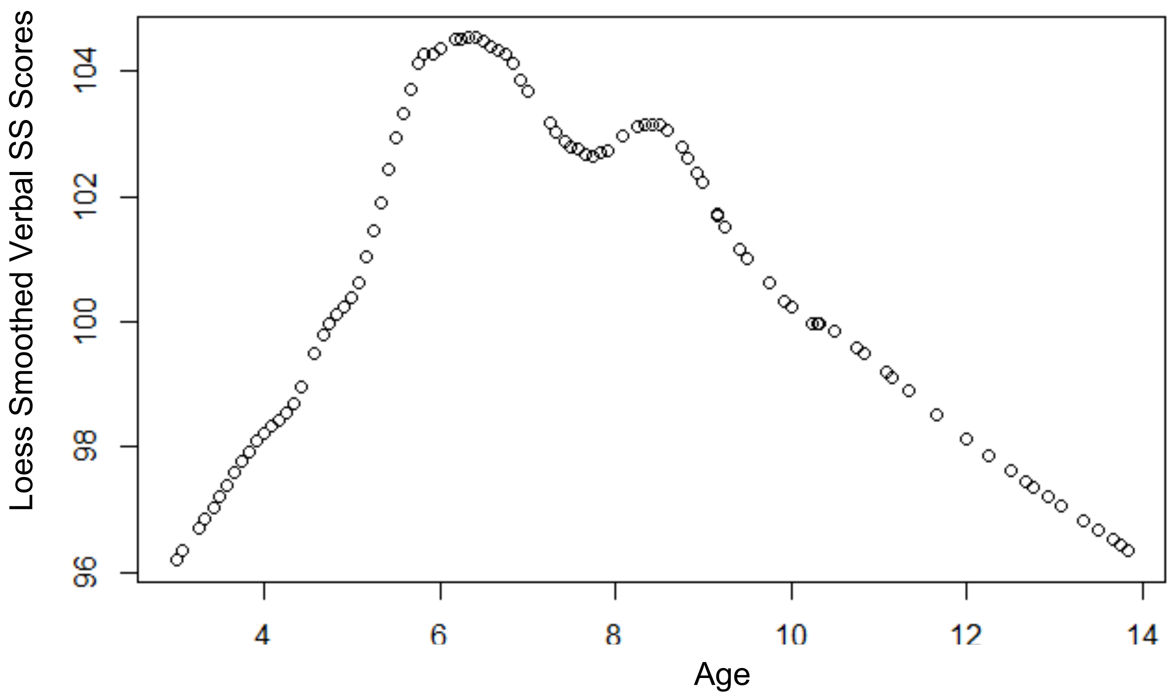
Curve of Lowess Smoothed GCA SS Scores of Individual Observations



Note. Data points are individual observations not grouped by participants. Lowess smoothing with smoother span = 1/3, "robustifying" iterations = 5, delta = .01 * range of x.

Figure 4

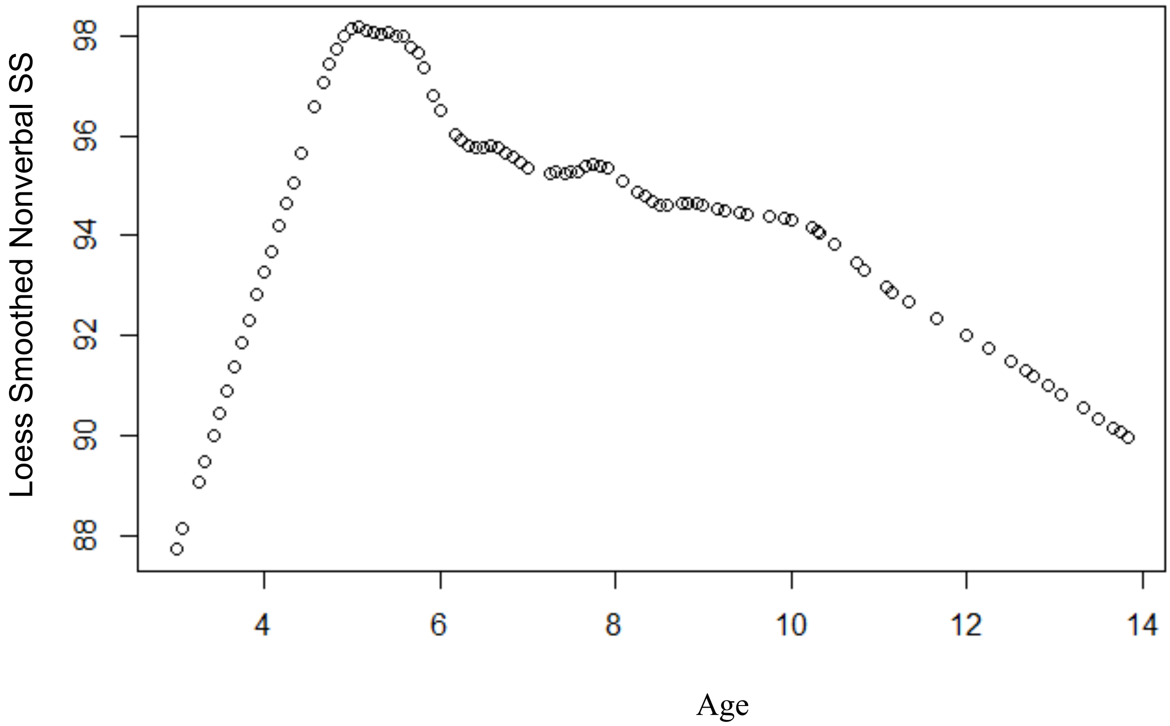
Curve of Lowess Smoothed Verbal SS Scores of Individual Observations



Note. Data points are individual observations not grouped by participants. Lowess smoothing with smoother span = 1/3, "robustifying" iterations = 5, delta = .01 * range of x.

Figure 5

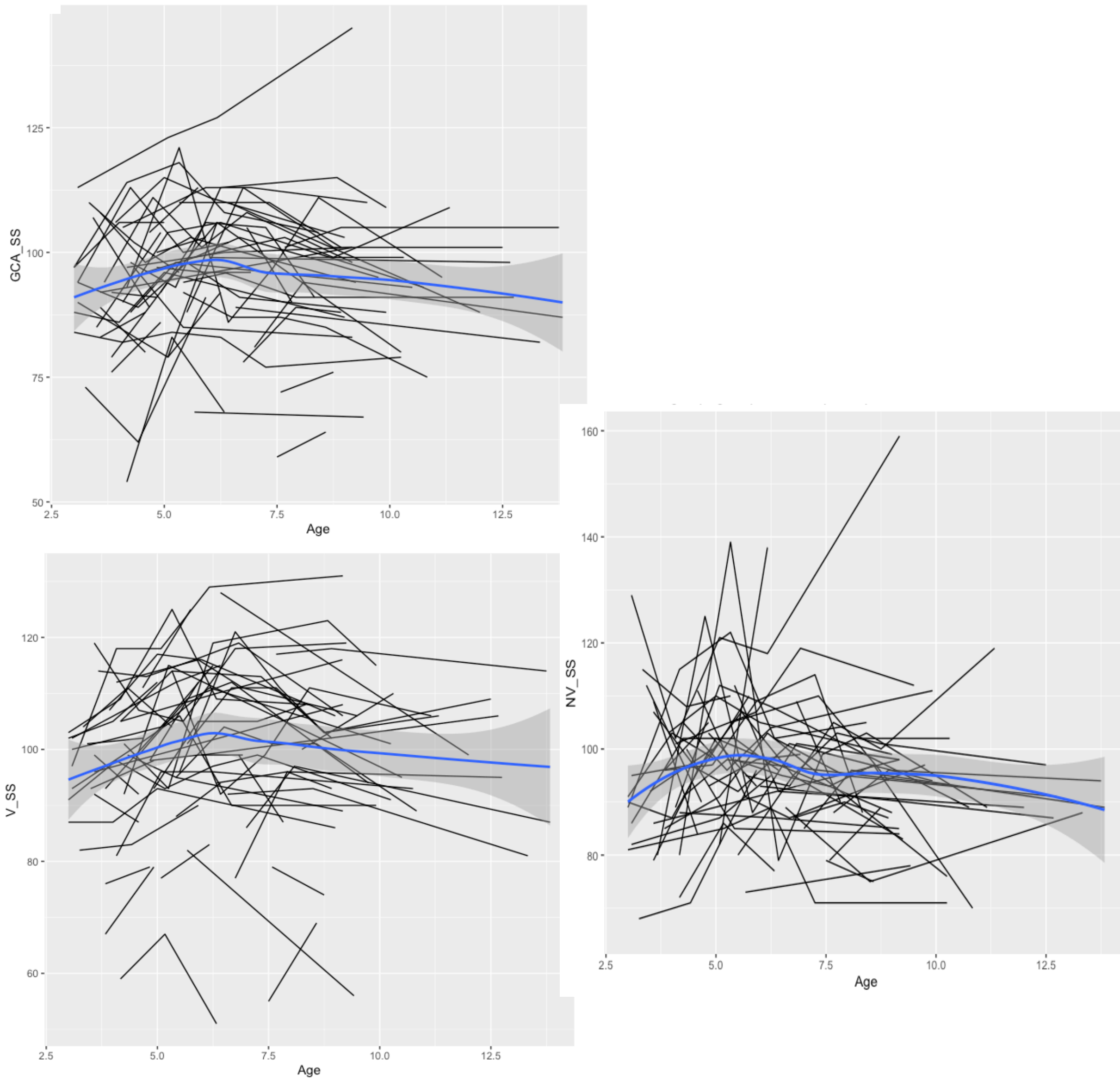
Curve of Lowess Smoothed Nonverbal SS Scores of Individual Observations



Note. Data points are individual observations not grouped by participants. Lowess smoothing with smoother span = 1/3, "robustifying" iterations = 5, delta = .01 * range of x.

Figure 6

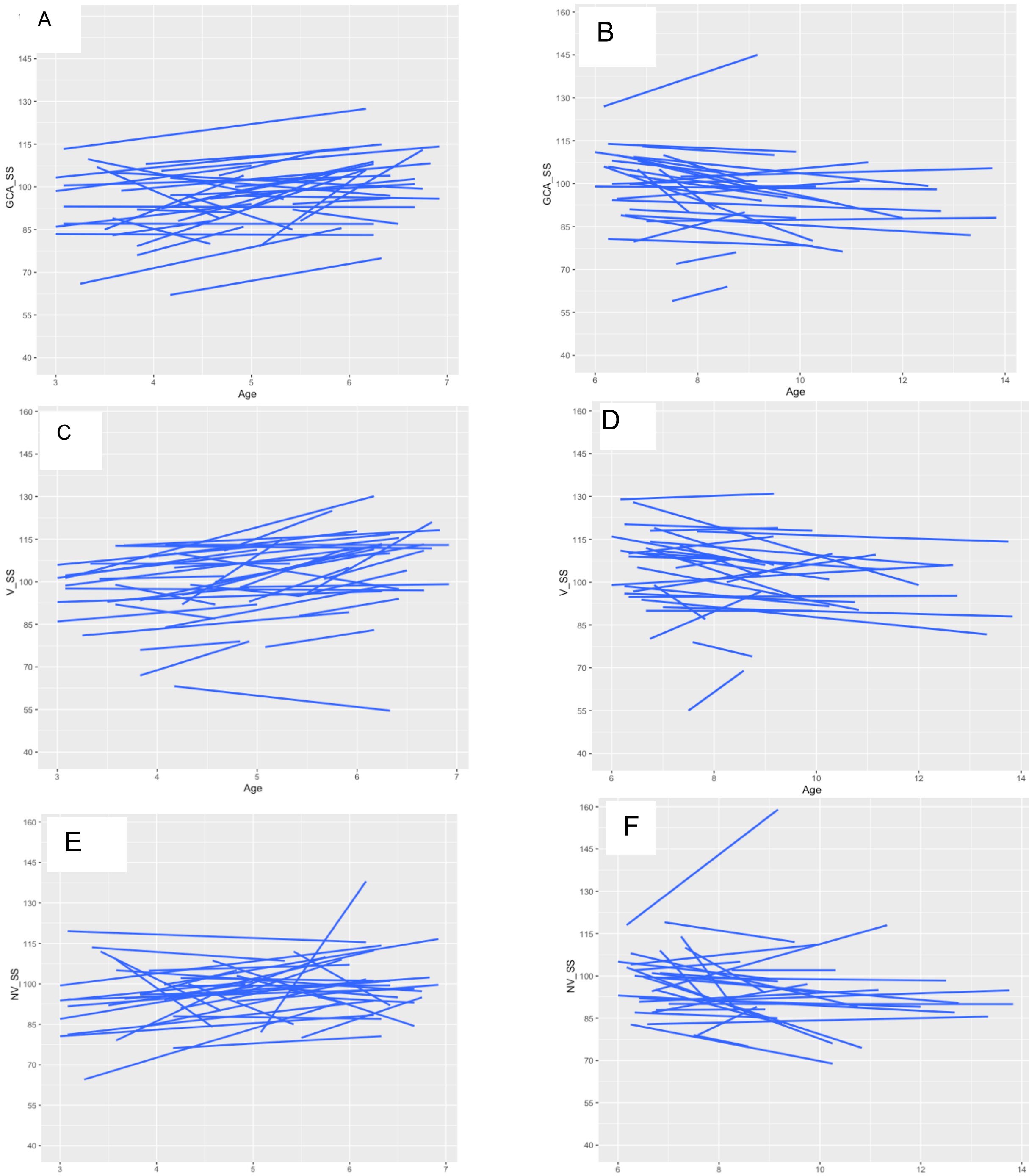
Grouped (by Participant) Spaghetti Plots of GCA SS, Verbal SS, Nonverbal SS Scores



Note. Blue line is Lowess smoothed mean curve. Panel A depicts GCA SS scores, Panel B depicts Verbal SS scores, and Panel C depicts Nonverbal SS scores.

Figure 7

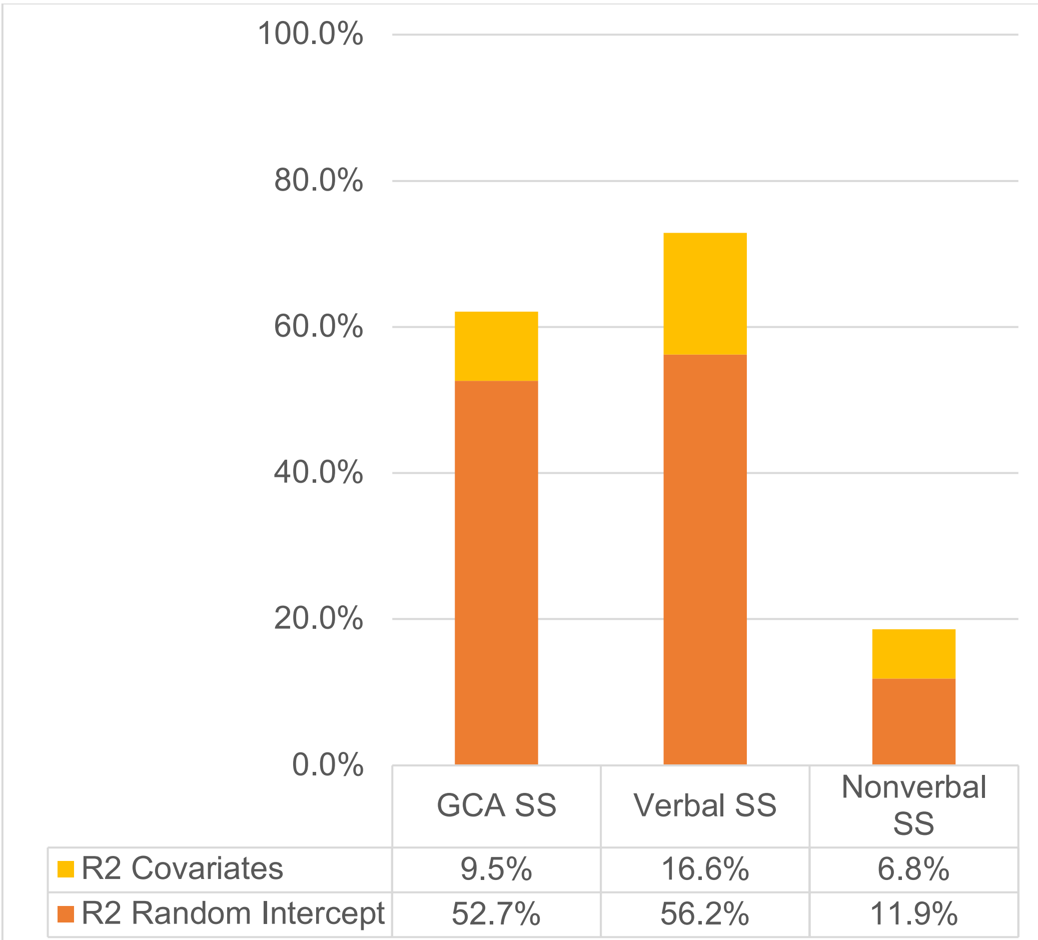
Individually Fitted Linear Models for GCA SS, Verbal SS and Nonverbal SS scores for Preschool and School-age Years Separately



Note. Panels on left show linear trends from Ages 3–6, panels on the right show linear trends Ages 6–13, Panels A and B show GCA SS linear trends, Panels C and D show Verbal SS linear trends and Panels E and F show Nonverbal SS linear trends.

Figure 8

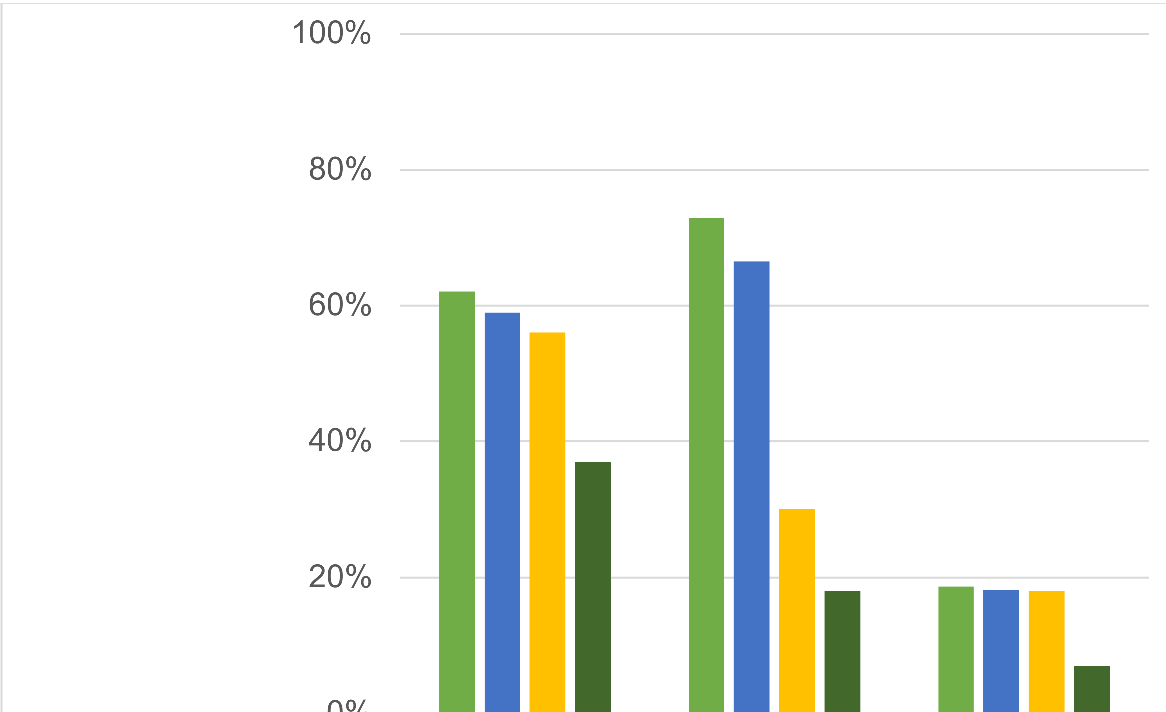
Linear Mixed Model Effect Sizes for Model Components by Criterion Variable



Note. Results from models without influential values (3 outliers removed for GCA models, 4 outliers for Verbal models and 6 for Nonverbal models). Nonverbal scores were log-transformed because residuals were not homogenous. "R2 Random Intercept" reflects $R^2_t{}^{(m)}$ (the proportion of the total variance accounted for by random intercept) in the final M4a model. "R2 Covariates" reflects $R^2_t{}^{(f)}$ which is the proportion of total variance accounted for by fixed effects (age and all other covariates in M4a) in the M4a models.

Figure 9

Comparison of Effect Sizes of Bivariate Correlations and Linear Mixed Models by Criterion Variable



	GCA	Verbal	Nonverbal
Final M4a model	62%	73%	19%
Empty Model M1	59%	66%	18%
correlation 3-4 to 9-13	56%	30%	18%
correlation 5-6 to 9-13	37%	18%	7%

Note. Results from models without influential values for both M1 and M4 (3 outliers removed for GCA models, 4 outliers for Verbal models and 6 for Nonverbal models), Nonverbal scores were log-transformed because residuals were not homogenous., Model M1 is the unconditional means model with the only parameter as the random intercept, and Nonverbal models demonstrated many values (>10) that had influential values, and depending on how many influential values were removed, total R varied from 19-40% see Table 12. "Final M4a" is the $R^2{}_t{}^{(fvm)}$ of the final M4a linear mixed models which reflects the proportion of variance accounted for by the entire final Model, "Empty Model M1" is the intra-class correlation coefficient of M1 and "Correlation 3–4 to 9–13" and "Correlation 5–6 to 9–13" reflect bivariate correlation coefficient based R^2 values.

References

Acosta, M. T., Gioia, G. A., & Silva, A. J. (2006). Neurofibromatosis type 1: New insights into neurocognitive issues. *Current Neurology and Neuroscience Reports, 6*(2), 136-143. https://doi.org/10.1007/s11910-996-0036-5

Baker, C. T., Sontag, L. W., & Nelson, V. L. (1958). Individual and group differences in the longitudinal measurement of change in mental ability. *Monographs of the Society for Research in Child Development, 23*(2), 11-85.

Bayley, N. (1949). Consistency and variability in the growth of intelligence from birth to eighteen years. *The Pedagogical Seminary and Journal of Genetic Psychology, 75*(2), 165-196. https://doi.org/http://dx.doi.org/10.1080/08856559.1949.10533516

Brothers, K. B., Glascoe, F. P., & Robertshaw, N. S. (2008). PEDS: Developmental milestones—an accurate brief tool for surveillance and screening. *Clinical Pediatrics, 47*(3), 271-279. https://doi.org/10.1177/0009922807309419

Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7*(1), 41-52. https://doi.org/10.1037/0894-4105.7.1.41

Conners, C. K. (1997). *Conners' parent rating scale revised*. Multi- Health Systems.

Coutinho, V., Kemlin, I., Dorison, N., de Villemeur, T. B., Rodriguez, D., & Dellatolas, G. (2016). Neuropsychological evaluation and parental assessment of behavioral and motor difficulties in children with neurofibromatosis type 1. *Research in Developmental Disabilities, 48*, 220-230. https://doi.org/10.1016/j.ridd.2015.11.010

Crockett, B. K., Rardin, M. W., & Pasewark, R. A. (1975). Relationship between WPPSI and stanford-binet IQs and subsequent WISC IQs in headstart children. *Journal of Consulting and Clinical Psychology, 43*(6), 922. https://doi.org/10.1037/0022-006X.43.6.922Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121-136. https://doi.org/10.1080/15248371003699969

Cutting, L. E., Huang, G., Zeger, S., Koth, C. W., Thompson, R. E., & Denckla, M. B. (2002). Growth curve analyses of neuropsychological profiles in children with neurofibromatosis type 1: Specific cognitive tests remain 'spared' and 'impaired' over time. *Journal of the International Neuropsychological Society, 8*(6), 838-46. https://doi.org/10.1017/S135561770286012X

Davis, O. S., Haworth, C. M., & Plomin, R. (2009). Dramatic increase in heritability of cognitive development from early to middle childhood: An 8-year longitudinal study of 8,700 pairs of twins. *Psychological Science, 20*(10), 1301-1308. https://doi.org/10.1111/j.1467-9280.2009.02433.x

DeRight, J., & Carone, D. A. (2015). Assessment of effort in children: A systematic review. *Child Neuropsychology, 21*(1), 1-24. https://doi.org/10.1080/09297049.2013.864383

Dovis, S., Van der Oord, S., Wiers, R. W., & Prins, P. J. (2012). Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? the effects of money and computer-gaming. *Journal of Abnormal Child Psychology, 40*(5), 669-681. https://doi.org/10.1007/s10802-011-9601-8

Elliott, C., D. (2007). *Differential ability scales second edition introductory and technical handbook*. PsychCorp.

Erdoğan-Bakar, E., Cinbiş, M., Ozyürek, H., Kiriş, N., Altunbaşak, S., & Anlar, B. (2009). Cognitive functions in neurofibromatosis type 1 patients and unaffected siblings. *The Turkish Journal of Pediatrics, 51*(6), 565-571.

Ferrari, M., & Sternberg, R. J. (1998). The development of mental abilities and styles. In W. Damon (Ed.), *Handbook of child psychology: Vol. 2. cognition, perception, and language* (pp. 899-946). John Wiley & Sons.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). 8 linear mixed effects models. *Applied longitudinal analysis* (2nd ed., pp. 189-240). John Wiley & Sons.

Foley, M., McClowry, S. G., & Castellanos, F. X. (2008). The relationship between attention deficit hyperactivity disorder and child temperament. *Journal of Applied Developmental Psychology, 29*(2), 157-169. https://doi.org/10.1016/j.appdev.2007.12.005

Gardner, M. K., & Clark, E. (1992). In Sternberg R. J., Berg C. A.(Eds.), *The psychometric perspective on intellectual development in childhood and adolescence*. Cambridge University Press, New York, NY.

Hindley, C. B., & Owen, C. F. (1978). The extent of individual changes in I.Q. for ages between 6 months and 17 years, in a british longitudinal sample. *Child Psychology & Psychiatry & Allied Disciplines, 19*(4), 329-350. https://doi.org/10.1111/j.1469-7610.1978.tb00480.x

Hollingshead, A. B. (1975). *Four factor index of social status.* Unpublished manuscript.

Honzik, M. P., Macfarlane, J. W., & Allen, L. (1948). The stability of mental test performance between two and eighteen years. *The Journal of Experimental Education, 17*(2), 309-324. https://doi.org/10.1080/00220973.1948.11010388

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

2005Hyman, S. L., Shores, A., & North, K. N. (2006). Learning disabilities in children with neurofibromatosis type 1: Subtypes, cognitive profile, and attention-deficit-hyperactivity disorder. *Developmental Medicine & Child Neurology, 48*(12), 973-977. https://doi.org/10.1111/j.1469-8749.2006.tb01268.x

Jacobson, N. S., & Truax, P. (1992). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19. https://doi.org/10.1037/0022-006X.59.1.12

Klein-Tasman, B. P., Janke, K. M., Luo, W., Casnar, C. L., Hunter, S. J., Tonsgard, J., Trapane, P., van der Fluit, F., & Kais, L. A. (2014). Cognitive and psychosocial phenotype of young children with neurofibromatosis-1. *Journal of the International Neuropsychological Society, 20*(1), 88-98. 10.1017/S1355617713001227

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

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education, 6*(1), 8. https://doi.org/10.1186/s40536-018-0061-2

Lorenzo, J., Barton, B., Arnold, S. S., & North, K. N. (2013). Cognitive features that distinguish preschool-age children with neurofibromatosis type 1 from their peers: A matched case-control study. *The Journal of Pediatrics, 163*(5), 1479-1483. https://doi.org/10.1016/j.jpeds.2013.06.038

Lorenzo, J., Barton, B., Arnold, S. S., & North, K. N. (2015). Developmental trajectories of young children with neurofibromatosis type 1: A longitudinal study from 21 to 40 months of age. *The Journal of Pediatrics, 166*(4), 1006-1012. https://doi.org/10.1016/j.jpeds.2014.12.012

Loy, A., & Hofmann, H. (2014). HLMdiag: A suite of diagnostics for hierarchical linear models in R. *Journal of Statistical Software, 56*(5), 1-28. http://hdl.handle.net/10.18637/jss.v056.i05

McCall, R. B., Appelbaum, M. I., & Hogarty, P. S. (1973). Developmental changes in mental performance. *Monographs of the Society for Research in Child Development, 38*(3), 1-84. https://doi.org/10.2307/1165768

North, K. (2000). Neurofibromatosis type 1. *American Journal of Medical Genetics, 97*(2), 119-127. https://doi.org/10.1002/1096-8628(200022)97:2<119::AID-AJMG3>3.0.CO;2-3

North, K., Joy, P., Yuille, D., Cocks, N., & Hutchins, P. (1995). Cognitive function and academic performance in children with neurofibromatosis type 1. *Developmental Medicine & Child Neurology, 37*(5), 427-36. https://doi.org/10.1111/j.1469-8749.1995.tb12026.x

Nupan, M. M. T., Van Meerbeke, A. V., Cabra, C. A. L., & Gomez, P. M. H. (2017). Cognitive and behavioral disorders in children with neurofibromatosis type 1. *Frontiers in Pediatrics, 5*, 227. https://doi.org/10.3389/fped.2017.00227

Payne, J. M., Pickering, T., Porter, M., Oates, E. C., Walia, N., Prelog, K., & North, K. N. (2014). Longitudinal assessment of cognition and T2 ‐ hyperintensities in NF1: An 18 ‐ year study. *American Journal of Medical Genetics Part A, 164*(3), 661-665. https://doi.org/10.1002/ajmg.a.36338

Ribeiro, M. J., dâ€™Almeida, O. C., Ramos, F., Saraiva, J., Silva, E. D., & Castelo-Branco, M. (2014). Abnormal late visual responses and alpha oscillations in neurofibromatosis type 1: A link to visual and attention deficits. *Journal of Neurodevelopmental Disorders, 6*(1), 4. https://doi.org/10.1186/1866-1955-6-4

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods, 24*(3), 309-338. https://doi.org/10.1037/met0000184

Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development, 64*(1), 80-97. 10.1111/j.1467-8624.1993.tb02896.x.

Sangster, J., Shores, E. A., Watt, S., & North, K. N. (2010). The cognitive profile of preschool-aged children with neurofibromatosis type 1. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence, 17*(1), 1-16. https://doi.org/10.1080/09297041003761993

Sansavini, A., Pentimonti, J., Justice, L., Guarini, A., Savini, S., Alessandroni, R., & Faldella, G. (2014). Language, motor and cognitive development of extremely preterm children: Modeling individual growth trajectories over the first three years of life. *Journal of Communication Disorders, 49*, 55-68. https://doi.org/10.1016/j.jcomdis.2014.02.005

Scheider, W., Perner, J., Bullock, M., Stefanek, J., & Ziegler, A. (1999). Development of intelligence and thinking. In F. E. Weinert, & W. Schneider (Eds.), *Individual development from 3 to 12 findings from the munich longitudinal study* (pp. 9-27). Cambridge University Press.

Schneider, W., & Bullock, M. (2010). Development of intelligence and thinking. *Human development from early childhood to early adulthood* (pp. 17-44). Psychology Press.

Scime, M., & Norvilitis, J. M. (2006). Task performance and response to frustration in children with attention deficit hyperactivity disorder. *Psychology in the Schools, 43*(3), 377-386. https://doi.org/10.1002/pits.20151

Singer, J., & Willet, J. (2003). *Exploring longitudinal data on change. applied longitudinal data analysis*. New York: Oxford University Press, Inc.

Tabachnick, B., & Fidell, L. (2013a). 15 multilevel linear modeling. *Using multivariate statistics* (6th ed., pp. 786-861). Pearson.

Tabachnick, B., & Fidell, L. (2013b). 5 multiple regression. *Using multivariate statistics* (6th ed., pp. 117-196). Pearson.

Wessel, L. E., Gao, F., Gutmann, D. H., & Dunn, C. M. (2012). Longitudinal analysis of developmental delays in children with neurofibromatosis type 1. *Journal of Child Neurology, 28*(12), 1689-1693. https://doi.org/10.1177/0883073812462885

West, B. T., Welch, K. B., & Galecki, A. T. (2014a). 2 linear mixed models: An overview. *Linear mixed models: A practical guide using statistical software* (2nd ed., pp. 9-58). Taylor & Francis.

West, B. T., Welch, K. B., & Galecki, A. T. (2014b). *Linear mixed models: A practical guide using statistical software* (2nd ed.). Taylor & Francis.

Wilson, R. S. (1974). Twins: Mental development in the preschool years. *Developmental Psychology, 10*(4), 580-588. https://doi.org/10.1037/h0036596

CURRICULUM VITAE

Gregor Nathanael Schwarz

EDUCATION

PhD, University of Wisconsin-Milwaukee (Expected August 2020)
Clinical Psychology (APA-accredited training program)
Advisor: Bonita Klein-Tasman
Cumulative GPA: 3.91
Preliminary exam passed: 6/13/2018
Dissertation proposed: 9/7/2018
Dissertation defended: 7/14/2020Dissertation title: *Stability and Predictive Value of Intellectual Functioning in Neurofibromatosis Type 1 Beginning in the Preschool Years*

MS, University of Wisconsin-Milwaukee (May 2016)
Clinical Psychology (APA-accredited training program)
Thesis proposed: 12/03/2015
Thesis defended: 03/04/2016
Thesis title: *Relations Between Lab-Based and Parent-Reported Executive Functioning in Children and Adolescents with Williams Syndrome*

BA, University of Virginia, Awarded with Highest Distinction (May 2013)
Major: Psychology
Grade Point Average: 3.73

AA, Piedmont Virginia Community College (May 2011)
Program of Study: Liberal Arts
Grade Point Average: 3.90

HONORS and AWARDS

The Phi Beta Kappa Society	June 2013
UWM Chancellor's Graduate Student Award	August 2013
Williams Syndrome Association Travel Award	July 2014

PROFESSIONAL AFFILIATIONS

American Psychological Association
Association for Psychological Science

CLINICAL EXPERIENCE

08/19 to 8/20

Pre-Doctoral Psychology Intern
Rogers Memorial Health; Child Adolescent Day Treatment and Partial Hospitalization Program (West Allis, WI)
Supervisors: Nancy J. Goranson, Psy.D., Kristin Miles, Psy.D.

Training Experience:

- Intervention: As member of the Early Adolescent DBT Partial Hospitalization Program (EAPHP) team and Adolescent DBT Intensive Outpatient Program provided case management, individual therapy, group therapy to 12–18-year-old patients with high suicide risk. Interventions include dialectical behavior therapy and cognitive-behavioral therapy.
- Assessment (10–18 year-olds): conducted therapy progress assessments for pre-adolescents and adolescents in the DBT Intensive Outpatient Programs, conducted psychological assessments including psychological testing (WISC-V, WIAT-III), report writing and feedback sessions, conducted social services intake assessments, observed psychological diagnostic evaluation interviews
- Participated in consultation team meetings, staffing meetings, supervision of supervision meetings and peer group supervision meetings. Attended weekly internship didactics.
- Engaged in program development by assisting Early Adolescent PHP team to revise programming handouts, diary cards, check in/check out cards, and participated in and lead DBT consultation team meetings for the EAPHP team.
  - Conducted psychological diagnostic evaluations (at intake) for entire CADT/PHP unit, provided crisis intervention and consultation for entire CADT/PHP unit and shadowed attending psychologist.

Therapy Practicum Student 06/18 to 06/19
Center for Behavioral Medicine (Brookfield, WI)
Supervisor: Henry Boeh, PhD

Training Experience:

- Providing adult outpatient dialectical behavior therapy (individual and group therapy; 1x advanced/stage II group) and phone coaching. Participating in adult consultation team meetings.
- Assist and lead adolescent dialectical behavior therapy stage 1 groups, participating in adolescent consultation team meetings, and providing phone coaching to parents of adolescents.

Therapy Practicum Student 05/17 to 05/18
Center for Behavioral Medicine (Brookfield, WI)
Supervisor: Kim Skerven, PhD

Training Experience:

- Provided comprehensive individual and group adult outpatient dialectical behavior therapy (1x stage I group, 1x advanced/ stage II group) and phone coaching. Participated in adult consultation team meetings.
- Conducted comprehensive psycho-diagnostic assessment (SCID-I and SCID-II) and provided individual skills training.
- Participated in adolescent consultation team meetings and provided phone coaching to parents of adolescents.

Therapy Practicum Student 08/16 to 07/17
Rogers Memorial Hospital; Day Treatment Unit (West Allis, WI)
Supervisor: Nancy J. Goranson, PsyD

Training Experience:
- Provided child and adolescent intensive outpatient therapy, primarily with 8–13-year-old clients.
- Assisted and led group therapy, conducted individual therapy sessions, observed suicide risk assessments, and participated in consultation team meetings and peer group supervision meetings.
- Interventions included cognitive-behavioral therapy and dialectical behavior therapy-informed skills training.

Therapy Practicum Student 08/16 to 06/19
University of Wisconsin-Milwaukee; Psychology Clinic
Supervisor: Shawn P. Cahill, PhD

Training Experience:
- Providing adult outpatient therapy for anxiety disorders, depression, and ADHD.
- Conducting intake assessments and individual therapy sessions.
- Interventions include cognitive-behavioral therapy and dialectical behavior therapy skills.

Therapy Practicum Student 09/15 to 08/16
University of Wisconsin-Milwaukee; Psychology Clinic
Supervisors: Bonita P. Klein-Tasman, PhD; Christopher Martell, PhD

Training Experience:
- Provided adolescent and adult outpatient therapy for anxiety disorders.
- Conducted intake assessments, individual therapy sessions, and family therapy sessions.
- Interventions included cognitive-behavioral therapy.

Graduate Assistant 08/14 to 06/19
University of Wisconsin-Milwaukee; Child Neurodevelopment Research Lab
Supervisor: Bonita P. Klein-Tasman, PhD

Training Experience:
- Psychoeducational assessment of young children with NF1 and adolescents with WS, scoring, and report writing.

Assessment Practicum Student 08/14 to 05/15
University of Wisconsin-Milwaukee; Psychology Clinic
Supervisors: Kristin D. Smith, PhD; Han Joo Lee, PhD

Training Experience:
- Psychoeducational assessments and psychosocial/diagnostic interviews with child, adolescent, and adult clients, scoring, and report writing

First Year Practicum Student 08/13 to 05/14
University of Wisconsin-Milwaukee
Supervisor: Bonita P. Klein-Tasman, PhD

Training Experience:

- Practice in unstructured and structured interviewing and test administration; conducted primarily on peers and volunteers.

Residential Caregiver 10/06 to 03/12
Innisfree Inc. (Charlottesville, VA)

Experience:

- Cared for adults with intellectual disabilities, including individuals on the autism spectrum and with Down syndrome.
- Worked in therapeutic workshops with adults with intellectual disabilities, including individuals on the autism spectrum and with Down syndrome.
- Employed full-time from 10/06 to 07/10

RESEARCH EXPERIENCE

Graduate Research Assistant (Multiple Studies) 06/13 to present
University of Wisconsin-Milwaukee; Child Neurodevelopment Research Lab
Faculty Advisor: Bonita Klein-Tasman, PhD

Duties:

- Administer DAS, WIAT, NEPSY, Purdue Pegboard, and KBIT-2 tests to participants, prepare data for analysis, conduct, and assist with data analyses, and assist in writing manuscripts.
- Conduct parent interviews (MINI, SIB-R, KDBDS) prepare data for analysis, conduct and assist with data analyses, and assist in writing manuscripts.
- Assist with pilot data collection and development of treatment manuals
- Participants include young children (3 to 6 years) with NF1 and children and adolescents with Williams Syndrome.

Undergraduate Research Assistant and Honors Thesis 08/11 to 04/13
University of Virginia; Early Development Lab
Faculty Advisor: Angeline S. Lillard, PhD
Co-investigator: Matthew Lerner, MA

Duties:

- Implemented study design, collected heart rate and electrodermal skin conductance values (Biopac system), analyzed data, authored thesis manuscript (APA style).
- Recruited, scheduled, and ran participants through the two-part, 1.5-hour study with free play sessions.
- Administered ToM Assessment Scale, KBIT-2, and Strange Stories (advanced ToM measure)
- Coded participant responses (ToM), coded play, and entered data into SPSS.
- Participants included 3–8-year-old typically developing children and one child with ASD.

Undergraduate Research Assistant 05/12 to 01/13

University of Virginia; Virginia Affective Neuroscience Lab
Faculty Advisor: James Coan, PhD

Duties:

- Recruited, phone-screened, and scheduled participants
- Ran 2.5-hour study, including administration of electric shock, Stroop task, and questionnaires, installation of ERP caps, collection of ERP data, and entry of questionnaire data.
- Trained new research assistants on all study-related responsibilities
- Participants: couples, opposite sex friends

PEER-REVIEWED PUBLICATIONS

Brei, N. G., Klein-Tasman, B. P., **Schwarz, G. N.**, & Casnar, C. L. (2014). Language in young children with neurofibromatosis-1: Relations to functional communication, attention, and social functioning. *Research in Developmental Disabilities*, 35, 2495–2504.

Brei, N. G., **Schwarz, G. N.**, & Klein-Tasman, B. P. (2015). Predictors of parenting stress in children referred for an Autism Spectrum Disorder diagnostic evaluation. *Journal of Developmental and Physical Disabilities,* 27(5), 617–635.

BOOK CHAPTERS

Janke, K., **Schwarz, G. N.**, & Klein-Tasman, B. P. (2016). Mental Health in Developmental Disabilities. In H. Friedman (Ed.), *Encyclopedia of Mental Health* (2nd ed.). Waltham, MA: Academic Press.

POSTER PRESENTATIONS

National Meetings

Schwarz, G. N., Casnar, C. L., Yund, B. D., Lee, K. M., Brei, N. G., & Klein-Tasman, B. P. (2019, April). *Stability and Predictive Value of Intellectual Functioning in Neurofibromatosis Type 1 Beginning in the Preschool Years.* Poster abstract submitted for the 52nd Annual Gatlinburg Conference, San Antonio, TX.

Lee, K. M., Yund, B. D., **Schwarz, G. N.**, Glad, D., & Klein-Tasman, B. P. (2019, February). *Longitudinal Examination of Problem Behaviors in Children with Neurofibromatosis Type 1.* Poster submitted for the Annual Meeting of the International Neuropsychological Society, New York, NY.

Kirkman, M., Davine, T., **Schwarz, G. N.**, Lee, C. B., Wessels, K., & Skerven, K. (2018, November). *Skill Use Moderates the Relationship Between Emotion Dysregulation and Borderline Symptoms.* Poster to be presented at the Annual Conference of the International Society for the Improvement and Teaching of Dialectical Behavior Therapy, Washington, D.C.

Schwarz, G. N., Bennett, D., Mervis, C. B., & Klein-Tasman, B. P. (2017, March). *Relations Between Brief Parent Ratings and DCCS Lab-based Performance for Youth with Williams Syndrome.* Poster presented at the 50th Annual Gatlinburg Conference, San Antonio, TX.

Schwarz, G. N., Smith, K. D., Bennett, D., Mervis, C. B., & Klein-Tasman, B. P. (2016, March). *Gender Differences in Relations Between Emotion Regulation, Inhibition, and Adaptive*

Functioning in Children and Adolescents with Williams Syndrome. Poster presented at the 49th Annual Gatlinburg Conference, San Diego, CA.

Schwarz, G. N., Bennett, D., Mervis, C. B., & Klein-Tasman, B. P**.** (2015, February). *Relations between lab-based and parent-reported executive functioning in children and adolescents with Williams syndrome.* Poster presented at the annual meeting of the International Neuropsychological Society, Denver, CO.

Smith, K. D., **Schwarz, G. N.**, Mervis, C. B., Bennett, D., & Klein-Tasman, B. P. (2014, June). *Relations Between Emotion Regulation and Adaptive Functioning in Children and Adolescents with Williams Syndrome.* Poster presented at the bi-annual meeting of the Williams Syndrome Association, Garden Grove, CA.

Regional Meetings

Miller, A. A., **Schwarz, G. N.**, Brei, N. G., Lee, H., & Klein-Tasman, B. P. (2017, April). *Reliability of a Lab-Based Measure of Response Inhibition in Youth with Williams Syndrome.* Poster presented at the 31st National Conference on Undergraduate Research, Memphis, TN.

Hayward, A. M., Hermsmeier Fiscus, E., Brei, N. G., Lee, H. J., **Schwarz, G. N.**, & Klein-Tasman, B. P. (2016, March). *Online Response Inhibition Training for Children with Williams Syndrome: Patterns of Practice Performance.* Poster presented at the 30th National Conference on Undergraduate Research, Asheville, NC.

Raicu, A., Lesch, L. M., Brei, N. G., Lee, H. J, **Schwarz, G. N.**, & Klein-Tasman, B. P. (2016, March). *Acceptability of Web-Based Response Inhibition Training in Children with Williams Syndrome*. Poster presented at the 30th National Conference on Undergraduate Research, Asheville, NC.

Rivera, K. M., Helms, M. I., Casnar, C. L., Brei, N. G., **Schwarz, G. N.**, & Klein-Tasman, B. P. (2015, April). *Examining Psychosocial Functioning in Young Children with Neurofibromatosis Type-1 Using the BASC-II Content Scales*. Poster presented at the 29th National Conference on Undergraduate Research, Cheney, WA.

Wilson, A. E., Basche, K. E., Rivera, K. M., Brei, N. G., **Schwarz, G. N.**, Lee, H., & Klein-Tasman, B. P. (2015, April). *A Pilot Study of Response Inhibition in Adolescents with Williams Syndrome Using a Go/No-Go Task*. Poster presented at UWM's 7th Annual Undergraduate Research Symposium, Milwaukee, WI.

Bauer, K. A., **Schwarz, G. N.**, & Klein-Tasman, B. P. (2014, April). *Relations of Parent-reported Attention Problems to a Lab-Based Inhibition Task in Young Children with Neurofibromatosis-I*. Poster presented at the 28th National Conference for Undergraduate Research Symposium, Milwaukee, WI.

Rivera, K. M., Helms, M. I., Casnar, C. L., Brei, N. G., **Schwarz, G. N.**, & Klein-Tasman, B. P. (2014, April). *Examination of BASC-II Content Scales in Young Children with Neurofibromatosis-1.* Poster presented at the UWM's 6th Annual Undergraduate Research Symposium, Milwaukee, WI.

Anglin, T., Basche, K., Casnar, C. L., Brei, N. G., **Schwarz, G. N.**, & Klein-Tasman, B. P. (2014, April). *Social Skills of Young Children with NF1: Relations to Attention Problems and Cognitive Functioning.* Poster presented at UWM's 6th Annual Undergraduate Research Symposium, Milwaukee, WI.

Schwarz, G. N. (2013, May). *Play as a Potential Regulator of Emotion.* Thesis presented at the University of Virginia L. Starling Reid Undergraduate Psychology Conference, Charlottesville, VA.

TEACHING EXPERIENCE

Independent Instructor & Graduate Teaching Assistant Spring 2019
Psychology 260 - Child Psychology (online)
University of Wisconsin-Milwaukee

Independent Instructor & Graduate Teaching Assistant Fall 2018
Psychology 260 - Child Psychology (online)
University of Wisconsin-Milwaukee

Independent Instructor & Graduate Teaching Assistant Spring 2018
Psychology 260 - Child Psychology (online)
University of Wisconsin-Milwaukee

Independent Instructor & Graduate Teaching Assistant Fall 2017
Psychology 260 - Child Psychology (online)
University of Wisconsin-Milwaukee

Independent Instructor & Graduate Teaching Assistant Spring 2017
Psychology 260 - Child Psychology (online)
University of Wisconsin-Milwaukee

Independent Instructor & Graduate Teaching Assistant Fall 2016
Psychology 260 - Child Psychology (online)
University of Wisconsin-Milwaukee

Graduate Teaching Assistant Spring 2015
Psychology 260 - Child Psychology (online)
University of Wisconsin-Milwaukee
Instructor: Christina Casnar

Graduate Teaching Assistant Fall 2014
Psychology 260 - Child Psychology (online)
University of Wisconsin-Milwaukee
Instructor: Dr. Kristin Smith

Graduate Teaching Assistant Spring 2014
Psychology 260 - Child Psychology
University of Wisconsin-Milwaukee
Instructor: Dr. Kristin Smith

Graduate Teaching Assistant Fall 2013
Psychology 260 - Child Psychology

University of Wisconsin-Milwaukee
Instructor: Dr. Kristin Smith

OTHER RELATED SKILLS and ACHIEVEMENTS

- Professional-level proficiency in all Microsoft Office Software
- Program proficiency in SPSS and R
- Basic structural equation modeling skills
- Basic Linear Mixed Modeling skills

SEMINARS, SPECIALIZED TRAINING, and CERTIFICATIONS

Introduction to R: Data Management and Analysis Fall 2015
University of Wisconsin-Milwaukee
Workshop series on introduction to R statistical software
Lecturer: David Armstrong, PhD

Behavior Activation for Depression Spring 2016
University of Wisconsin-Milwaukee
Day workshop on behavioral activation as treatment for depression
Lecturer: Christopher Martell, PhD, ABPP

Mental Health Services for Transgender Clients Spring 2017
Center for Behavioral Medicine (Brookfield, WI)
Workshop on mental health services for transgender clients
Lecturer: Gregory Simons, PhD

Eating Disorders Summer 2017
University of Wisconsin-Milwaukee
Graduate seminar on the assessment and treatment of eating disorders
Lecturer: Stacey Nye, PhD, FAED

Group Therapy Summer 2017
University of Wisconsin-Milwaukee
Graduate seminar on implementing group therapy
Lecturer: Stacey Nye, PhD, FAED

Introduction to Dialectical Behavior Therapy Summer 2017
Center of Behavioral Medicine (Brookfield, WI)
Workshop series introducing Dialectical Behavior Therapy
Lecturers: Kim Skerven, PhD, Henry Boeh, PhD, Neal Moglowsky, LPC, Mary-Catherine Nimphius, LPC, and Megan Schiferl, LPC

Issues in Mental Health Assessment and Service for Trans Students Fall 2017
University of Wisconsin-Milwaukee
Workshop on assessment and treatment of transgender students
Lecturer: Barry Schreier, PhD

Mindfulness Symposium: Harnessing Mindfulness – Fitting the Practice to the Person Summer 2018
Medical College of Wisconsin (Milwaukee, WI)
Day workshop on clinical decision making of when and how to use mindfulness practices in mental health treatment
Lecturer: Ronald Siegel, PsyD

Dialectical Behavior Therapy Foundations August 2019
Rogers Behavioral Health – Rogers University
Week-long 40-hour seminar about providing DBT in acute levels of care
Lecturers: Melisa Nelson, PhD, LPC, NCC; Erik Ulland, MD, Jolie Fritz, LPC-IT, SAC-IT, Nicole Nevaranta, Psy.D., Carissa Buchanan, M.S., Kristine Kim, Psy.D.