

Document Type: IRB Protocol

Official Study Title: Effectiveness and Predictors of Intervention Response for a
Technology-based Reading Intervention in the Home

NCT Number: NCT04568824

Unique Protocol ID: H20-0089

Document Date:10/25/2022

Neurocognitive Factors in EdTech Treatment Response

Site Principal Investigator: Michael Milham, MD, PhD

I. Protocol Summary

Background

Reading Disability (RD) is the most common learning disability, affecting 10 – 15% of school age children. It incurs major functional impairments at all stages of life. A wealth of data documents lifelong disadvantages in educational and occupational attainment. Problematically, current evidence-based reading interventions largely rely on services by trained specialists, either in well-resourced classrooms or clinical settings. As such, under-resourced schools (or countries) often are unable to provide reading interventions for their students. The significance of this dilemma is compounded when considering that children from lower socioeconomic status (SES), and children with other serious comorbid behavioral health conditions, may have more severe or complex RD profiles¹⁻³. Thus, the children most in need are the least likely to have access to evidence-based treatment.

Recent studies have supported the efficacy of technology-based reading interventions, particularly GraphoLearn, which focuses on teaching phonological skills and letter sound correspondence, though with notable variation in effect sizes. As highlighted in a recent meta-analysis, sample sizes tend to be small, and there is marked variation in key methodological components across studies, particularly gameplay dosage (which is well below expert recommendations in most studies), control group (e.g., active versus passive), and level of adult involvement in administration. Most relevant, the utility of GraphoLearn has not been adequately examined outside the confines of controlled school-settings, thus failing to inform on the value of home administration – the major positive feature of the GraphoLearn. Similarly, possible predictors of negative outcomes for home administration are unexplored (e.g., compliance with prescribed gameplay dosage, which, as noted in our preliminary data, can vary).

Significance

The proposed research is motivated by two significant overarching problems: (1) opportunities for reading intervention are not available to the majority of children in need of help^{4,5}, and (2) across children there is substantial variability in reading intervention gains⁶⁻⁸. Factors that predict intervention response remain poorly understood, with many potentially relevant factors (e.g., environment, mental health) understudied⁹. To address these issues, our proposed project leverages the availability of a large, diverse, and well characterized sample of reading disabled children. It aims to establish the effectiveness of GraphoLearn, an alternative, affordable technology-based, reading intervention administered in the home. The project also aims to identify multi-level predictors of intervention response.

The overarching tenets guiding this work are as follows:

Educational Technology (EdTech) is a potentially valuable way of increasing accessibility to treatment for those with limited resources. EdTech treatments, also called “serious games”, rely on computers,

tablets and smartphones to deliver interventions¹⁰. These treatments deliver educational content and interventions to individuals in situ (e.g., school, home), without the need for expert supervision. Over the past decade, there has been an explosion of EdTech development; however, there is a paucity of high-quality research on the effectiveness of these games¹¹. From a resource perspective, delivering services via an app in the home is ideal, as it removes dependencies on school supports, resources and infrastructures, which can vary substantially. However, the home setting presents unique challenges, as the lack of structure or monitoring by a trained provider may decrease the level of engagement and compliance^{5,12,13}.

Gamification has the potential to increase engagement and compliance with intervention. Engagement is a key challenge for the delivery of any remedial effort, let alone those being delivered in unstructured or unsupervised settings (e.g., the home). Gamification is a commonly cited means of improving engagement, as it can help motivate learners through the creation of a positive, reward-driven learning environment, in which feedback can be easily and immediately provided^{14,15}. The inclusion of appealing graphics is an additional benefit of gamification that can serve to increase engagement and increase the feasibility of at home interventions¹⁶.

While studies of GraphoLearn in laboratories, and controlled school settings, provide evidence of its efficacy, its performance when delivered in the home has received little attention to date. GraphoLearn, formerly known as GraphoGame¹⁷, fosters phonological awareness and teaches letter-sound correspondences. To our knowledge, it is the best studied serious reading game. It is a noncommercial product developed by reading researchers at the University of Jyväskylä and has been extensively studied in Finnish and more recently in English and several other languages¹⁸. Across studies, in multiple languages, GraphoLearn has been found to be superior to various other control conditions (e.g., music training, math games, typical classroom instruction)^{19,20}. However, effect sizes for reading achievement have been highly variable, from negligible to very large²¹. Further, the effectiveness of GraphoLearn, and the generalizability of the findings remain unclear due to several study limitations: (1) small sample sizes; (2) administration limited to supervised school environments rather than in the home (86% per recent meta-analysis); (3) relatively short durations of exposure (typically less than 15 hours, which is far below expert recommendations for conventional intervention), and (4) homogeneous, European samples, that tend to exclude individuals with psychiatric comorbidities. These factors have also limited the exploration of factors (person-specific and environmental) that may contribute to the effectiveness of GraphoLearn. Thus, real-world testing, with appropriately large samples, is essential to establishing the usefulness of GraphoLearn.

A range of individual-specific factors (e.g., behavioral, cognitive, mental health, brain, environment) may impact the effectiveness of reading interventions, with some having particular relevance to real-world settings (i.e., compliance). It is well recognized that not all children make significant gains following reading interventions, even under optimal treatment conditions⁶. What is less well established are the participant- and environment-level factors that contribute to response variability⁸. Relevant studies have found that both domain-specific factors (level of children's phonological awareness, rapid naming skill) and domain general factors (IQ, executive functioning skills) were significant positive predictors of

treatment response^{9, 22-25}. The few recent studies of environmental factors have also found significant, albeit inconsistent, relationships with SES; Some have reported that intervention gains for low SES children were equivalent to those of high SES children⁸, while others found that they were relatively better for low SES children²⁶. With respect to mental health factors, we know of very few studies²⁷, despite reports of moderate associations between anxiety and reading skills^{20,30}. Together, these findings foster our investigation of participant-level and environmental factors that may contribute to EdTech reading intervention effects. In particular, there is a need for examining mental health and environmental factors, which have been understudied, and may have unique predictive value for response to home-delivered interventions.

Brain-based measures have potential utility for enhancing explanatory models of treatment outcomes.

A growing number of studies show that neurobiological measures make unique contributions in predictive models of reading outcomes, suggesting that these are sensitive measures of plasticity. This holds true for both task-based imaging measures that assess functionality during reading (and related) tasks, as well as for structural imaging of the neuroanatomy that supports these functions. Two studies from our investigative team demonstrate this predictive utility: 1) Co-I Hoeft and colleagues found that greater pretreatment activation of right Inferior Frontal Gyrus (IFG) during reading and greater white matter integrity (Fractional Anisotropy) in the right superior longitudinal fasciculus (SLF) predicted amount of reading gain over two years, beyond the behavioral predictors²⁹; and, 2) Preston and colleagues found that the degree to which children activated the same left hemisphere language-associated regions (IFG, Inferior Parietal Cortex [IPC]) during both print and speech processing predicted reading gains two years later, above and beyond initial reading level³⁰. These findings suggest that both initial integrity in the left hemisphere (LH) reading circuit and in alternative regions that support reading success in the right hemisphere (RH) are predictive of positive reading intervention outcomes²⁹⁻³². Moreover, other work suggests that neurobiological measures may be sensitive to other factors of interest in the current proposal; for example, a recent study found that low SES RD children showed greater cortical thickening from pre-to-post-test (relative to high SES children), suggesting that neurobiological response to intervention may differ as a function of SES in children with RD²⁶. These findings support the current study's focus on selected neurobiological predictors of treatment response, including task-based electroencephalogram (EEG) measures and structural MRI measures.

Overall Goal of Study

This project seeks to determine the effectiveness of in-home administered reading remediation (GraphoLearn), in a large, diverse, sample of children with RD. To this end, we compare the performance of children randomized to GraphoLearn or to an active control (math game), for 12 weeks, and followed beyond treatment over another 12 weeks (retention period). To explore treatment response at the behavioral and neurobiological levels, we will obtain pre- and post-intervention measures of reading and reading relevant skills and EEG/ERP measures of print and speech processing. In addition, comprehensive assessments will be obtained, which include detailed demographics, parenting style, and children's IQ, medical history, mental health status, and social skills. Critically, beyond determining the overall effectiveness of GraphoLearn (for which we expect there to be substantial individual variability), the sample diversity and availability of extensive behavioral and neurobiological data on participants in this

study will permit the development of explanatory models of predictive factors for GraphoLearn intervention outcomes.

II. Study Objectives

Evaluate the effectiveness of GraphoLearn in a large sample of children with RD. This project addresses limitations inherent in remedial interventions by implementing a large-scale, randomized controlled trial (RCT) of home administered GraphoLearn in 450 RD children (boys and girls, ages 6.0-11.0). To accomplish this goal rapidly and with minimal cost, we will recruit participants enrolled in the Healthy Brain Network [HBN] (Advarra Protocol Number Pro00012309). HBN is an ongoing study of mental health and learning disorders in children, ages 5.0-21.0, whose family have concerns about behavior and/or learning (target $n = 10,000$; current $n = 3000+$). The HBN includes comprehensive psychiatric, cognitive, EEG and multimodal MRI characterizations for all participants, providing the proposed study with rich descriptive data. Reading disabilities will be identified using the TOWRE (standard score < 85). Children will be assigned to a 12-week/20-hour program of GraphoLearn or an “active” control condition, group matched for IQ and reading-level. The control will consist of a math game, similar in style to GraphoLearn, to control for non-specific effects of the intervention, such as time spent on the program, attention, and other factors. Children randomized to the control group will be offered GraphoLearn following the study due to ethical considerations. Consistent with real-world conditions, non-study related interventions during the course of the study period will be permitted. They will be documented for comprehensive sample description, and for analyzing predictors of response.

Children’s reading level will be assessed before, during (week 6), and at the end of, the active study (week 12). Extended effects will be evaluated 12 weeks post treatment (week 24). We hypothesize statistically significant differences in reading level in favor of the GraphoLearn group over the control group at weeks 12 and 24. Should the hypothesis be supported, it will have important public health implications for children with RD (there is no claim that the program will be a substitute for individualized, and expert-supervised, reading remediation).

Identify participant-related and environment-level factors that are significantly associated with GraphoLearn outcomes. The identification of participant-level and environment-level factors that predict treatment response is an important topical goal but has not been considered for GraphoLearn. For specialist-based interventions, a consistent finding is that older children, and children with language or cognitive impairments, are less likely to benefit. Here, we leverage the diversity of co-morbid clinical presentations among poor readers in HBN to identify putative predictors of treatment response. In line with recent studies suggesting the added value of brain-based features for prediction of intervention outcomes, we also leverage the availability of anatomical MRI data in the development of predictive models. Only MRI data previously /planned to be collected as part of the HBN will be used.

The extensive phenotypic and imaging characterizations included in the HBN sample are uniquely suited for exploring an extensive set of participant and environmental factors that may predict treatment outcomes. Random Forest Regression models will be used to predict differences in intervention response to GraphoLearn, based on an a priori set of baseline demographic, cognitive, emotional, brain and home environment variables. Assuming that compliance (i.e., actual time spent playing GraphoLearn relative to

prescribed) is found to be a strong predictor of response, exploratory analyses will also look at predictors of compliance to further increase real-world value of this work.

III. Study Design

A total of 450 children from the HBN will be randomly assigned to a 12-week GraphoLearn intervention group (Group 1) or to a Math Control group (Group 2) in a between-groups design and evaluated immediately before (Time 1 – T1) and after the intervention (Time 2 – T2). To assess the extended differential effects of the intervention, children will then be followed-up after a 12-week post-intervention retention period (Time 3 – T3). A standardized battery of reading and math skills, as well as neurobiological (EEG) measures of orthographic and phonological sensitivity, will be administered at timepoints Time 1, 2 and 3. An additional time point (T4) is included for Group 2 (i.e., control group) participants who opt to also complete GraphoLearn. Participant-level and environment-level measures from the HBN are obtained prior to study enrollment and will be used to generate comprehensive predictive models of treatment outcome. In addition to in-person assessments at Time 2, we will conduct a brief (under 10 minute) video-based (i.e., Zoom) assessment at the 6-week, mid-intervention, point. It will consist only of the Test of Word Reading Efficiency (TOWRE), Grapho-Strain Task, and two subtests of the Comprehensive Test of Phonological Processing (CTOPP).

Project coordination will take place at the centrally located Child Mind Institute Headquarters, in midtown Manhattan (101 E 56th Street, New York, NY, 10022). Data collection will take place at our Child Mind Institute Headquarters, in midtown Manhattan and our satellite offices located in Harlem (286 Malcolm X Boulevard, 2nd Floor, New York, NY).

Strategies for Recruitment and Retention

All recruitment will be from HBN participants meeting the inclusion/exclusion criteria specified below (the TOWRE, WISC V, and comprehensive psychiatric evaluation are all included in the HBN protocol). Recruitment will occur at two different times: 1) Prior to enrollment in the HBN through external advertising of an HBN and GraphoLearn study progression and 2) During HBN protocol after HBN enrollment. HBN evaluations comprise four visits, each lasting 3-3.5 hours, and each with a specific focus: visit 1: intake and cognitive testing; visit 2: multimodal MRI imaging; visit 3: language and learning; visit 4: mental health. All research data generated through HBN is openly shared in a de-identified form. One month following completion of the four sessions, participant families are provided a comprehensive mental health and learning diagnostic report and treatment recommendations. Families will be approached at the end of the fourth HBN visit, as the assessments needed to determine eligibility will be completed and scored by this time. Information about the proposed study will be included in welcome packets given at HBN visit 1. During visit 4, caregivers of eligible children will be introduced to our research coordinator (who will be co-located with the HBN team); families and children will be informed about the study and enrolled if they wish – thus enabling a relatively seamless connection between studies. All subject and recruitment materials will be submitted for review prior to use.

For our external advertisements, we will advertise for children aged 6.0 to 10.0 only given the Healthy Brain Network can up to 6 months to complete. This recruitment strategy ensures that those participants

who are externally recruited and must complete the Healthy Brain Network are still within the age range once they are determined eligible for the GraphoLearn study.

The HBN employs a community-referred recruitment model. Advertisements encourage participation of families who have concerns about psychiatric symptoms or learning difficulties in their child. The 'announcements' are distributed to community members, educators and local care providers, as well as directly to parents via email lists and events. The advertisements highlight the potential value of participation for children who may require school-based accommodations (e.g., comprehensive diagnostic report, actionable treatment recommendations).

The HBN has enrolled participants at a rate of 90-100 participants per month over the past two years, a rate expected to increase to 125-150 a month with its expansion to a new site in Harlem in June 2019. To date, since 2016, study non-completion has consistently ranged between 10 and 12%. Examination of the current HBN sample ($n = 3,000$ children ages 5-21) revealed that 32.4% of children aged 6.0-10.0 years old meet the proposed study's criterion for reading disability (defined here as TOWRE TWRE SS < 85). A conservative yearly enrollment estimate for the project years (2020-2022) is 1,500 children per year; based on this estimate, and trends over the past 4 years, we expect that 750 per year will be enrolled in our targeted age range, yielding 720 children who will meet all study criteria over the enrollment period (TOWRE TWRE SS < 85, IQ > 75, normal vision hearing, willing and able to comply). Consequently, to reach the study target, we need to enroll 62% of the eligible HBN participants to reach our target $n = 450$ – a goal that is expected to be easily achieved in families who are actively seeking help for their children. Additionally, should there be any shortfalls in recruitment, we also have the option of recruiting children who have completed the HBN protocol prior to the proposed project period. Thus, while our primary strategy is to focus recruitment on children currently being enrolled in the study, we have the option to recruit previously complete participants in the case of any shortfalls in recruitment from the sample.

Collectively, the PIs and co-investigators have extensive experience with longitudinal studies, and recent experience with intervention studies, including reading interventions with EdTech. We will optimize retention by hiring experienced staff and maintaining frequent (i.e., weekly) 5-minute phone check-ins with participants and caregivers, over the 12-week GraphoLearn training period, and monthly during the retention period. Contact will be maintained through multiple avenues, including phone, email, and text, and be personalized to caregiver preferences. Despite efforts, we estimate and account for 10% attrition at each assessment Time point.

Inclusion Criteria

1. Age: 6.0 to 11.0 years old
2. Sex: male and female
3. HBN: currently or previously enrolled in HBN
4. Reading, language, or learning difficulties: To meet this criterion, children should meet one of the following:
 1. a standard score ≤ 85 , on the Test of Word Reading Efficiency³³ (TOWRE), Total Word Reading Efficiency (TWRE)

2. a standard score <90, on the Test of Word Reading Efficiency (TOWRE), Total Word Reading Efficiency (TWRE), AND a learning disability or presumptive diagnosis of learning disability in reading, language, or writing
3. a standard score <90 on the Test of Word Reading Efficiency (TOWRE), Total Word Reading Efficiency (TWRE), AND a very poor to a below-average standard score on the Comprehensive Test of Phonological Processing (CTOPP) Phonological Awareness composite
4. a standard score between 90 and 100 (inclusive) on the Test of Word Reading Efficiency (TOWRE), Total Word Reading Efficiency (TWRE), AND an average standard score on the Comprehensive Test of Phonological Processing (CTOPP) Phonological Awareness composite AND significant parent or teacher concerns about reading or phonological processing AND a learning disability or presumptive learning disability that includes reading, language, or writing.
5. IQ: To meet this criterion, children should meet one of the following:
 1. Full Scale IQ (FSIQ) on the Wechsler Intelligence Scale for Children (WISC V)³⁴ ≥ 70
 2. IQ Composite Standard Score on the Kaufman Brief Intelligence Test-II (K-BIT II) ≥ 70 for children 5 years old.
6. Language: Native speakers of English

Exclusion Criteria

1. Vision or hearing impairment, per the Snellen Test and audiometer measurement
2. Current exclusionary psychiatric conditions will be (not others):
 1. moderate to severe autism (levels 2-3)
 2. major depressive disorder
 3. suicidal or homicidal ideation
 4. psychotic disorder
 5. manic episode
3. Any neurologic impairment that limits the ability to use a keyboard or mouse, or to touch a screen.

Randomization and Group Assignment

The study design calls for two group conditions – 1) the GraphoLearn intervention and 2) an active math control. The two groups will be matched for time of year at study entry, which will occur during three possible time-windows or “waves” (Oct/Nov, Feb/March, July/August), (Table 1). At each entry window, a stratified randomization strategy that subgroups new enrollees based on age (e.g., one-year bins) and IQ will be used to avoid unintended group differences in these variables. We will rely on randomization to balance other variables (e.g., internalizing behaviors, externalizing behaviors, SES). Should unexpected

Year 1			Year 2			Year 3		
Oct/Nov	Feb/March	July/August	Oct/Nov	Feb/March	July/August	Oct/Nov	Feb/March	July/August
W1; N = 50	W2; N = 50	W3; N = 50	W4; N = 50	W5; N = 50	W6; N = 50	W7; N = 50	W8; N = 50	W9; N = 50

group differences emerge during Year 1, we will consider a more complex stratification procedure for subsequent years.

Study Treatments

GraphoLearn Reading Intervention. GraphoLearn is an evidence-based treatment, delivered as an engaging computer game. Players match auditory targets (e.g., phonemes, rimes) to visual targets (single letters, letter sequences, words). The same stimuli and stimulus types are presented in multiple graphical settings to provide variety that keeps players engaged. For example, players might select the ball containing the letter sequence corresponding to the auditorily presented rime among a series of balls falling from the top of the screen. The game provides trial-wise feedback on accuracy and provides corrective feedback by guiding players to select the correct mapping. The game logs accuracy and latencies to all trials, and players are rewarded every few minutes with in-game tokens, virtual stickers, etc. GraphoLearn is adaptive such that players need to achieve at least 80% accuracy before introducing a new complexity level. The complexity of the items within each level is ordered such that at each level, the most frequent and regular mappings are introduced first based on measures such as orthographic/phonological neighborhood size and morphological family size. GraphoLearn allows the children to practice and reinforce lessons at their own individual trial pace and provides a record of performance progress that can be used to guide analyses. Before beginning the intervention, children and their family will be introduced to the concepts and trained in the game by staff who are knowledgeable of reading theory and GraphoLearn. We will emphasize to caregivers the importance of their engagement in their children's game play and introduce them to the Likert-type caregiver engagement questionnaires (outlined below) that they will be instructed to complete each week. GraphoLearn does not meet criteria to be submitted to the FDA as a software medical device.

Active Math Control. Recent meta-analysis indicates that findings for GraphoLearn are similar whether one uses an active or waitlist control²¹. We selected an active control to maximize the specificity of the treatment outcomes; to this end, math games are among the most commonly used. The Math Control game is a noncommercial, computerized game, appropriate for children ages 6 and up^{35,36}. Game sessions support learning numerical mathematical skills and cognition related to mathematical skills. In addition to math, this game contains training tasks for visuospatial working memory, spatial visualization, and visuospatial reasoning. The overall theme of the game and feedback style are similar to those in GraphoLearn.

Intervention Schedule

Children will be told to play the games for 20 minutes per day, 5 days per week, for 12 weeks, on tablets provided by the research team. With this design, children have a total active intervention time of 20 hours for each game (GraphoLearn, Math Control game).

Compliance

Based on recent meta-analysis²¹ that notes the potential importance of adult supervision in EdTech intervention, we will ask caregivers to monitor administration as much as possible. Despite our efforts to encourage compliance, it is expected that there will be some non-compliance (see Innovation, as well as Design Alternatives and Considerations). GraphoLearn software logs information about when and for how long children play the game. We will monitor compliance online (via GraphoLearn whenever possible and

via cell phone-based app) and provide feedback at weekly phone check-ins. The primary variables we will use to index compliance are: 1) total hours played, 2) average time played per week, and 3) number of days played per week. We include these measures as predictors of outcome.

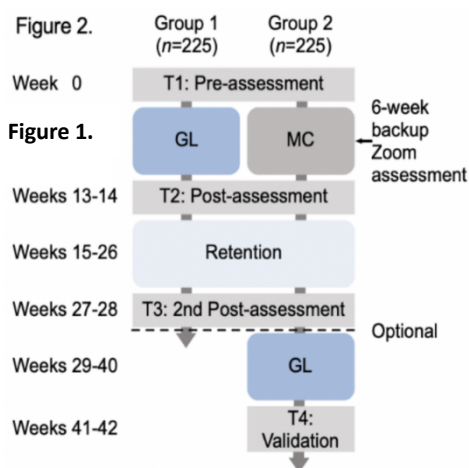
Gameplay Monitoring

Given previous findings which indicate that level of caregiver and/or teacher engagement affects EdTech intervention outcomes¹³, we will gather this information daily via cell phone with the MindLogger app platform developed by EdTech experts at CMI (<https://mindlogger.org>) for the purposes of ecological momentary assessment (aka daily diary cards or experience sampling). Currently there are no terms of use or privacy policy for users to agree to when using MindLogger. The MindLogger platform is owned and operated by the Child Mind Institute and all data will be stored internally. The app will prompt parents daily to have their child complete the GL schedule, as well to ask about the level of caregiver involvement. This includes a survey modeled on a 5- point Likert-type questionnaire developed by the makers of GraphoLearn and previously used to establish the importance of adult involvement¹³. The first question asks about general circumstances under which children played the game over the week (1 = child played alone, 2 = occasional adult visits in the room where the game was used, 3 = adult present in the room, 4 = adult present in the room and actively observing the child, providing help if needed, 5 = adult actively provided help), the second asks caregivers to evaluate how often they provided positive feedback and encouragement to the child during training (5-point scale, ranging from constantly to never). We include these measures as predictors of outcome.

IV. Methodology

Assessment Schedule

Assessments at each assessment time point (T1 - T3 [and T4 for Group 2]) are identical. They include tests of reading and reading-relevant skills (and math for exploratory analyses), as well as experimental EEG tasks. Time 1 assessment establishes baseline performance for all children. Following Time 1, Group 1 receives GraphoLearn (reading intervention), and Group 2 receives the active control (math intervention). Time 2 assessment serves to quantify the effects of the GraphoLearn intervention relative to the active control group. Following Time 2 is a 12-week no-treatment retention period. Time 3 assessment will quantify the extended effects of GraphoLearn. For Group 1, this is the end of the study. At this point, Group 2 will be offered a 12-week GraphoLearn intervention which would be followed by a Time 4 (T4) assessment (Figure 1) to validate treatment effects in this group.



In addition to in-person assessments at Time 2, we will conduct a brief (under 30 minutes) video-based (i.e., Zoom) assessment at the 6-week, mid-intervention, point. It will consist of the Test of Word Reading Efficiency (TOWRE;

detailed below), Grapho-Strain Task, and two subtests of the Comprehensive Test of Phonological Process (CTOPP). The 6-week assessment will serve two purposes: one, it will inform whether effects, if found at the end of the intervention, are detectable after 6 weeks; and two, importantly, it will provide data for participants who may drop out between weeks 7 and 12, protecting against complete data loss of the primary outcome measure.

Participant Compensation

Participants will be reimbursed at a rate of \$50 per imaging session. All participants will be expected to attend three sessions (baseline, immediate post-intervention assessment, retention assessment); participants assigned to the control condition will be asked to attend an additional assessment after being given the reading intervention. Participants will also be reimbursed \$25 for transportation to each session.

Pre- and Post-Intervention Study Measures

Assessments to be completed by study participants are outlined below and described in detail within the attached table of study measures, Attachment A.

Standardized measures of reading and language skills. These measures, given at initial intake, post-GraphoLearn intervention, and following the 12-week retention period, will enable evaluation of GraphoLearn treatment outcomes. Measures were selected because they (or similar UK-normed versions) have been shown to be sensitive to treatment gains following GraphoLearn intervention in previous studies^{19,20}. They have been widely used in educational research and have excellent psychometric properties, including high reliability and construct validity. These assessments are also ideal for comparing repeated measures across children of a broad age range because they utilize the same test for diverse ages. Participants will complete assessments of the following reading and reading-related skills: (1) Timed word and nonword reading, which will be assessed with the Test of Word Reading Efficiency, phonemic decoding efficiency (PDE) and sight word efficiency (SWE) subtests; these comprise the total word reading efficiency (TWRE) measure, TOWRE-II³³, which will be our primary treatment outcome measure; as well as the Woodcock Johnson Tests of Achievement (WJ-IV), Letter Word ID and Word Attack subtests, which are untimed³⁷; (2) Phonological processing, including Phonological Awareness, Phonological Memory, and Rapid Naming, which will be assessed with the Comprehensive Test of Phonological Processing (CTOPP-II)³⁸; (3) Spelling, which will be assessed with the WJ-IV³⁷; and, (4) Vocabulary (Time 1 only for prediction of outcome), which will be assessed with the Peabody Picture Vocabulary Test (PPVT)³⁹.

Additional measures of reading and language skills. In addition to the measures above, participants will complete the following assessments: Moby.Read Oral Reading Fluency Benchmark Assessment, a self-administered online assessment that is automatically scored⁴⁰, Cross-Modal Matching Task, a simple cross-modal matching of spoken letters, word, and pseudowords with print targets in a two-alternative forced choice paradigm⁴¹, and Strain Naming Task, a word naming task with a manipulation of items' frequency, imageability, and consistency⁴², and Grapho-Strain Task, a word reading task consisting of 160 trials presented to children in a fixed order. Words are taken directly from the GraphoLearn intervention.

Participant's Parent or Legal Guardian will also complete the COVID-19 Home Literacy Study from GAAB LAB, an online questionnaire administered to learn more about the effects of COVID-19 on literacy environment at home (see Attachment A for additional information).

Standardized measures of math. Although the math intervention is an active control here for our primary intervention of interest (reading), we include a standardized assessment of math skills with excellent psychometric properties at each testing time point for future exploratory analyses of response to math intervention; The KeyMath-3⁴³.

EEG measures. We use two well-studied EEG/ERP measures that are sensitive to different reading sub skills and have been shown to correlate with reading outcomes.

1. AV EEG Task: to assess participant's audiovisual print-speech integration, we will administer an Audio Visual (AV) Letter Sounds Task⁴⁴.
2. MMN EEG Task: to index phoneme discrimination, another precursor to successful reading that also increases with reading skill, we will utilize two standard/deviant oddball paradigms to elicit mismatch negativity (MMN) responses⁴⁵.

Additional Predictor Measures

In addition to measures utilized to capture intervention response, we will include a range of measures already obtained through HBN participation as well as measures obtained at the beginning of the GraphoLearn protocol for the purposes of developing predictive models. These measures will serve to provide more comprehensive assessments, and some are hypothesized to be associated with intervention outcomes. These measures include demographic and familial predictors, handedness, psychoactive medications, parenting, mental health, attention and social skills, social functioning, emotional and behavioral problems, self-Efficacy, and neuroimaging measures. Questionnaires to be completed by participants, parents/guardians, and teachers, if applicable, are outlined in Attachment B.

V. Data Collection and Management Procedures

GraphoLearn data will be collected manually, via paper, or electronically at Child Mind Institute's Midtown Location. Data may also be collected via an online participant portal.

1. *Electronic:* Electronic data will be collected via NextGen, the Collaborative Informatics and Neuroimaging Suite (COINS) and Research Electronic Data Capture (REDCap). NextGen (<http://www.nextgen.com/>) will provide the infrastructure for both the collection of applicable phenotypic assessments and the participant portal. All data collected electronically will be stricken from any participant PHI. Participants will be given a random unique identifier, generated directly from the electronic data capture system or the Global Unique Identifier (GUID) Tool at the start of participation. The GUID Tool is a customized software application that generates a Global Unique Identifier for each study participant. The GUID is made up of random alpha-numeric characters and is NOT generated from PHI. As such, it has been approved by the NIH Office of General Counsel. All data will be labeled with this unique identifier. Applicable participants will

receive a unique login and password to enter on the participant portal that serves as a secure, HIPAA-compliant, easy way for participants, and parents of participants, to complete relevant study measures assessments that are not collected during the face-to-face session. The Participant Portal is equipped with a variety of security features, including 128-bit encryption, firewall, anti-virus software, and intrusion detection software. Messaging features are also secure. All electronic data collected as part of GraphoLearn is kept on password-protected computers. Only approved researchers will be granted with logins for electronic data entry, review, and processing.

2. *Paper:* Paper records, including consent, assent, and HIPAA forms, and paper-CRFs will be kept in participant files in locked file cabinets. These materials are additionally kept behind locked doors. Participant data collected via paper-CRFs will not contain any participant PHI. These data forms will be de-identified and contain only the participant's unique identifier. Additionally, data collected via paper will be input into the electronic data system. Therefore, all participant data will be contained within a single, secure, electronic system to facilitate data sharing efforts of the project.

In all cases, if data is collected via paper or electronic means, access to study data is limited to research staff at the Child Mind Institute. Data is fully de-identified and entered on paper forms or electronically using the previously generated unique identifier. These unique identifiers are also used for digital data transmission, storage, and analysis. Protected health information will not be divulged, even to possible future consultants. Instead, participant identifiers will be used. Primary data storage will be via the Child Mind Institute secure data servers or Mind Research Network (MRN) in New Mexico. Any data stored in MRN's COINS database will be completely secure as this system is designed with security and restricted/controlled access features. External access to the system is restricted via a VPN account with MRN or through a firewall rule for limited IP addresses within a collaborating institution.

Pre-Distribution Data Preparation: Prior to distribution, all phenotypic (i.e., psychological, self-report, etc.) information will be entered by GraphoLearn researchers and clinicians and subjected to cross-validation to protect against human error.

Data Distribution: via International Neuroimaging Data-sharing Initiative (INDI) Distribution Formats. Consistent with the model setup by the PI, Dr. Michael Milham for the INDI effort, imaging data will be distributed via the NITRC-based 1000 Functional Connectomes Project (FCP)/INDI website and its ancillary sharing venues (e.g., MRN's COINS). Users will receive meta-data as .csv files. No PHI will be contained within this study data. All data will be labeled with the participant's unique identifier.

Phenotypic data will be shared via resources developed by the PI. Data will be preprocessed, with all identifying information removed, prior to integration into this framework. As in the process with FCP/INDI, users will receive meta-data as .csv files which in no circumstance contain any participant PHI. All data will be labeled with the participant's unique identifier.

Protection of Privacy: All participants will give informed consent for sharing of their data prior to participation in the study. The FCP/INDI interfaces are fully anonymized data repositories. Consistent with

the FCP/INDI protocol previously employed, the following steps will be carried out to ensure that no identifying or protected health information is released:

- Random 1-2-month lags between acquisition date and distribution will further protect against identification.
- Random ID numbers will be generated and assigned to each dataset prior to release.

Data Use Agreement: The project has elected to adopt a modified version of both the IRB-approved CMI Biobank and NKI-RS Data Use Agreement (DUA) requirement for members of the scientific community to gain access to the datasets generated. As in the enhanced NKI-RS, this decision was motivated by the following factor:

- The dataset in this study is being obtained from participants within the NYC metropolitan area. As a result of this, a participant's residential location is identifiable to the level of a city, thus prohibiting complete de-identification in accord with HIPAA's 18 protected health identifiers.
- We believe the concurrently supplied psychometric data in the HBN phenotypic protocol engenders an increased risk of identification beyond that of the traditional 18 protected health identifiers specified by HIPAA. This further supports the need of a data use agreement.

Data Safety Monitoring

This protocol and its component procedures (including the GraphoLearn intervention) present minimal risk to participants, and adverse events or other problems are not anticipated. Given the minimal risk, PI Milham will take primary responsibility for carrying out the DSMP. The PI will provide continuous oversight and evaluation of study conduct, recruitment, retention, and occurrence of adverse events. These include regular monitoring of reading intervention compliance and of data quality. In the event that ongoing site monitoring identifies threats to data quality or participant safety, the PI and such project personnel as deemed appropriate will confer outside regular review to address those issues.

Procedures to monitor confidentiality of participant data

Participant confidentiality will be protected by keeping all signed consent forms and any paper data in secure, locked file cabinets, with code numbers used in place of names on all data. No information will be given to anyone (to the extent permitted by law) about participants without a signed release by the child's guardians. Data collected for the study will be entered into a secure HIPAA compliant database—NextGen—on a rolling basis, as it is acquired. All project staff will be educated about the importance of strictly protecting participants' rights to confidentiality. In the unlikely event of a breach of confidentiality, specific procedures will be followed. The participant will be informed of this breach, as will the Advarra IRB and the IRB at the University of Connecticut.

VI. Potential Benefits and Risks

Potential Harms/Risks and Inconveniences

Risks associated with all tasks are minimal, however, participants may become frustrated, bored, or fatigued during test sessions or participation in the EdTech intervention. Participants may become bored or fatigued during behavioral assessment as well.

EEG: The risks involved in performing the EEG are minimal, and the procedures used are standard in clinical and basic research settings. While the procedure is non-invasive, children may experience discomfort toward the end of the session due to the fit of the net on their scalp. The cap is thoroughly cleaned after each use. Stimuli in all tasks will be presented at comfortable intensity levels. There is a risk of boredom and/or tiredness.

Benefits

Individuals who participate may see significant increases in their reading skills following treatment. Participants in the study will be compensated for their time and transportation expenses with standard monetary amounts for each visit approved by the Child Mind Institute Institutional Review Board and the University of Connecticut Institutional Review Board. All participants will be provided with the results of standardized behavioral assessment battery, which may provide minimal benefit. Additional benefits that may develop from the research are likely to be realized by society at large and may include better understanding of the effectiveness of this treatment and the behavioral and neurobiological predictors of treatment response. The potential benefits to children and families and the contributions to scientific research on comprehension ability are outweighed by the minimal risks involved.

Risk/Benefit Analysis

There is no more than minimal risk of physical or psychological damage associated with this study. Proper training of research staff and prior familiarization of subjects with the experimental methods will serve as the primary protection against risk. In addition, to protect against fatigue participants will be given breaks between testing sessions and longer assessments. Risks associated with confidentiality are mitigated by appropriate data storage and identification procedures.

1. Eye injuries: The risks inherent in the visual procedures are no greater than those encountered during an individual's normal reading or recreational activities.
2. Ear injuries: The sound pressure levels to be used in the auditory experiments will be set at comfortable levels that pose no threat of injury to the ear.
3. Fatigue: Participants will make 3-4 visits to the lab on separate days over the course of 29-44 weeks. No session will last more than 3 hours, and participants will be provided with ample breaks within sessions.
4. Risks to subject confidentiality will be minimized by adopting suitable data storage procedures. All data for each subject will be identified by numerical ID. Data collected for the study will be entered into a secure HIPAA compliant database—NextGen—on a rolling basis, as it is acquired. Reports based on these data will not identify subjects by name. The principal investigators are responsible for monitoring the data and conducting performance safety reviews (see data and safety monitoring plan). No physical, emotional, psychological, social stigmatization, economic or legal risk is anticipated. The principal investigators will evaluate any unanticipated adverse event and determine whether the adverse event affects the Risk/Benefit ratio of the study and whether

modifications to the protocol or consent form are required. A summary of any adverse events will be reported to the IRB, at minimum, when annual re-approval of the protocol is sought. The summary will include the number of subjects enrolled and a summary of graded adverse events to date.

Additional Protections for Children Involved as Subjects in Research. The proposed study involves no more than minimal risk. As described above, participants will give written assent and consent will be solicited from their parents or guardians.

Procedures to report any Adverse Event during the study

In the unlikely case that serious and unanticipated, study-related adverse events or unanticipated problems involving risks to subjects or others do occur, these will be reported in writing within 48 hours to the Advarra IRB and the IRB at the University of Connecticut. The PIs will promptly apprise fellow investigators and study personnel of such events that occur during the conduct of this research project. This communication will take place by email from the PIs. The funding agency for the study, the National Institutes of Health (NICHD), will be informed of any adverse event that jeopardizes the health of subjects within 10 business days of the event becoming known to the PIs.

References:

1. Duncan GJ, Magnuson K. Socioeconomic status and cognitive functioning: moving from correlation to causation. *Wiley Interdiscip Rev Cogn Sci*. 2012;3(3):377-386. doi:10.1002/wcs.1176
2. Francis DA, Caruana N, Hudson JL, McArthur GM. The association between poor reading and internalising problems: A systematic review and meta-analysis. *Clin Psychol Rev*. 2019;67:45-60. doi:10.1016/j.cpr.2018.09.002
3. Becker N, Vasconcelos M, Oliveira V, et al. Genetic and environmental risk factors for developmental dyslexia in children: systematic review of the last decade. *Dev Neuropsychol*. 2017;42(7-8):423-445. doi:10.1080/87565641.2017.1374960
4. Fuchs LS, Fuchs D. Creating Opportunities for Intensive Intervention for Students with Learning Disabilities. *TEACHING Exceptional Children*. 2009;42(2):60-62. doi:10.1177/004005990904200207
5. Landi N, Cutting LE. Global approaches to early learning research and practice: integrative commentary. *New Dir Child Adolesc Dev*. 2017;2017(158):105-114. doi:10.1002/cad.20223
6. Torgesen JK. Individual differences in response to early interventions in reading: the lingering problem of treatment resisters. *Learning Disabil Res Pract*. 2000;15(1):55-64. doi:10.1207/SLDRP1501_6
7. Lovett MW, Frijters JC, Wolf M, Steinbach KA, Sevcik RA, Morris RD. Early intervention for children at risk for reading disabilities: the impact of grade at intervention and individual differences on intervention outcomes. *J Educ Psychol*. 2017. doi:10.1037/edu0000181
8. Morris RD, Lovett MW, Wolf M, et al. Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *J Learn Disabil*. 2012;45(2):99-127. doi:10.1177/0022219409355472

9. Frijters JC, Lovett MW, Steinbach KA, Wolf M, Sevcik RA, Morris RD. Neurocognitive predictors of reading outcomes for children with reading disabilities. *J Learn Disabil.* 2011;44(2):150-166. doi:10.1177/0022219410391185
10. Susi T, Johannesson M, Backlund P. Serious games: An overview. 2007.
11. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. *J Educ Psychol.* 2013;105(2):249-265. doi:10.1037/a0031311
12. Landi N, Cutting LE. Global approaches to early learning research and practice: integrative commentary. *New Dir Child Adolesc Dev.* 2017;2017(158):105-114. doi:10.1002/cad.20223
13. Ronimus M, Lyytinen H. Is School a Better Environment Than Home for Digital Game-Based Learning? The Case of GraphoGame. *Human Technology.* 2015;11(2):123-147. doi:10.17011/ht/urn.201511113637
14. Buckley P, Doyle E. Gamification and student motivation. *Interactive Learning Environments.* 2016;24(6):1162-1175. doi:10.1080/10494820.2014.964263
15. Dicheva D, Dichev C, Agre G, Angelova G. Gamification in education: A systematic mapping study. *Educational Technology & Society.* 2015;18(3):75-88.
16. Muntean CI. Raising engagement in e-learning through gamification. *Proc 6th International Conference on Virtual Learning ICVL.* 2011;1.
17. Lyytinen H, Erskine J, Kujala J, Ojanen E, Richardson U. In search of a science-based application: a learning tool for reading acquisition. *Scand J Psychol.* 2009;50(6):668-675. doi:10.1111/j.1467-9450.2009.00791.x
18. Richardson U, Lyytinen H. The GraphoGame Method: The Theoretical and Methodological Background of the Technology-Enhanced Learning Environment for Learning to Read. *Human Technology.* 2014;10(1):39-60. doi:10.17011/ht/urn.201405281859
19. Bhide A, Power A, Goswami U. A Rhythmic Musical Intervention for Poor Readers: A Comparison of Efficacy With a Letter-Based Intervention. *Mind Brain Education.* 2013;7(2):113-123. doi:10.1111/mbe.12016
20. Kyle F, Kujala J, Richardson U, Lyytinen H, Goswami U. Assessing the Effectiveness of Two Theoretically Motivated Computer-Assisted Reading Interventions in the United Kingdom: GG Rime and GG Phoneme. *Read Res Q.* 2013;48(1):61-76. doi:10.1002/rrq.038
21. McTigue EM, Solheim OJ, Zimmer WK, Uppstad PH. Critically reviewing graphogame across the world: recommendations and cautions for research and implementation of computer-assisted instruction for word-reading acquisition. *Read Res Q.* May 2019. doi:10.1002/rrq.256
22. Cho E, Roberts GJ, Capin P, Roberts G, Miciak J, Vaughn S. Cognitive Attributes, Attention, and Self-Efficacy of Adequate and Inadequate Responders in a Fourth Grade Reading Intervention. *Learn Disabil Res Pract.* 2015;30(4):159-170. doi:10.1111/ldrp.12088
23. Fletcher JM, Stuebing KK, Barth AE, et al. Cognitive correlates of inadequate response to reading intervention. *School Psych Rev.* 2011;40(1):3-22
24. Ron Nelson J, Benner GJ, Gonzalez J. Learner Characteristics that Influence the Treatment Effectiveness of Early Literacy Interventions: A Meta-Analytic Review. *Learning Disabil Res Pract.* 2003;18(4):255-267. doi:10.1111/1540-5826.00080
25. Al Otaiba S, Fuchs D. Characteristics of children who are unresponsive to early literacy intervention. *Remedial and Special Education.* 2002;23(5):300-316. doi:10.1177/07419325020230050501

26. Romeo RR, Christodoulou JA, Halverson KK, et al. Socioeconomic status and reading disability: neuroanatomy and plasticity in response to intervention. *Cereb Cortex*. 2018;28(7):2297-2312. doi:10.1093/cercor/bhx131
27. Grills-Taquechel AE, Fletcher JM, Vaughn SR, Stuebing KK. Anxiety and reading difficulties in early elementary school: evidence for unidirectional-or bi-directional relations? *Child Psychiatry Hum Dev*. 2012;43(1):35-47. doi:10.1007/s10578-011-0246-1
28. Haft SL, Duong PH, Ho TC, Hendren RL, Hoeft F. Anxiety and Attentional Bias in Children with Specific Learning Disorders. *J Abnorm Child Psychol*. 2019;47(3):487-497. doi:10.1007/s10802-018-0458-y
29. Hoeft F, McCandliss BD, Black JM, et al. Neural systems predicting long-term outcome in dyslexia. *Proc Natl Acad Sci USA*. 2011;108(1):361-366. doi:10.1073/pnas.1008950108
30. Preston JL, Molfese PJ, Frost SJ, et al. Print-Speech Convergence Predicts Future Reading Outcomes in Early Readers. *Psychol Sci*. 2016;27(1):75-84. doi:10.1177/0956797615611921
31. Aboud KS, Barquero LA, Cutting LE. Prefrontal mediation of the reading network predicts intervention response in dyslexia. *Cortex*. 2018;101:96-106. doi:10.1016/j.cortex.2018.01.009
32. Hoeft F, Ueno T, Reiss AL, et al. Prediction of children's reading skills using behavioral, functional, and structural neuroimaging measures. *Behav Neurosci*. 2007;121(3):602-613. doi:10.1037/0735-7044.121.3.602
33. Torgesen JK, Wagner R, Rashotte C. Test of Word Reading Efficiency:(TOWRE-2). *Test of Word Reading Efficiency:(TOWRE-2)*. 2012.
34. Kaufman AS, Raiford SE, Coalson DL. *Intelligent Testing with the WISC-V*. John Wiley & Sons; 2015.
35. Nemmi F, Helander E, Helenius O, et al. Behavior and neuroimaging at baseline predict individual response to combined mathematical and working memory training in children. *Dev Cogn Neurosci*. 2016;20:43-51. doi:10.1016/j.dcn.2016.06.004
36. Räsänen P, Salminen J, Wilson AJ, Aunio P, Dehaene S. Computer-assisted intervention for children with low numeracy skills. *Cogn Dev*. 2009;24(4):450-472. doi:10.1016/j.cogdev.2009.09.003
37. Schrank FA, Wendling BJ. The Woodcock–Johnson IV. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. 2018
38. Wagner R, Torgesen J, Rashotte C, Pearson NA. *CTOPP-2: Comprehensive Test of Phonological Processing–Second Edition*. Austin, TX: Pro-ed; 1999.
39. Dunn LM, Dunn LM. Peabody picture vocabulary test. Pearson Assessments. Minneapolis, MN. 2007.
40. <https://www.analyticmeasures.com/moby-read>
41. Del Tufo SN, Frost SJ, Hoeft F, et al. Neurochemistry Predicts Convergence of Written and Spoken Language: A Proton Magnetic Resonance Spectroscopy Study of Cross-Modal Language Integration. *Front Psychol*. 2018;9:1507. Published 2018 Sep 4. doi:10.3389/fpsyg.2018.01507
42. Siegelman N, Rueckl JG, Steacy LM, et al. Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. *J Mem Lang*. 2020;114:104145. doi:10.1016/j.jml.2020.104145
43. Connolly AJ, Nachtman W, Pritchett EM. Keymath diagnostic arithmetic test. *Keymath diagnostic arithmetic test*. 1976
44. Kronschnabel, J., Brem, S., Maurer, U., & Brandeis, D. (2014). The level of audiovisual print-speech integration deficits in dyslexia. *Neuropsychologia*, 62, 245–261. <https://doi.org/10.1016/j.neuropsychologia.2014.07.024>

45. Hämäläinen JA, Landi N, Loberg O, Lohvansuu K, Pugh K, Leppänen PHT. Brain event-related potentials to phoneme contrasts and their correlation to reading skills in school-age children. *Int J Behav Dev.* 2018;42(3):357-372. doi:10.1177/0165025417728582

Attachment A: Table of Study Measures

Domain	Study Measure	Abbreviation	Subtests	Description	Collection Timepoints		References
					Group 1	Group 2	
Standardized Measures of Reading and Language Skills	Test of Word Reading Efficiency	TOWRE 2	Form A: Phonemic Decoding Efficiency (PDE) Sight Word Efficiency (SWE)	The Test of Word Reading Efficiency Second Edition (TOWRE 2) is a measure of an individual's ability to pronounce printed words (Sight Word Efficiency) and phonemically regular non-words (Phonemic Decoding Efficiency) accurately and fluently. The TOWRE 2 provides an efficient means of monitoring the growth of two kinds of word reading skill that are critical in the development of overall reading ability. The Sight Word Efficiency (SWE) subtest assesses the number of real words printed in vertical lists that an individual can accurately identify within 45 seconds. The Phonemic Decoding Efficiency (PDE) subtest measures the number of pronounceable non-words presented in vertical lists that an individual can accurately decode within 45 seconds.	T1	T1	Torgesen JK, Wagner R, Rashotte C. Test of Word Reading Efficiency:(TOWRE-2). Test of Word Reading Efficiency:(TOWRE-2). 2012.
			Form B: Phonemic Decoding Efficiency (PDE) Sight Word Efficiency (SWE)		T1.5 T2 T3	T1.5 T2 T3 T4	
	Woodcock-Johnson IV Tests of Achievement	WJ IV ACH	Letter-Word Identification Word Attack Spelling	The Woodcock-Johnson IV Tests of Achievement (WJ IV ACH) measures academic achievement in reading, mathematics, writing, and cross-domain cultures. Letter-Word Identification is part of the Standard Reading Battery and measures an individual's ability to identify letters and words. Word Attack reflects the importance of mapping phonology onto orthography in the development and evaluation of Basic Reading Skills.	T1 T2 T3	T1 T2 T3 T4	Schrank FA, Wendling BJ. The Woodcock-Johnson IV. Contemporary Intellectual Assessment: Theories, Tests, and Issues. 2018.
	Comprehensive Test of Phonological Processing	CTOPP 2	Phonological Awareness Nonword Repetition Rapid Naming	The Comprehensive Test of Phonological Processing (CTOPP 2) is a measure of phonological awareness, phonological memory, and naming. The CTOPP 2 can be used to help evaluate phonological processing abilities as a prerequisite to reading fluency.	T1 T2 T3	T1 T2 T3 T4	Wagner R, Torgesen J, Rashotte C, Pearson NA. CTOPP-2: Comprehensive Test of Phonological Processing-Second Edition. Austin, TX: Pro-ed; 1999.

Domain	Study Measure	Abbreviation	Subtests	Description	Collection Timepoints		References
					Group 1	Group 2	
	Peabody Picture Vocabulary Test	PPVT 4		The Peabody Picture Vocabulary Test Fourth Edition (PPVT 4) measures the receptive (hearing) vocabulary of children and adults. It is a norm-referenced, wide-range instrument that is untimed and individually administered. It is available in two parallel forms, designated as Form A and Form B. Each form contains training items followed by 228 test items, each consisting of four full color pictures arranged on a page. The examinee selects the picture that best illustrates the meaning of a stimulus word spoken by the examiner.	T1 T2 T3	T1 T2 T3 T4	Dunn LM, Dunn LM. Peabody picture vocabulary test. Pearson Assessments. Minneapolis, MN. 2007.
Additional Measures of Reading and Language	Moby.Read Oral Reading Fluency Benchmark Assessment	Moby.Read		Moby.Read is a self-administered oral reading fluency (ORF) benchmark assessment in English that runs in a Chrome browser or on an iPad and is automatically scored. Moby.Read uses automatic speech recognition (ASR) and natural language processing (NLP) technologies to score the tests. Test results are available online shortly after the assessment is complete. Test scores and student recordings provide educators with immediate, accurate data that can support effective strategies for improving student performance.	T1 T2 T3	T1 T2 T3 T4	https://www.analyticmeasures.com/moby-read
	Cross-Modal Matching Task	AV		The Cross Modal Matching Task entails a simple cross-modal matching of spoken letters, word, and pseudowords with print targets in a two-alternative forced choice paradigm. On a given trial the participant hears a letter name, a CVC pseudoword, or a CVC word followed by two printed tokens. The participant chooses the matching token and both latency and accuracy are measured and aggregated by condition and by confusability of foils (as foils are either highly or weakly confusable with the target).	T1 T2 T3	T1 T2 T3 T4	Del Tufo SN, Frost SJ, Hoeft F, et al. Neurochemistry Predicts Convergence of Written and Spoken Language: A Proton Magnetic Resonance Spectroscopy Study of Cross-Modal Language Integration. Front Psychol. 2018;9:1507. Published 2018 Sep 4. doi:10.3389/fpsyg.2018.01507

Domain	Study Measure	Abbreviation	Subtests	Description	Collection Timepoints		References
					Group 1	Group 2	
	Strain Naming Task	Strain		The Strain Naming task is a word naming task with with a manipulation of items' frequency, imageability, and consistency. It has a focus on individuals' reliance on regularities in the mapping between print and speech versus associations between print and meaning. The word naming task includes 160 trials presented to children in a fixed order. In each trial, a fixation cross appears for 500 ms, and then is replaced by a monosyllabic word presented in the center of the screen until response. Participants are asked to read out loud each word as accurately and quickly as possible.	T1 T2 T3	T1 T2 T3 T4	Siegelman N, Rueckl JG, Steacy LM, et al. Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. J Mem Lang. 2020;114:104145. doi:10.1016/j.jml.2020.104145
Standardized Measures of Math	KeyMath 3 Diagnostic Assessment	KeyMath	Numeracy Addition and Subtraction Mental Computation and Estimation Applied Problem Solving Foundations of Problem Solving Algebra Geometry Measurement Data Analysis and Probability	The KeyMath 3 Diagnostic Assessment is a comprehensive, norm-referenced measure of essential mathematical concepts and skills. This untimed, individually administered assessment covers the complete range of math concepts and skills, from early experiences with rote and rational counting through experiences with factoring polynomials and solving linear equations.	T1 T2 T3	T1 T2 T3 T4	Connolly AJ, Nachtman W, Pritchett EM. Keymath diagnostic arithmetic test. Keymath diagnostic arithmetic test. 1976.
	Audio Visual Letter Sounds Task	AV EEG		The Audio Visual (AV) Letter Sounds Task consists of 570 trials. On each trial, participants see a letter and/or hear the name of a letter. Stimuli are presented in four conditions: (a) Visual-only (e.g., d), (b) Auditory-only ("dee"), (c) AV-congruent (d+"dee"), and (d) AV-incongruent (b+"dee"). During the task, participants monitor for the letter n, pressing a button whenever they see n and/or hear "en".	T1 T2 T3	T1 T2 T3 T4	Kronschabel, J., Brem, S., Maurer, U., & Brandeis, D. (2014). The level of audiovisual print-speech integration deficits in dyslexia. Neuropsychologia, 62, 245–261. https://doi.org/10.1016/j.neuropsychologia.2014.07.024

Domain	Study Measure	Abbreviation	Subtests	Description	Collection Timepoints		References
					Group 1	Group 2	
EEG/ERP Measures	Mismatch Negativity Paradigm Speech Sound Discrimination Task	MMN EEG		To index phoneme discrimination, a precursor to successful reading that increases with reading skill, two standard/deviant oddball paradigms will be utilized to elicit mismatch negativity (MMN) responses. MMN responses are sensitive to reading skill, predictive of longitudinal reading outcomes, and predict pre-to-post treatment response. In this experiment, participants passively listen to CVs that differ in voice onset time (VOT) and place of articulation (/ba/ vs. /pa/ and /ga/ vs. /da/).	T1 T2 T3	T1 T2 T3 T4	Guttorm TK, Leppänen PH, Poikkeus AM, Eklund KM, Lyytinen P, Lyytinen H. Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. <i>Cortex</i> . 2005;41(3):291-303. doi:10.1016/s0010-9452(08)70267-3
Questionnaires	COVID-19 Home Literacy Study			The COVID-19 Home Literacy Study from GAAB LAB is an online questionnaire administered to learn more about the effects of COVID-19 on literacy environment at home.	T1	T1	https://www.gaablab.com/covid-19-hle

Attachment B: Table of Questionnaires

Below is a comprehensive list of the questionnaires we will administer to participants, their parents/guardians, and teachers if applicable, upon enrolling in the study.

Child Battery					
Measure	Abbreviation	Description	Reference	Age	Time to Complete
NIH Toolbox	NIHTX	NIH Toolbox is a multidimensional set of brief measures assessing cognitive, emotional, motor, and sensory function from ages 3 to 85, meeting the need for a standard set of measures that can be used as a "common currency" across diverse study designs and settings.	Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH Toolbox for Assessment of Neurological and Behavioral Function. <i>Neurology</i> , 80, S2-S6.	All participants	30 minutes
Edinburgh Handedness	Edinburgh	The Edinburgh Handedness Inventory is a measurement scale used to assess the dominance of a person's right or left hand in everyday activities. The inventory can be used by an observer assessing the person, or by a person self-reporting hand use.	Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. <i>Neuropsychologia</i> , 9(1), 97-113.	All participants	5 minutes
Youth Self Report	YSR	Youth (ages 11-18) rate themselves for how true each item is now or has been during the past 6 months, using the same 3-point scale as the CBCL 6-18 and the TRF. YSR also has 14 socially desirable items that most youths endorse about themselves.	Achenbach, T. M. (1991). Manual for the Youth Self-Report and 1991 profile. Burlington, VT: University of Vermont, Department of Psychiatry.	Participants ages 11	10 minutes
Mood and Feelings Questionnaire – Long Version	MFQ	The MFQ consists of a series of descriptive phrases regarding how the subject has been feeling or acting recently. Codings reflect whether the phrase was descriptive of the subject most of the time, sometimes, or not at all in the past two weeks.	Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F., & Silver, D. (1995) The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. <i>International Journal of Methods in Psychiatric Research</i> , 5, 237 - 249.	Participants ages 8+	10 minutes

Screen for Child Anxiety Related Disorders - Self-report	SCARED	The SCARED is a child and parent self-report instrument used to screen for childhood anxiety disorders including general anxiety disorder, separation anxiety disorder, panic disorder, and social phobia. In addition, it assesses symptoms related to school phobias.	Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. <i>Journal of the American Academy of Child and Adolescent Psychiatry</i> , 38(10), 545-553.	Participants ages 8+	10 minutes
Conners ADHD Rating Scales - Self Report Short Form	Conners	The Conners ADHD Rating Scale is an instrument that uses self-report ratings to help assess attention-deficit/hyperactivity disorder (ADHD) and evaluate problem behavior in children and adolescents.	Conners, C.K. (2001). <i>Conners' Rating Scales-Revised</i> . North Tonawanda, NY: Multi-Health Systems, Inc.	Participants ages 8+	10 minutes
Extended Strengths and Weaknesses Assessment of Normal Behavior-Self Report	E-SWAN Self	The E-SWAN is a self-report instrument used to help assess multiple DSM Disorders and evaluate problem behavior in children and adolescents. The disorders include Depression, Disruptive Mood Dysregulation Disorder (DMDD), Social Anxiety, and Panic Disorder.	Developed by the Child Mind Institute	All participants	15 minutes
GRIT Scale	GRIT	The GRIT Scale measures trait-level perseverance and passion for long-term goals.	Duckworth, A.L., Peterson, C., Matthews, M.D., & Kelly, D.R. (2007). Grit: Perseverance and passion for long-term goals. <i>Journal of Personality and Social Psychology</i> , 9, 1087-1101. http://www.sas.upenn.edu/~duckworth/images/Grit%20JPSP.pdf	Participants ages 8+	3 minutes
General Self Efficacy	GSE	The scale was created to assess a general sense of perceived self-efficacy with the aim in mind to predict coping with daily hassles as well as adaptation after experiencing all kinds of stressful life events.	Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, <i>Measures in health psychology: A user's portfolio. Causal and control beliefs</i> (pp. 35-37). Windsor, UK: NFER-NELSON.	Participants ages 8+	3 minutes

Internet Addiction Test	IAT	The Internet Addiction Test (IAT) is a 20-item scale that measures characteristics and behaviors associated with compulsive use of the Internet that includes compulsivity, escapism, and dependency. Questions also assess problems related to addictive use in personal, occupational, and social functioning.	Young, K. S. (1998) Internet addiction: The emergence of a new clinical disorder. <i>CyberPsychology and Behavior</i> , 1(3), 237-244.	All participants	5 minutes
Internet Use Questionnaire	IUQ SR	The internet use questionnaire assesses which types of activities participants engage in online, and the frequency of participation in these activities	Project Developed	All participants	5 minutes
Alabama Parenting Questionnaire	APQ	The APQ measures is a 42-item questionnaire measuring five domains of parenting that are relevant to the etiology and treatment of child externalizing problems: positive involvement, supervision/monitoring, use of positive discipline, consistency in the use of such discipline, and use of corporal punishment.	Essau, C.A., Sasagawa, S., & Frick, P.J. (2006). Psychometric properties of the Alabama Parenting Questionnaire. <i>Journal of Child and Family Studies</i> , 15, 597-616.	All participants	15 minutes

Parent/Guardian Battery					
Measure	Abbreviation	Description	Reference	Age	Time to Complete
Barratt Simplified Measure of Social Status	Barratt	This measure is built on the work of Hollingshead (1957, 1975) who devised a simple measure of Social Status based on marital status, retired/employed status (retired individuals used their last occupation) educational attainment, and occupational prestige. This is a measure of social status, which is a proxy for socio-economic status. This is not a measure of social class, which is best seen as a cultural identity.	Barratt, W. (2006). The Barratt simplified measure of social status (BSMSS): Measuring SES. Unpublished manuscript. http://socialclassoncampus.blogspot.com/2012/06/barratt-simplified-measure-of-social.html	Parents/guardians of all participants	5 minutes

Child Behavior Checklist	CBCL	The CBCL is a device by which parents or other individuals who know the child well rate a child's problem behaviors and competencies. The CBCL can also be used to measure a child's change in behavior over time or following a treatment. It consists of 118 items related to behavior problems, which are scored on a 3-point scale ranging from not true to often true of the child. There are also 20 social competency items used to obtain parents' reports of the amount and quality of their child's participation in sports, hobbies, games, activities, organizations, jobs and chores, friendships, how well the child gets along with others and plays and works by him/herself, and school functioning.	Achenbach, T. M. (1991). Integrative Guide to the 1991 CBCL/4-18, YSR, and TRF Profiles. Burlington, VT: University of Vermont, Department of Psychology.	Parents/guardians of all participants	20 minutes
Mood and Feelings Questionnaire	MFQ	The MFQ consists of a series of descriptive phrases regarding how the subject has been feeling or acting recently. Codings reflect whether the phrase was descriptive of the subject most of the time, sometimes, or not at all in the past two weeks.	Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F., & Silver, D. (1995). The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. <i>International Journal of Methods in Psychiatric Research</i> , 5: 237-249.	Parents/guardians of all participants	10-15 minutes
Screen for Child Anxiety Related Disorders - Parent report	SCARED	The SCARED is a child and parent self-report instrument used to screen for childhood anxiety disorders including general anxiety disorder, separation anxiety disorder, panic disorder, and social phobia. In addition, it assesses symptoms related to school phobias.	Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. <i>Journal of the American Academy of Child and Adolescent Psychiatry</i> , 38(10), 1230-1236.	Parents/guardians of all participants	10 minutes
Extended Strengths and Weaknesses Assessment of Normal Behavior-Parent Report	E-SWAN Parent	The E-SWAN is a parent report instrument used to help assess multiple DSM Disorders and evaluate problem behavior in children and adolescents. The disorders include Depression, Disruptive Mood Dysregulation Disorder (DMDD), Social Anxiety, and Panic Disorder.	Developed by the Child Mind Institute	Parents/guardians of all participants	15 minutes

Autism Spectrum Screening Questionnaire	ASSQ	The ASSQ is a 27-item checklist for completion by lay informants when assessing symptoms characteristic of Asperger's Syndrome or high-functioning autism spectrum disorders.	Ehlers S., Gillberg C., Wing, L. (1999). A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. <i>Journal of Autism Developmental Disorders</i> 29, 129-41.	Parents/guardians of all participants	15 minutes
Social Communication Questionnaire	SCQ	The SCQ consists of forty yes-or-no questions, which a parent can complete in around ten minutes. It is brief and easy to understand, yet provides valuable information on a child's body movements, use of language or gestures, and style of interacting. Usually, the SCQ is used as a screening instrument to help identify which children may have an autism spectrum disorder. It is not meant to provide a detailed diagnosis, but to indicate whether a child needs a more careful and in-depth evaluation.	Chandler, S., Charman, T., Baird, G., et al (2007). Validation of the Social Communication Questionnaire in a population cohort of children with autism spectrum disorders. <i>Journal of the American Academy of Child and Adolescent Psychiatry</i> , 46, 1324 -1332.	Parents/guardians of all participants	10 minutes
Social Aptitudes Scale	SAS	The SAS is a 10-item parent-report scale that assesses a child's behavior in different social situations compared to other children of the same age. Items are rated from above to below average.	Liddle, Elizabeth B., Martin J. Batty, and Robert Goodman. "The social aptitudes scale: an initial validation." <i>Social psychiatry and psychiatric epidemiology</i> 44.6 (2009): 508-513.	Parents/guardians of all participants	5 minutes
WHO Disability Assessment Schedule – Parent Report	WHODAS-P	The WHODAS is a generic assessment instrument for health and disability. It assesses function across domains of cognition, mobility, self-care, getting along, life activities, and participation.	Gold, Liza H. "DSM-5 and the assessment of functioning: the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0)." <i>Journal of the American Academy of Psychiatry and the Law Online</i> 42.2 (2014): 173-181.	Parents/guardians of all participants	5 minutes
Strengths and Difficulties Questionnaire (with impact supplement)	SDQ	A brief behavioral screening questionnaire with 25 items on psychological attributes divided between 5 scales. The extended version includes an impact supplement that asks if the caregiver if the participant has experienced a problem and the degree of chronicity, distress, social impairment, and burden to others. The questionnaire is administered during the initial telephone screen with the caregiver of a child or	Goodman, R. 1997. The Strengths and Difficulties Questionnaire: A Research Note. <i>Journal of Child Psychology and Psychiatry</i> , 38, 581-586; Goodman, R. 1999.	Parents/guardians of all participants	10 minutes

		adolescent participant. For clinical assessment, SDQ helps to determine how the protocol assessments are completed and the professionals involved.			
Social Responsiveness Scale-2	SRS-2	The second edition of the widely administered Social Responsiveness Scale (SRS-2; Constantino & Gruber, 2012) maintains continuity with the original instrument as an efficient quantitative measure of the various dimensions of interpersonal behavior, communication, and repetitive/stereotypic behavior characteristic of autism spectrum disorders (ASD).	Constantino, J.N., Davis, S.A., Todd, R.D., Schindler M.K., Gross, M.M., Brophy, S.L., et al. (2003). Validation of a brief quantitative measure of autistic traits: comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. <i>Journal of Autism Developmental Disorders</i> . 33, 427-433.	Parents/guardians of all participants	15 minutes
Affective Reactivity Index-Parent	ARI-P	The Affective Reactivity Index- Self Report is a concise scale for the dimensional assessment of irritability.	Stringaris, A., Goodman, R., Ferdinando, S., Razdan, V., Muhrer, E., Leibenluft, E., & Brotman, M. (2012). The Affective Reactivity Index: A concise irritability scale for clinical and research settings. <i>Journal of Child Psychology and Psychiatry</i> . 53(11) pp. 1109-1117.	Parents/guardians of all participants	5-7 minutes
Parent-Child Internet Addiction Test	PCIAT	The Internet Addiction Test (IAT) is a 20-item scale that measures characteristics and behaviors associated with compulsive use of the Internet that include compulsivity, escapism, and dependency. Questions also assess problems related to addictive use in personal, occupational, and social functioning.	Young, K. S. (1998) Internet addiction: The emergence of a new clinical disorder. <i>CyberPsychology and Behavior</i> , 1(3), 237-244.	Parents/guardians of all participants	5-10 minutes
Internet Use Questionnaire	IUQ P	The internet use questionnaire assesses which types of activities participants engage in online, and the frequency of participation in these activities	Project Developed	Parents/guardians of all participants	5 minutes

Positive Behavior Scale	PBS	The Positive Behavior Scale was developed for the New Chance study to assess positive behaviors in a population of educationally and economically disadvantaged children and youth. The scale, which includes 25 items, is divided into three subscales: (a) social competence and sensitivity (e.g., gets along well with other children, shows concern for other people's feelings); (b) compliance and self-control (e.g., thinks before he/she acts, usually does what I tell him/her); and (c) autonomy (e.g., tries to do things for him/herself, is self-reliant).	Quint, J. C., Bos, J. M., & Polit, D. F. (1997). New Chance. Final Report on a Comprehensive Program for Young Mothers in Poverty and Their Children.	Parents/guardians of all participants	5 minutes
Child Flourishing Scale	CFS	The child flourishing scale measures indicators of positive development of children and adolescents.	National Survey of Children's Health 2016: http://childhealthdata.org/learn-about-the-nsch/topics_questions/2016-nsch-guide-to-topics-and-questions	Parents/guardians of all participants	1 minute
Alabama Parenting Questionnaire – Parent Report	APQ – Parent	The APQ measures is a 42-item questionnaire measuring five domains of parenting that are relevant to the etiology and treatment of child externalizing problems: positive involvement, supervision/monitoring, use of positive discipline, consistency in the use of such discipline, and use of corporal punishment.	Essau, C.A., Sasagawa, S., & Frick, P.J. (2006). Psychometric properties of the Alabama Parenting Questionnaire. <i>Journal of Child and Family Studies</i> , 15, 597-616.	Parents/guardians of all participants	15 minutes
Negative Life Events Scale – Parent Report	NLES - Parent	This scale focuses on family, community, and school-based stressors that children may be exposed to and are known to be related to negative physical and mental health outcomes. Parents are asked to rate whether the event occurred, whether the child was aware, and the child's level of resulting stress. Administration of this checklist includes the General Coping Efficacy sub-scale.	Sandler, I.N., Wolchik, S.A., Braver, S.L. & Fogas, B. (1991). Stability and quality of life events and psychological symptomatology in children of divorce. <i>American Journal of Community Psychology</i> , 19(4), 501-520.	Parents/guardians of all participants	10 minutes

Distress Tolerance Scale	DTS	This 15-item self-report questionnaire measures emotional distress tolerance.	Simons & Gaher. (2005) The distress tolerance scale: development and validation of a self-report measure. <i>Motivation and Emotion</i> , 29(2), 83-102.	Parents/guardians of all participants	5 minutes
Parenting Stress Index Fourth Edition	PSI	This inventory is designed to evaluate the magnitude of stress in the parent-child system, focusing on three major domains of stress: (1) child characteristics, 2) parent characteristics, and 3) situational/demographic life stress.	Abidin, R.R. (2012). <i>Parenting stress index, fourth edition (PSI-4)</i> . Lutz, FL: Psychological Assessment Resources.	Parents/guardians of all participants	15minutes

Teacher Battery					
Measure	Abbreviation	Description	Reference	Age	Time to Complete
Child Behavior Checklist – Teacher Report Form	CBCL - TRF	The CBCL is a device by which parents or other individuals who know the child well rate a child's problem behaviors and competencies. The CBCL can also be used to measure a child's change in behavior over time or following a treatment. It consists of 118 items related to behavior problems, which are scored on a 3-point scale ranging from not true to often true of the child. There are also 20 social competency items used to obtain parents' reports of the amount and quality of their child's participation in sports, hobbies, games, activities, organizations, jobs and chores, friendships, how well the child gets along with others and plays and works by him/herself, and school functioning.	Achenbach, T. M. (1991). <i>Integrative Guide to the 1991 CBCL/4-18, YSR, and TRF Profiles</i> . Burlington, VT: University of Vermont, Department of Psychology.	Teachers of all participants	20 minutes
Strengths and Weaknesses of Attention Deficit/Hyperactivity Disorder Symptoms and Normal Behavior Scale	SWAN-TRF	The SWAN ADHD Rating Scale is an instrument that uses observer ratings to help assess attention-deficit/hyperactivity disorder (ADHD) and evaluate problem behavior in children and adolescents.	Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., . . . McCleary, R. (2001). The SWAN rating scale. Available from http://www.adhd.net .	Teachers of all participants	10 minutes

