Neural Mechanisms of Successful Intervention in Children with Dyslexia

Study Protocol and Statistical Analysis Plan

NCT04323488

October 29, 2018

SPECIFIC AIMS - Neural mechanisms of successful intervention in children with dyslexia

Reading instruction prompts the emergence of neural circuits that are specialized for rapidly translating printed symbols into sound and meaning. Understanding how these circuits differ in children with dyslexia, and change with learning, is an important scientific challenge that holds practical implications for education. The proposed research employs frequent longitudinal measurements over an intensive summer intervention program for children with dyslexia to: (a) determine how brain structure and function change in response to reading instruction; and (b) investigate neurobiological factors that predispose a child to struggle or succeed in the intervention. Thus, this proposal seeks to determine both how education shapes brain development, and how a child's unique neurobiology predicts educational outcomes.

Aim 1) White matter plasticity and learning - White matter was previously considered static infrastructure; it is now known that the thickness of the myelin sheath, the caliber of axons, and density of glial cells change with learning. Recent studies in mouse models demonstrate the central role of white matter plasticity in the learning process, and highlight the importance of understanding white matter plasticity in humans[1–3]. Reading interventions provide a powerful tool to study experience-dependent plasticity in the human brain. Based on novel quantitative MRI and diffusion MRI techniques developed by the PI and collaborators[4–8], it is now possible to quantify changes in cell density, intra-axonal water and myelination at millimeter resolution.

Approach: Forty children with dyslexia (ages 9-11y), will be recruited for a highly effective intervention program that involves **intensive** (160 hours over 8 weeks) one-on-one training in phonological and orthographic processing skills[9–11]. Based on dense longitudinal measurements collected before, during, and after the intervention, we will model the time-course of white matter plasticity associated with improvements in reading skills and investigate the biological mechanisms that underlie differences in learning among children. **Aim 1** targets three questions: ***1a:*** *Does intervention cause short-term, transient changes in the white matter, or long-term remodeling of the brain circuit?* ***1b:*** *Do different cellular properties of the white matter display different time-courses of plasticity?* ***1c:*** *Is white matter plasticity coupled to behavioral improvement?*

Aim 2) Bottom-up and top-down computations in the reading circuit - When our eyes fixate upon a word, a cascade of neural processes is initiated, beginning in the visual system and progressing through a series of computations that translate the visual representation into sound and meaning. The visual word form area (VWFA)[12–15] is the intersection of vision and language, and has direct connections to visual cortex and language areas. Based on meta-analysis[16], the VWFA is the most common location of neural deficits in dyslexia. We have developed the first model of neural computations performed by the VWFA, characterizing both: (a) the automatic, or bottom-up, response of neurons that encode words, and (b) the effect of top-down signals that modulate the VWFA response[17]. Based on this model, we will investigate how computations and connectivity of the VWFA differ in children with dyslexia, and change in response to reading instruction.

Approach: FMRI experiments measuring stimulus-driven (bottom-up) neural tuning properties and top-down, task-dependent modulation of the VWFA (adapted from ref[17]) will be collected longitudinally over the course of the intervention program. By measuring changes in neural computations over a period of targeted training, while tightly controlling both the stimulus properties and cognitive task demands, we will address three broad questions: ***2a:*** *To what extent do VWFA deficits, and learning-induced changes, reflect bottom-up neural tuning properties versus functional connectivity between visual and phonological processing regions?* ***2b:*** *Do VWFA changes occur in concert with changes in phonological processing regions?* ***2c:*** *Do changes in white matter connectivity increase the efficiency of top-down signals and improve functional connectivity between visual and phonological processing regions?*

Aim 3) Neural biomarkers of learning outcomes - Even in a controlled and intensive learning environment, some children show substantial (>2 SD) improvements in reading skill, while others show limited change. What biological factors predispose a child to excel or struggle when provided a high-quality intervention? Data from multiple labs show converging evidence that tissue properties of the arcuate fasciculus predict the likelihood that a child will struggle learning to read. However, one of the limitations of past research is the lack of control over educational experience, a problem that is addressed by the summer intervention program employed here.

Approach: Pre-intervention MRI measurements will be examined as predictor variables (in combination with behavioral measures) for individual differences in intervention learning rate, and long-term, post-intervention outcomes. Previous research and pilot data support 3 hypotheses: ***3a:*** *Poor integrity of the arcuate fasciculus will predict resistance to remediation above and beyond behavioral measures.* ***3b:*** *The severity of neural processing deficits in components of the reading circuitry will predict learning outcomes.* ***3c:*** *Combining behavioral measures and data from the different MRI modalities with machine learning will lead to more accurate predictions than any single modality on its own.*

Significance

Overview: The proposed experiments combine a highly effective dyslexia intervention program[9,10] with a sequence of cutting-edge neuroimaging measurements to study the neurobiology of learning in the context of a controlled and intensive educational program. The methodology builds from measurements of brain circuit structure (**Aim** 1), to neural computation (**Aim 2**), and then combines the data from the first two aims to examine the predictive value of brain measurements as biomarkers for future learning (**Aim 3**). Integrating these three measurement modalities – behavior, structure and function – will motivate a new circuit-level understanding of the neurobiological basis of learning to read, and the underpinnings of learning difficulties in dyslexia.

Promoting the virtuous cycle between neuroscience and education: Interdisciplinary research at the intersection of education and neuroscience promises to catalyze discovery in both fields. On the one hand, neuroimaging is a powerful tool for understanding the mechanisms that underlie the successful treatment of learning disabilities such as dyslexia. A deeper understanding of the neurobiological underpinnings of individual differences in learning will pave the way for innovative treatment programs that are personalized to a child's unique pattern of brain maturation. On the other hand, intervention programs offer a powerful tool for studying basic mechanisms of experience-dependent plasticity in the human brain and making causal inferences about how changes to the environment prompt reorganization of brain circuits. Much of what we know about plasticity comes from research in model organisms that show dramatically different developmental time-courses from humans. By following children longitudinally through targeted education programs, we can answer fundamental questions about how experience shapes brain development and forge new links between neuroscience theory and education practice.

Understanding individual differences in reading outcomes: Learning to read is at the foundation of academic success. Formal reading instruction begins in kindergarten and by fourth grade every child is expected to fluidly decode written text into sound and meaning. While the focus of the first few years of elementary school is learning to read, children are quickly expected to use reading as a tool for learning across all academic domains: math is taught through word problems, history through textbooks, and reasoning skills through analyzing texts. Children who struggle learning to read quickly find themselves struggling across academic domains[18].

Fortunately, scientific research on the mechanisms underlying dyslexia has led to the development of effective intervention programs to improve readings skills in children with dyslexia[19–29]. Even though interventions are successful at the group level, there is always a subset of "*treatment resisters*" who do not show substantial improvements in reading scores[30]. On the other end of the spectrum, some children show rapid and transformative improvement in reading skill during an intervention. Behavioral measures have been established as useful predictors of intervention outcomes (for review see[31,32]), but even after accounting for baseline behavioral differences, there is still substantial unexplained variance. This fact has prompted great interest in exploring brain imaging measurements as potential biomarkers of children's future learning trajectories[33–43]. Here, we explore the hypothesis that an individual's response to intervention can be accurately predicted based on properties of their reading circuitry, and that *treatment resisters* can be identified based on specific deficits that impede the learning process. By collecting pre-intervention measurements of white matter connectivity (**Aim 1**), and neural response properties (**Aim 2**), we will endeavor to build a model of the subject-specific characteristics that predict success in the intervention program (**Aim 3**). Our approach will leverage the extensive behavioral literature[31,32,44] and test the hypothesis that specific neural biomarkers will improve prediction accuracy. Such a model holds practical value given that the costs of an intensive intervention program far exceeds that of an MRI scan. Moreover, by testing specific biomarkers from the literature in the context of an intervention study, we are uniquely poised to determine the factors that predict individual differences in learning, when children are in a controlled, and high-quality educational setting.

The role of white matter in reading development: Seminal work by Klingberg and colleagues first demonstrated the relationship between white matter tissue properties and reading skill[45]. Subsequently, dozens of studies have examined correlations between diffusion MRI measurements of the white matter and reading skill[33,35–38,42,46–56]. Based on this extensive literature, there are two fundamental, unanswered questions. First, do anatomy-behavior correlations reflect static traits that differentiate individuals, or do correlations arise due to differences in educational experience/environment? Our recent work suggests that the white matter is surprisingly plastic (see **Aim 1** pilot data), opening the possibility that anatomy-behavior correlations emerge as temporary states in a highly dynamic system[11]. And yet, baseline diffusion measurements also predict intervention outcomes (**Aim** 3 pilot data). The proposed research will clarify the role of educational experience

in shaping white matter tissue properties. Second, what is the underlying biology relating diffusion MRI measures to behavior? Diffusion measurements are incredibly sensitive to individual differences in tissue structure, but are not specific to any one biological property[35,57–60]. Based on our innovative quantitative MRI measurement protocol (**Aim 1**), we will clarify the biological source of white matter-behavior correlations, and group differences in dyslexia. This will have a significant impact on the reading literature, and more broadly in cognitive neuroscience for interpreting correlations between diffusion properties and behavior and establishing closer links to research in animal models.

Open science: The PI has made a strong commitment to open science through the development and support of one of the most widely used, open-source software packages for quantitative analysis of white matter tissue properties[6], and the release of analysis code and data to reproduce findings from published work in his lab[4,9–11,17,61]. All data from this grant will be made publicly available in order to facilitate scientific reproducibility, and to allow other researchers to tackle new and innovative questions that were not conceptualized in the original proposal[62]. This will represent one of the most innovative and significant public datasets targeting plasticity and learning and will be valuable to a broad collection of researchers.

Innovation

Understanding how the developing brain builds circuits to rapidly translate printed symbols into meaning is an important scientific challenge with practical implications for education. This proposal capitalizes on a cutting edge suite of measurement techniques and software algorithms that the research team has developed to model the biological processes that underlie learning to read[4–7,63]. Specific innovations in our methodology are highlighted alongside pilot data for each aim. Training the brain to decode text is a powerful paradigm to investigate mechanisms of plasticity. **By employing an intervention that is delivered in an intensive (20 hrs/week) one-on-one setting, we have the opportunity to investigate large-scale plasticity in the white matter that has not been observed in less intensive intervention paradigms** (see pilot data **Aim 1** and[11]). These measurements will answer fundamental questions about the nature of experience-dependent plasticity in the human brain and generate discoveries that can inform clinical and educational practice. Unlike previous reading intervention research, the primary focus of this proposal is to link learning mechanisms with the broader neuroscience literature; through close collaboration and data sharing with Ken Pugh (Haskins) and Richard Wagner (Florida State) we will also relate these data to cognitive models of the reading architecture.

Approach

Overview: The goal of this proposal is not to determine the efficacy of any specific intervention program but, rather, to capitalize on an intervention with proven efficacy as a tool to study the biological underpinnings of learning. Forty children with dyslexia will be recruited to participate the Lindamood-Bell *Seeing Stars* intervention, which combines training in phonological and orthographic processing (details can be found in our recent work[9,10] and the published manual[64]). Pilot data on 26 children with dyslexia demonstrate **highly significant** ($p<10^{-10}$) changes in reading skills as indexed by the Woodcock Johnson Basic Reading Skills composite, with average improvements of 0.9 SD over 8 weeks (**Figure 1**). Reading automaticity (Test of Word Reading Efficiency (TOWRE) $p<10^{-6}$) and Reading Fluency ($p<10^{-5}$) also show highly significant change, confirming the generalizability of the intervention.

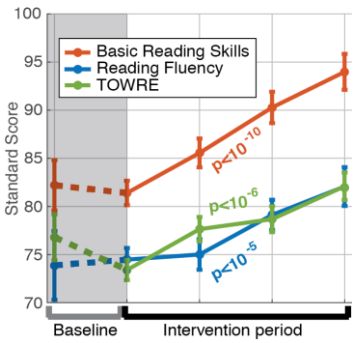

Figure 1: Reading skills improve substantially and systematically over the course of the *Seeing Stars* intervention and are stable during the baseline period.

Each child will participate in six measurement sessions. Two sessions will occur prior to the intervention and will serve as a control period (individual baseline); one session will be conducted at the mid-point of the intervention; one immediately post-intervention; two long-term follow up sessions will be conducted at 6- and 12-months post-intervention. Multiple baseline sessions will allow each child to serve as their own control, and intervention-driven change will be compared to changes observed during the control period[65]. For the aims of this proposal, multiple baseline sessions are preferable to a wait-list control group because it makes it possible to control for developmental effects within the same subjects that participate in the intervention. For example, given the focus on individual differences (**Aim 3**), it is important to consider each subject's baseline (pre-intervention) growth rate as a potential covariate for differences in growth during the intervention period. Additionally, two control groups (see below) will serve as a comparison for long-term intervention effects.

Behavioral assessment: The Woodcock Johnson IV (WJ, Basic Reading Skills and Reading Fluency), TOWRE-2, Gray Oral Reading Tests (GORT-5, Rate, Accuracy, Fluency, Comprehension), and the Test of Silent Reading Efficiency and Comprehension (TOSREC) will be used to assess reading automaticity, accuracy and comprehension at each session. Phonological awareness (PA), phonological memory (PM) and rapid naming (RN) will be assessed with the Comprehensive Test of Phonological Processing (CTOPP). An auditory phoneme categorization task will be used to gain additional insights into phonological processing[66]. Additionally, parent questionaires will be used to assess (1) home literacy environment, education history, and current education experience[67], and (2) ADHD symptoms using the Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior (SWAN) rating scale[68,69].

Participants and experimental design: Subjects will be recruited from the University of Washington Reading and Dyslexia Research Database, a subject pool run by the PI. On average, 125 subjects per year are recruited into the Database, and each subject receives an extensive battery of behavioral tests (including the WJ, TOWRE, CTOPP, and Wechsler Abbreviated Scales on Intelligence (WASI)), MRI screening, and practice at an MRI mock-scanner that includes training on holding still. Additionally, subjects fill out questionnaires on education history, home environment, socioeconomic status (SES) and ADHD symptoms. Although the aims of this proposal do not target the relationship between SES and response to intervention, obtaining this information will allow other researchers to explore this relationship when data are made public (e.g.[70]). This well-characterized subject pool makes it makes it possible to recruit matched samples of intervention and control subjects with specific behavioral and demographic characteristics. In our experience, there is very low subject attrition because all subjects have already visited the lab and are familiar with the research protocol.

Intervention subjects (**N=40**) will be children between **nine and eleven years of age** who are below the 25th percentile in timed (TOWRE composite) **and** untimed (WJ Basic Reading Skills) single word decoding abilities, and within the normal range (+/- 1 SD) on measures of general cognitive abilities (WASI). Diagnosis of ADHD will be an exclusion criterion due to concerns over motion. Pilot measurements on children 7-12 years of age demonstrate that the intervention is equally effective (d=~.9) across this age range[9]. By focusing on 9-11-year-old children less data will be lost due to motion and subjects will stay focused for longer scan sessions.

Intervention protocol: Intervention will be administered to participants in Lindamood-Bell learning centers in the Seattle area four hours a day, five days a week, for eight consecutive weeks during summer vacation. Trained tutors will work one-on-one with each child. Importantly, the scientific research will be completely independent of the administration of the intervention to eliminate any potential conflicts of interest. Lindamood-Bell will not have access to any information the PI's laboratory collects on the subjects, including subject attrition.

The Lindamood-Bell "*Seeing Stars*" instruction program teaches phonological and orthographic processing through a combination of mental imagery and sensory-motor learning. Children practice visualizing the orthography of words, starting from simple consonant-vowel syllables and systematically working into more complex consonant-vowel-consonant pairings. The instruction focuses on segmenting and blending strings of phonemes into words and visualizing the relationship between the articulatory elements, and their representation as visual symbols. Phonological awareness is systematically built through articulatory exercises where children are taught to attend to the relationship between motor movements of the mouth and tongue, and speech sounds in words, and the corresponding letters. As children build a stronger foundation of phonological and orthographic knowledge, additional practice with increasingly complex sight-word identification and phonological decoding is layered into the intervention.

Intervention fidelity and monitoring: The intervention curriculum is detailed in the publicly available *Seeing Stars* Teacher's Manual allowing the procedure to be precisely reproduced[64]. Fidelity to the published curriculum will be ensured in two ways. First, Lindamood-Bell has established a rigorous methodology for ensuring the quality and fidelity of instruction: (1) Each instructor completes a 3-week training program; (2) At each learning center a "coach" with multiple years of experience monitors each instructor's progress and provides feedback, ensuring that the instructor is properly progressing through the steps of the curriculum with their student; and (3) Each coach provides weekly reports to the regional director to ensure consistency of implementation across learning centers. This fidelity monitoring will be recorded and databased for each research study participant. Second, one session a week (for each child) will be recorded with a digital video camera and a member of the research team (independent of Lindamood-Bell) will score the fidelity of implementation for each session based on: (1) adherence to the procedures of the intervention using a 5-point implementation behavior scale and (2) engagement of the child. Child engagement will be measured using a 5-second partial interval time sampling procedure on the same videotaped session used for procedural fidelity. Reliability procedures will require research staff to rate a sample of training videos and for their ratings to correlate at $r \geq .80$ prior to data collection.

The researcher will provide regular feedback to the Lindamood-Bell regional director based on the independent assessment of fidelity so that instructors can receive coaching in the case of deviations from the published curriculum. Finally, scoring will be databased so that details of the intervention delivery can be examined as covariates in statistical models of intervention effects.

Control subjects: Multiple baseline measurements in each intervention subject will be used to establish the stability of our measurements, and control for the effects of repeated testing over the intervention period. Additionally, **two control groups** will be recruited to assess long-term change in the absence of intervention. Each control group will include 20 subjects **matched in terms of age, gender, non-verbal IQ and socioeconomic status**. One control group will also be matched in terms of reading skills (**dyslexic control**) and will serve as a comparison for post-intervention changes in children with dyslexia. The other control group will consist of typical readers and will serve as a comparison for typical development. Based on the typical-reading control group, we can assess neural differences in the subjects with dyslexia and examine "normalization" of deficits associated with dyslexia. Control groups will participate in four measurements sessions: (1) a baseline session followed by longitudinal visits at (2) two months (length of intervention), (3) eight months and (4) 14 months. Dyslexic control subjects will be offered the opportunity to participate as intervention subjects the following summer.

Statistics: Longitudinal change will be modeled with linear mixed effects (LME) models. LME models are preferable to repeated measures ANOVAs for two reasons: (1) subjects with missing data points do not need to be excluded; (2) non-linear effects can be modeled by adding higher-order polynomials terms. Model selection based on Bayesian information criteria will be used to determine the random effects structure[71–73].

Aim 1: White matter plasticity and mechanisms of learning

Significance: Interventions are a powerful tool to examine how an experimental manipulation of a child's environment impacts brain development. **Aim 1** seeks to determine how remediation of reading difficulties changes the wiring of the human brain, and to resolve the long-term time-course of plasticity and learning that is set in place by an intensive summer intervention[11]. **Aim 1** will answer basic scientific questions of how experience shapes brain development and generate a new understanding of the mechanisms underlying dyslexia remediation. Investigating how experience changes the white matter will provide critical data to interpret the hundreds of studies demonstrating correlations between white matter tissue properties and behavioral measures, and the dozens of studies showing correlations specific to reading skills (for reviews see[35,74–79]).

Innovation: Over the past decade, non-invasive techniques to measure the living human brain have dramatically improved. Research on experience-dependent plasticity that was once only feasible in animal models can now be conducted in living and behaving human subjects. Properties of cellular organization including the density of tissue macromolecules and the concentration of myelin can be accurately estimated using recently developed quantitative MRI techniques employed here[4,5]. Now, for the first time, we can begin to develop models of the biological processes that underlie the development of uniquely human skills such as reading. Here we harness new innovations in diffusion MRI and quantitative MRI acquisitions, combine these measures with biophysical modeling, and collect measurements longitudinally over a tightly controlled educational intervention, to pioneer a new understanding of the neurobiology of learning. The PI's laboratory is uniquely poised to implement these cutting-edge measurement techniques in an intervention study.

Relation to previous work: Keller and Just (2009) were the first to report intervention-driven changes in the white matter, establishing the feasibility of measuring white matter plasticity during a reading intervention[80]. But a number of questions were left unanswered by this original work: First, the behavioral effects were relatively small (significant improvement in pseudo-word but not real-word reading or comprehension), and the white matter changes were localized to a small region and not in any of the core reading tracts. Therefore, white matter plasticity is likely to be much more extensive during an intensive intervention program (see **Figure 2** and ref[11]). Also, the intervention was carried out over an extended period of time (six months during regular schooling), which introduces maturational and environmental confounds. Third, quantitative MRI methods capable of resolving different biological mechanisms in the white matter were not available at the time. Finally, long-term follow-ups were not conducted to ascertain the stability of learning effects. Thus, employing an intensive intervention over 8 weeks of summer, combined with innovative, biologically specific imaging methods and long-term follow-up measures and questionnaires will allow us to address significant, unanswered questions.

Quantitative MRI measurements of white matter tissue structure

Diffusion weighted MRI (dMRI) data acquisition – DMRI data will be acquired at 1.5 mm^3 spatial resolution (96 gradient directions, distributed across 2 b-values, b=711 and b=2855 s/mm^2). 64 diffusion directions will be sampled at the high b-value, 32 directions at the low b-value, and 8 images will be acquired without diffusion weighting (b=0). Half the b=0 images will be acquired with a reversed phase encoding direction to correct distortions due to field inhomogeneities[81,82]. This sequence was determined optimal for estimation of parameters of the neurite orientation dispersion and density imaging (**NODDI**) model, based on data from a Philips Achieva scanner[83]. The acquisition optimally balances (a) angular resolution/contrast to estimate fiber orientation distribution functions for tractography[84], (b) multiple shells to estimate tissue properties based on multiple compartment models[83] and (c) signal to noise ratio (SNR) given the time constraints of working with children.

Fiber tractography – Fiber orientation distribution functions will be estimated using constrained spherical deconvolution (CSD[85]) as implemented in the Diffusion in Python (DIPY) software package[86]. Optimal parameters for model fitting will be determined based on cross-validation[8]. Fibers will be estimated using probabilistic tractography. Given concerns over false positives in tractography (erroneous fibers) highlighted by us and others[87], the Linear Fascicle Evaluation (LiFE) algorithm will be used to control for false positives, and retain an optimized set of fibers[7]. Fascicles will then be identified within each individual's brain using the AFQ software package[6], and analyzed in relation to reading skills over the course of the intervention. AFQ measurements are highly reliable ($r = 0.93$[6]) and are, therefore, well suited for longitudinal studies.

Modeling tissue structure from dMRI measurements – The diffusion of water molecules probes the microscopic structure of brain tissue. Membranes create obstacles to the diffusion process, leading to declines in the rate of diffusion, and creating separate compartments of intra- and extra-cellular water that can be distinguished with dMRI. The multi-b-value, dMRI sequence of the proposed study will make it possible to estimate the proportion of water that is restricted within cell bodies and axons (neurites[83]), adding an additional layer of specificity to our biological interpretation. Mean diffusivity (**MD**) and fractional anisotropy (**FA**) will be computed to characterize Gaussian diffusion and intracellular water volume fraction (based on the NODDI model[83]) will be calculated to characterize changes in the amount of water that is contained within cellular structures. Intracellular volume fraction will change if axons grow larger in diameter in during learning.

T1 and MTV mapping – Our goal is to go beyond simply localizing learning-induced changes and work towards understanding the neurobiology of these changes. Diffusion measurements will be combined with quantitative MRI measurements of T1 relaxation rate and macromolecular tissue volume (**MTV,** introduced by Mezer, Yeatman and colleagues[5]). T1 measures of the longitudinal relaxation rate of hydrogen protons in a magnetic field and post mortem work has demonstrated that, in white matter, T1 is primarily driven by variation in myelin content[88]. The PI and collaborators have confirmed that qMRI measures of MTV are an accurate index of the true volume of tissue macromolecules within each voxel[5]. Thus, MTV can be used to directly monitor the creation of new tissue over the course of an intervention and, in combination with T1, can be used to make inferences as to the cellular composition (e.g., myelination) of tissue at millimeter resolution. Quantitative T1 and MTV mapping will be performed by collecting T1-weighted images (SPGRs) at multiple flip angles ($1mm^3$ resolution) and transmit coil inhomogeneity (B1) will be corrected based on unbiased T1 estimates from a low resolution spin-echo inversion recovery sequence[4,5]. T1 and MTV values will be mapped to fiber tracts using AFQ[4].

Combining qMRI and dMRI for a more accurate model of human brain tissue: The qMRI and dMRI measurements in **Aim 1** are independent acquisitions and we have demonstrated in previous work that these measures are sensitive to independent biological processes in the white matter[4]. These measures provide a means to differentiate changes in intracellular water fraction from changes in myelin. For example, although the branching of astrocyte processes has a large impact on diffusion[89], astrocytes are believed to have a smaller impact on T1 relaxation than do changes in myelin[4]. Given that we have confirmed that measures of T1 and diffusion have similar SNR[4], and similar sensitivity to developmental change, the presence of common versus differential effects in the two modalities would point to distinct biological underpinnings. An interpretation of differential effects in the two measures will have to acknowledge and account for differential biases (e.g., subject motion introduces bias into diffusion, but not qMRI measures [4]), but **such a finding would be novel in the literature**. In summary, combining multiple MRI techniques will lead to a more detailed description of underlying biological sources of white matter plasticity during learning and build new bridges with animal models.

Research questions and preliminary data

Aim 1.A: Does intervention cause short-term, transient changes in the white matter, or long-term remodeling of the brain circuit? Based on previous work, we find highly significant changes ($p<0.0001$) in diffusion properties for an extensive network of white matter tracts (**Figure 2 A,B,C**), and stability in a control group[11]. Surprisingly, even tracts beyond the core reading circuitry show dramatic learning-induced plasticity. Hence, there is no question that reading intervention causes rapid and widespread changes in white matter tissue properties. However, there are many potential interpretations for these changes that point to different underlying mechanisms. For example, we might interpret changes in mean diffusivity as reflecting increased "white matter integrity", and infer that white matter deficits in children with dyslexia are remediated by the intervention program[80]. Alternatively, we might posit that the changes in mean diffusivity reflect transient changes in oligodendrocyte precursor cells (and other glial cells) resulting from the new experience of the intervention, but not reflecting permanent remodeling of the brain circuit. Under this second interpretation, white matter plasticity is still critical to the learning process (in line with animal work[1,2,90]), but reflects a more nuanced mechanism than typically considered in human studies. To dissociate these alternative accounts of experience-dependent plasticity and dyslexia-remediation, we will analyze follow-up measurements at 6 and 12 months post-intervention to determine if: (a) diffusion properties return to baseline, (b) changes remain stable or continue to grow over time, or (c) diffusion properties remain stable or continue to grow **only in tracts that are critical for skilled reading** (e.g., left arcuate fasciculus and ILF), but return to baseline in other tracts (**Figure 2D**).

These data will offer a new understanding of white matter plasticity and learning by either: (1) confirming the stability of rapid, learning-induced changes, or (2) dissociating short-term and long-term mechanisms of change. **No study has examined the relationship between short-term and long-term mechanisms of plasticity**, meaning that findings will be novel, irrespective of how the results

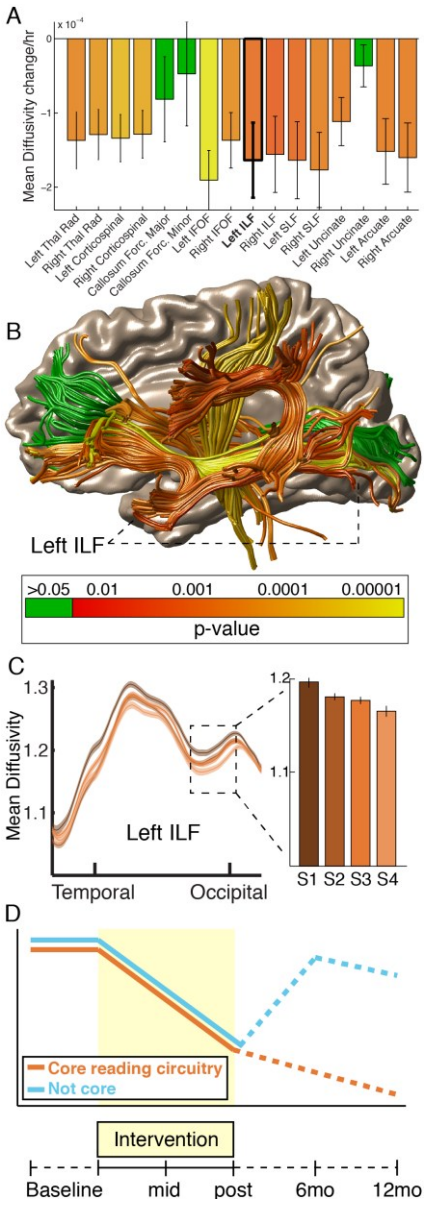

Figure 2: Diffusion measurements detect highly significant, intervention-driven white matter plasticity. LME models were used to model changes in mean diffusivity as a function of hours in the intervention (change/hr). **(A,B)** Significant linear growth was detected in a network of tracts including, but not limited to, the "core reading circuitry". Linear growth rates are shown for each tract, color-coded based on the p-value from the model. There is an extensive literature linking a core network of tracts (including the left arcuate fasciculus and left ILF) to reading skills. Surprisingly, our data demonstrate rapid, intervention-driven changes in these tracts, as well as many other tracts that are not conventionally associated with reading. Based on a simulation conducted with these data, the proposed sample size has excellent power to reliably measure white matter plasticity: assuming a 10% attrition rate, which is larger than the attrition in the pilot study, power = 0.98 for the left arcuate and 0.95 for the left ILF at $\alpha = 0.05$. **(C)** Mean diffusivity sampled along the trajectory of the ILF (from the occipital pole to the temporal pole) shows incremental growth between each measurement session, and along the full tract. Error bars represent +/- 1 SEM. Over the course of learning, many tracts show very similar growth trajectories to the one displayed for the ILF. **(D)** Why does such an extensive network show highly significant changes during the intervention? One hypothesis is that the structural changes seen during the intervention are transient and will return to baseline when the intervention is complete. Alternatively, we might hypothesize that only the tracts that are critical for maintaining the behavioral improvements, tracts that are considered to be the "core circuitry for reading" [34,35], will show sustained change or continued growth after the intervention is complete. This hypothesis is illustrated in **panel D**. During the baseline period we expect stability in the measurements: **19 control subjects not enrolled in the intervention showed <u>no change over the course of 4 measurement sessions.</u>** Hence, changes are not attributable to development or repeated scanning. During the intervention period we observe plasticity in the core reading circuitry (orange) and other, non-core tracts (blue). At 6-month and 12-month follow up sessions, our "transient change hypothesis" predicts that many of the measures will return to baseline and match the dyslexic control group. Critical tracts that support the behavior are predicted to show continued growth, and this growth will also be reflected in changes in quantitative T1 and MTV values (see **Aim 1.B**).

turn out. Given the high power for detecting short term learning effects (power = 0.95-0.99 at α=0.05), **Aim 1.A** is a low-risk, high-reward endeavor.

Aim 1.B: Do different properties of the white matter display different time-courses of plasticity? We measured rapid changes in mean diffusivity that emerge within just a few weeks of intervention (**Figure 2C**). This is in line with studies reporting mean diffusivity changes after hours of learning[89], and animal models indicate that these rapid changes are likely to reflect the morphology of glial cells. However, we cannot determine from diffusion measurements alone if changes in the diffusion signal are driven by glia (as suggested in the aforementioned work[89]), or if there are also changes in myelination and axon properties, as we might expect from recent animal work demonstrating that myelin plasticity is a critical component of learning[1,2]. By collecting long-term follow up measurements, and combining new, cutting-edge quantitative MRI pulse sequences and biophysical models, we will be able to better understand the underlying biology and time-course of the rapid white matter plasticity. We hypothesize that our data will fall in line with the sequence of changes that occur in mouse models of learning[1,2]: initial proliferation of glial cells throughout an extensive network, followed by long term remodeling of axons and myelin within the specific pathways that are critical for the behavior. The initial proliferation of glial cells would have a dramatic impact on mean diffusivity, but limited impact on T1, or MTV values[4,89]. Subsequent changes in myelin would have a dramatic impact on T1 and MTV values, but limited impact on m diffusivity. Though no MRI measure is specific to a single tissue type, combining multiple measures makes it possible to reason about the most likely biological source.

Aim 1.C: Does white matter plasticity track behavioral improvements? Seminal work by Barres and colleagues demonstrated that electrical activity on an axon causes oligodendrocytes to respond with increased wraps of myelin[91,92]. This work inspired a surge of interest in understanding the role of white matter plasticity in learning[75,93–95]. But an outstanding question is the extent to which changes in the white matter reflect the process of remodeling the circuit to support changes in behavior, versus simply experience dependent changes that do not directly relate to learning of the new behavior. Recent work published in *Science*, demonstrated that activity-dependent changes in myelin are fundamental for motor learning in a rodent model[1,2]. When receptors on oligodendrocytes were blocked, meaning that myelination remains constant but no new myelin is created in response to electrical signals, mice did not learn a complex motor task. While this work clearly demonstrates that myelin is critical for learning, **it does not indicate that all changes in the white matter are important for learning**. For example, one interpretation of our pilot data is that the new experiences associated with the intervention program promote cellular changes that are not tightly linked to behavioral improvements.

Aim 1.C will investigate the link between white matter plasticity and behavior. Three statistical approaches will be employed. First, LME models will be used to test whether changes in diffusion properties, MTV and T1 track behavioral changes during the intervention. To limit the number of statistical comparisons, we will perform a principal component analysis (PCA) of the reading scores and use the first PC as a general index of reading abilities. For any significant effect, we will conduct post-hoc analyses to determine if it is specific to certain measures of reading skills. Pilot data show a statistically significant, but modest, relationship between the time-course of MD change within a subject and improvements in reading score (r = -0.30 Arcuate, p = 0.003; r = -0.32 ILF, p = 0.006). Second, correlation analysis will be used to analyze whether changes in white matter tissue properties between: (a) the baseline period and the long-term follow-up sessions and (b) the immediate post-intervention session and the long-term follow-up sessions, correlate with individual differences in the amount of behavioral change. We hypothesize that the rapid changes in diffusion properties observed during the intervention are experience-dependent but do not reflect learning per se, and will only show a modest relationship to individual differences in learning. Instead it will be the longer-term changes in T1 values (indicative of myelination) that predict crystallized skill acquisition. Third, we will perform **an exploratory analysis** using latent change score modeling[96–98] to investigate whether the data support time-lagged relationships between growth in the white matter and growth in reading skills.

Alternative strategies

Aim 1 is grounded in methods for which the PI's lab has substantial expertise, and previous work indicates large effects with excellent statistical power. Additionally, there is superb support at the scanning facility, with an "on call" physicist, and technical staff that are trained to work with children. However, some of the more cutting-edge modeling approaches are still being refined in the literature (including work by our collaborators[99,100]) and are, therefore, open to alternative interpretations. We will confirm that our inferences do not depend on specific model assumptions by ensuring findings hold up under different modeling frameworks[101–105]. Additionally, even though the proposed interpretations of the qMRI measures are grounded in an extensive literature [4,5,88,99,106–114], these interpretations are simplifications of complex biophysical phenomena. For a comprehensive understanding of

the link between measured changes in MRI signals, and underlying biological sources, we will leverage simulations[115–118]. Simulations are a powerful tool to explore alternative interpretations of *in vivo* measurements. Finally, we do not anticipate issues with subject attrition: Before enrolling in the intervention, potential subjects undergo extensive behavioral testing, MRI screening, and practice at an MRI mock scanner. Power is greater than 0.95 assuming a 10% attrition rate, but **no subjects dropped out from the pilot study**.

Aim 2: Bottom-up and top-down computations in the reading circuitry

Significance: When our eyes fixate upon a word, a cascade of neural processes is initiated, beginning in the visual system and progressing through a series of computations that translate the visual representation into sound and meaning. In skilled readers, this process occurs in a fraction of a second, and engages a well-characterized network of brain regions[119–123]. In people with dyslexia, many regions within this circuit show lower levels of activation. However, simply knowing that a region is "under-activated" does not elucidate the nature of the aberrant neural computations associated with poor reading. For example, in skilled readers a region of ventral occipitotemporal (**VOT**) cortex, termed the visual word form area (**VWFA**), selectively responds to printed text. This "text-selective" response is considered a hallmark of literacy, and under-activation of this region is the most consistently reported neural deficit in people with dyslexia[16,122–130] (see[16,124] for meta-analysis). Preliminary data on a small group of children scanned before and after the *Seeing* Stars intervention confirms that the intervention prompts a reliable increase in text-selectivity within the VWFA **(Figure 3)**. This learning-induced change was present in every subject. The text-selective response in the VWFA is typically interpreted as indicating neural tuning for orthography[12,131–133], although it is also widely accepted that top-down signals contribute to the VWFA response[15,17,134,135]. Hence, there are different mechanisms that might contribute to differences in dyslexia, and changes with learning. By employing a paradigm that tightly controls both the visual properties of the stimulus, and the cognitive processes that are engaged by the subject[17], it will be possible to re-interpret the mechanisms underlying low levels of activation. This aim leverages the first computational model of the VWFA[17]. By capitalizing on longitudinal measurements over the intervention, we will then determine how the learning process shapes neural computations in VOT.

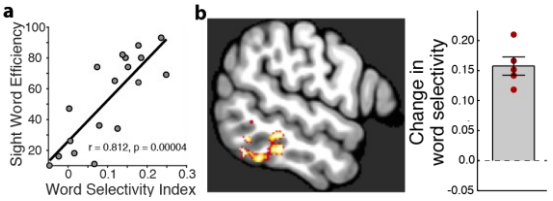

Figure 3: Learning-induced changes in the VWFA response to text. (a) Text-selective response in the VWFA, indexed as the difference in response to words compared to other images, is highly correlated with reading skill (N=18). (**b**), A voxelwise analysis shows a highly significant increase in text-selectivity localized to the typical location of the VWFA (p<0.001). The bar graph shows the change in selectivity for each individual subject (red dots, N=5).

Innovation: In **Aim 2** we employ an experimental paradigm that dissociates the automatic, bottom-up, stimulus-driven response evoked by seeing a word, from the top-down effect of the cognitive processes engaged by the subject (**Figure 4**). Our previous work has demonstrated that when subjects are engaged in a difficult task that diverts attention from the visual stimulus, the VWFA BOLD response can be accurately predicted with a three parameter visual encoding model (**Figure 4D**)[17]. This model posits specific neural computations governing the VWFA response. However, these "bottom-up" computations fail to predict the VWFA BOLD response when subjects engage in a task that requires them to attend to the visual stimulus. For example, when subjects engage in a one-back task (blue line, **Figure 4C**), the VWFA response is

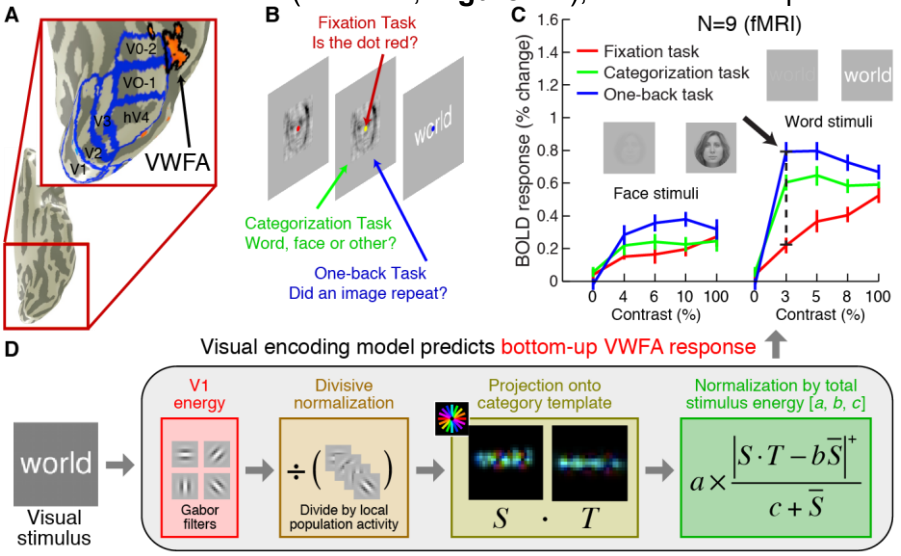

Figure 4: Bottom-up and top-down computations in the VWFA. (**A**) The VWFA is localized in ventral occipitotemporal cortex. (**B**) 9 Subjects were shown 22 categories of images, and engaged in three different tasks while viewing the images (Fixation, Categorization, and One-back tasks). (**C**) Average BOLD responses are shown for the VWFA, demonstrating that the VWFA is sensitive to visual properties of the image, and cognitive demands of the task. The VWFA selectively responds to text in the absences of attention (Fixation-task), and the response is amplified in the One-back task. **(D)** A three-parameter model predicts the bottom-up VWFA response based on a series of computations that neurons perform on the visual image. **Note**: our previous work separated the VWFA response into bottom-up vs. top-down and here we specifically consider the nature of the top-down signals.

amplified by 400% above the model prediction (black arrow, **Figure 4C**). Hence, there are separable components of the VWFA response: (1) bottom-up neural tuning properties and (2) top-down, task-dependent modulation. We refer to this task-dependent modulation of visual cortex as "top-down" since it reflects the cognitive demands of the task, but there are other interpretations that are consistent with the experimental design such as automatic versus goal-directed processing of the stimulus. Therefore, knowing that the VWFA response level differs between subjects, or changes with learning, could reflect a variety of mechanisms. Our innovative experimental approach will (a) provide new insights into the mechanisms underlying neural deficits in children with dyslexia[16,125,127] and (b) dissociate intervention driven changes in cognitive processing, from intervention driven changes in the automatic processing of text.

Relation to previous work: There is an extensive literature demonstrating that phonological processing regions in temporo-parietal cortex (TPC) show improved responses after intervention[136–142]. Some of these studies also report changes in VOT cortex[138] but, since the myriad of fMRI intervention studies employ cognitive tasks that require reading and/or phonological processing, it is unclear whether the improved VOT response reflects: (a) top-down signals from phonological processing regions, (b) improved automatic response to words, or (c) a combination of multiple effects. Our study fills an important gap by: (a) targeting changes in specific computations within VOT cortex, (b) employing a unique array of tasks and stimuli designed based on a computational model of the VWFA. These data will also allow us to test the hypothesis that top-down signals from TPC during phonological encoding play an important role in tuning the bottom-up VOT cortex response to words[126,143]. Moreover, of the 22 fMRI intervention studies reviewed in[142], most involve ~1hr/day of intervention making us uniquely poised to resolve the importance of intervention dosage for changing the VOT response.

Functional MRI measurements of bottom-up and top-down computations in the reading circuit

Data processing – A fundamental challenge in the longitudinal analysis of fMRI data is poor alignment due to EPI distortions. By collecting field maps at the beginning and end of each session, we will correct for field inhomogeneity, allowing for alignment of longitudinal data within ~1mm[144,145]. Distortion corrected fMRI data will be aligned to each subject's MPRAGE scan and resampled to the cortical surface[146]. A general linear model (GLM) will be fit to the data with GLMdenoise[147] leading to improved SNR.

Experiment 1 (Localizer scan) – All analyses will be conducted within each individual's native space, and without the use of smoothing kernels, allowing us to make precise inferences about neural response properties that are not confounded by anatomical differences among subjects. The use of an individual subjects approach will also allow us to identify mechanisms of change that would be obscured by a conventional group-average, voxelwise analysis (e.g., changes in the size of the VWFA relative to the immediately adjacent fusiform face area (FFA) vs. changes in VWFA selectivity for words).

Word, face and object selective regions[14,148–151] will be localized on each individual's cortical surface based on an optimized, block design, localizer[152]. Based on these data we can determine whether the learning process affects the size, or location, of category selective regions, and localize regions of interest (ROIs) that will be analyzed in the subsequent experiments. Having distortion-corrected localizer data collected over time will allow for three analysis strategies: (1) If there are not systematic changes in the boundaries of each region, then ROIs will be defined by concatenating the data across the sessions to maximize SNR; (2) If the VWFA does not exist in session 1 data, and emerges by the final scan as reading skills improve, then ROIs will be identified based on the final scan data so that we can determine how response properties in this patch of cortex change over the course of learning; (3) If the size, location and boundaries of category selective regions change during learning, then we will model these changes using deformation fields to better understand learning-induced cortical reorganization. Finally, the use of template ROIs remains a viable alternative strategy.

Experiment 2: Dissociating bottom-up encoding from top-down modulation – Data will be analyzed within regions of interest (see above) and, as an alternative strategy, group analysis will be used to examine effects that might occur outside of our ***a priori*** ROIs (see alternative strategies). Five image categories (words, faces, abstract objects, symbol strings and foreign characters will be presented at fixation (4 degrees visual angle, 800ms image presentation, 200ms ISI). Each image category will be presented at 2 contrast levels (4% and 100%) for a total of 10 stimulus conditions. Stimuli will be organized into 4 second blocks, and the order will be randomized with blanks to maximize our SNR for estimating the response to each stimulus type[17,152,153]. In recent work, we have developed a paradigm to separate bottom-up (stimulus-evoked) and top-down (task-dependent) signals in the VWFA[17]. On alternating runs with ***the exact same stimuli***, children will engage in a *Fixation* task or *One-back task*. The Fixation task is designed to isolate the automatic, bottom-up response: the subject makes judgments about the color of rapidly changing fixation dot, ignoring the presented images, with task difficulty

controlled by a staircase procedure to maintain an 82% level of performance. The O*ne-back task* is to press a button whenever any image appears twice in succession, and therefore directs attention to the images and, for words, engages the broader network of language regions involved in reading. We will employ the *One-back* task in this proposal, rather than a reading-specific task, because all children are able to perform the One-back task regardless of reading skill. The use of a reading-specific task would add the confound of differences in task performance: for example, in a severely dyslexic child, it would be unclear if lack of top-down modulation reflects a general deficit in top-down signaling, or the inability to perform the reading-related task. There is a large text-selective response in both tasks (**Figure 4**).

Experiment 3: Rhyming task – In order to (a) isolate top-down signals due to phonological processing and (b) examine functional connectivity between the VWFA and TPC, we will employ the rhyming task of Hoeft and colleagues[38,125]. This task involves rhyme judgments on visually presented words. By comparing the VWFA response in our novel visual encoding task to the VWFA response on this well-studied phonological processing task over the course of intervention we can test the hypothesis that top-down phonological signals are critical for VWFA tuning[126]. Moreover, we can relate functional connectivity between the VWFA and TPC to changes in white matter connectivity.

Research questions

Aim 2.A: To what extent do deficits in ventral regions reflect automatic, stimulus-driven, neural tuning properties versus top-down signals? Multiple competing mechanisms have been proposed to explain text-selective responses in VOT: (1) the text-selective response reflects tuning properties of neurons in visual cortex that are involved in automatic word recognition[131,154]; (2) the text-selective response reflects top-down signals from language regions[15,155]. These alternate explanations of computations in the VWFA provide two very different interpretations for under activation of this region in people with dyslexia. Under activation might result from (a) poor tuning of neurons for text or (b) deficiencies in top-down signals originating from language regions. To achieve a more complete understanding of neural deficits in dyslexia, it is essential to tightly control both the cognitive task engaged by the subject, as well as properties of the visual stimulus to disambiguate visual encoding from task-related activation. **No previous study has examined VWFA response properties using a fixation task that minimizes top-down signals**. Though this group comparison is not the primary motivation for the proposed experiments (see 2.B and 2.C), the data will provide a more nuanced and mechanistic understanding of previously reported VWFA deficits in dyslexia.

Aim 2.B: To what extent does intervention change the bottom-up, or top-down response in the VWFA? There is general agreement that a successful intervention improves the response to words in VOT cortex[138,156]. In an intervention study by Shaywitz and colleagues, intervention effects were initially observed in TPC, and changes in VOT were observed in long-term follow-up measurements[157]. This seminal study lead to the hypothesis that the process of remediating reading disabilities first improves phonological processing circuits that subsequently tune VOT for automatic word recognition. This hypothesis was formalized in a model by Pugh and colleagues, which posits a central importance of TPC in the development of VOT[126,143,158]. In line with this model, we hypothesize intervention-driven changes will initially manifest as top-down, task-dependent signals from phonological processing regions in TPC and no change in the bottom-up VOT response. As reading fluency improves towards the end of the intervention, and over the subsequent year, changes will crystalize as an improvement in the automatic, bottom-up response to words. In other words, we hypothesize that both **bottom-up and top-down signals play crucial roles at different points in the learning process**.

Pilot data in literate adults demonstrates that the bottom-up VWFA response is highly selective for words over other visual stimuli even when subjects are performing the Fixation task and ignoring the words (**Figure 4 red line**). Pilot data in a group of children performing a visual detection task on words, faces and objects confirms a strong correlation between reading skill and VWFA selectivity for words, even in the absence of tasks requiring reading or phonological processing (**Figure 3a**). In a small sample of 5 dyslexic subjects who initially did not show a text-selective response, the *Seeing Stars* intervention induced a reliable increase in the text-selective response **for every subject (Figure 3b)**; the proposed experiments and modeling will allow us to determine the extent to which this change reflects bottom-up or top-down signals as well as resolving the time-course of learning effects. LME models will be used to analyze longitudinal changes in the fMRI data, and to link changes in neural responses to changes in behavior. The pilot data confirms that we can expect an increase in the response to words over the course of the learning period (mean % BOLD increase = 0.15, SD = 0.034, bias corrected SD = 0.036, power > 0.95 at α=0.05), and the specific experiments will provide a nuanced understanding of the neural computations that produce this change in BOLD amplitude.

Aim 2.C: **Relating changes in structure to changes in function.** ***Do changes in white matter connectivity increase the efficiency of top-down signals and improve functional connectivity between visual and phonological processing regions?*** The tight control provided by a within-subjects experimental paradigm provides a unique opportunity to model the relationship between white matter tissue properties and cortical response properties in the reading circuitry. For example, we might posit that increases in top-down, task-dependent modulation of the VWFA arises from increased efficiency of white matter connections. Indeed, we have demonstrated a direct connection from the VWFA to TPC, and inferior frontal cortex (IFC)[159], meaning that top-down signals might emanate from these classic language circuits. In our previous work, we introduced a method to model how specific brain regions modulate the VWFA response[17]. Here we will capitalize on this approach to model the strength of top down signals from (a) TPC, (b) posterior parietal cortex and (c) IFC to the VWFA. We will then examine whether a subject's intervention-driven change in 3 white matter tracts predict changes in top down signal strength: (a) the posterior arcuate fasciculus, (b) the vertical occipital fasciculus and (c) the arcuate fasciculus. Modeling the relationship between the white matter and cortical computation is of great theoretical importance and we consider sub-aim 2.C as a high-risk high-reward endeavor. This level of risk for a sub-aim is appropriate given the exceptional statistical power for each of our other aims and sub-aims.

Alternative strategies

Group analysis: Our approach benefits from the increased precision, specificity, and power afforded by an ROI analysis[160,161]. But voxel-wise group analysis remains a perfectly viable alternative strategy. A study-specific template will be created with the ANTS toolbox[162], de-noised data will be normalized to the template and voxelwise statistics will be computed in line with our aims. This alternative approach may lead to discoveries in other components of the reading circuitry beyond our specific hypotheses about VOT cortex.

Aim 3 – Neural biomarkers of learning outcomes

Significance: Given the commitment that is required for a successful intervention, ideally, we would have accurate methods to predict the amount of improvement a child a child is likely to show. Predicting a child's response-to-intervention (RTI) has been a major focus of the behavioral intervention literature. Numerous predictor variables have been examined with PA and RN emerging as the most consistent predictors[31,32]. However, despite progress in understanding the child characteristics that predict intervention success or lack thereof, prediction accuracy is far from perfect. With substantial room for improvement, many labs have turned to neuroimaging as a potential means to improve prediction. For example, in our pilot data, dMRI measures of the arcuate predict RTI better than PA or RN (**Figure 4B**). In the long run, an understanding of the mechanisms that predict individual differences in learning will pave the way for innovative, personalized intervention programs that specifically target these mechanisms.

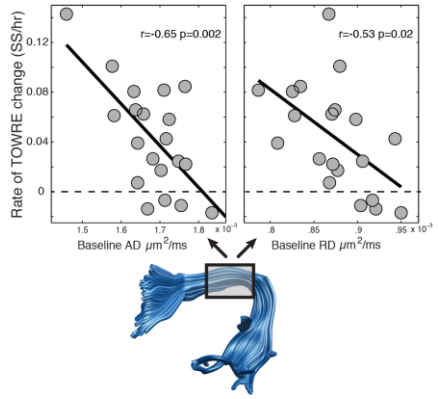

Figure 4: Arcuate diffusivity predicts reading improvement. Examining the region of the arcuate identified by Wang and colleagues [36], we find that pre-intervention diffusion properties predict each child's rate of improvement on the TOWRE. Children with lower diffusivity showed greater RTI. Using a step-wise, multivariate regression model to compare arcuate diffusivity, PA, and RN as predictors of RTI, only arcuate diffusivity was retained as a significant predictor. In other words, this biomarker subsumes the predictive value of the most widely used behavioral predictors.

In the neuroimaging literature, convergent findings, across multiple labs, have identified tissue properties of the arcuate fasciculus and superior longitudinal fasciculus (SLF) as potential biomarkers for dyslexia by demonstrating that tissue properties: (a) correlate with pre-reading skills such as phonological awareness[23], (b) differentiate preschool children and infants at risk for dyslexia from not-at-risk individuals[33,41,163], (c) develop more rapidly in good versus poor reading children[36,37], and (d) predict future reading gains in elementary school children with poor reading skills[36,38]. What is particularly noteworthy is that arcuate/SLF tissue properties predict future reading development better than any behavioral measure[36,38]. **However, none of these studies have assessed whether tissue properties are predictive of behavioral gains in an intervention setting.** Assessing the predictive power of this biomarker in a controlled intervention setting is both important for ruling out the possibility that predictions were influenced by environmental differences among children, and for the practical purpose of developing tools to predict a child's likelihood of success before the substantial commitment involved in this popular intervention program. The successful development of a predictive model can then be extended to other intervention programs through our sharing of reproducible analysis tools[6,61].

Innovation: The concept of "neuroprognosis", or using brain measurements to make useful and prognostic predictions, has generated a lot excitement, and a flurry of publications in the field of educational neuroscience (for review see[39]). However, one of the limiting factors is that there is always substantial variability in the quality, and type, of education that children have access to. This sets an upper-bound on the potential accuracy of any prediction: even if the critical brain property is identified for a particular academic skill, a substantial portion of the variance in outcome measurements will be due to differences in a child's environment. Indeed, roughly 50% of the variance in reading skills is the product of environmental factors[164]. The current proposal benefits from an intervention design in which environmental factors are controlled across subjects, allowing us to test whether there are neuroanatomical factors, unique to the child, that predispose them to succeed in the intervention program. Our approach to prediction will begin by examining biomarkers that are motivated by the literature and pilot data. After establishing the predictive value of these theory-driven biomarkers, we will use **cutting-edge machine learning algorithms to investigate whether a multivariate, data-driven approach can improve prediction accuracy**. Our approach to neuroprognosis benefits from unique tools we have developed to quantify tissue properties in the individual, and leverages collaboration with machine learning experts to explore the high-dimensional space afforded by these data.

Research questions and preliminary data

Aim 3.A: Tissue properties of the arcuate as a biomarker for response to intervention: The aforementioned studies demonstrating that arcuate tissue properties are predictive of longitudinal growth employed the AFQ software package[33,36], which is developed by the PI's lab. Hence, the proposed research team is optimally suited to follow up on these findings, develop methods for automated, brain-based predictions of reading outcomes, and disseminate these tools to the field. Wang and colleagues reported that a specific region of the arcuate is predictive of reading outcomes when children are followed longitudinally in a conventional school setting[36]. Examining this same region of the arcuate, we find that **each subject's pre-intervention diffusivity values predict their improvement during the intervention above and beyond baseline behavioral measures (Figure 4)**. These data are a promising example demonstrating that biomarkers identified in one laboratory, can be replicated and extended by another lab, and lend support to the perspective that anatomical properties of the arcuate play a role in determining how easily children learn reading skills. **Such an extension is made possible through open sharing of code[6,165], and is facilitated by the open science approach of the PI's lab.** Pilot data clearly establishes the feasibility of predicting learning outcomes based on properties of the arcuate (power = 0.93 for RD; 0.99 for AD at $\alpha = 0.05$). The critical question is whether this biomarker predicts long-term follow-up measurements (above and beyond variance predicted by baseline behavioral data), since long-term achievement is the ultimate goal of an intervention. This type of long-range, post intervention prediction has not been examined before in the neuroimaging literature. Cross validation will be used to test out of sample generalization and assess whether biologically specific qMRI parameters improve prediction accuracy.

Aim 3.B: Does the severity of neural processing deficits predict resistance to remediation? Even though the *Seeing Stars* intervention produces highly significant changes in reading skills, there is variability among children, with some children increasing by more than 2 standard deviations, and others showing limited growth. Is variability in response to intervention linked to the severity of underlying neural deficits? We will focus on three regions implicated in dyslexia: the temporoparietal region involved in phonological processing[166,167], the ventral occipitotemporal VWFA, and the inferior frontal region that has been proposed as a compensatory mechanism in dyslexia[38,125–127]. The pre-intervention response profile in each of these regions will be used to index the magnitude of neural deficits in each subject and will be examined as a predictor of RTI. **Aim 3.B** is considered an exploratory analysis due to lack of supporting pilot data. However, the data for **Aim 3.B** will already be collected for **Aim 2**, and it is of theoretical importance to know whether more severe neural deficits, in specific components of the reading circuitry, are resistant to remediation.

Alternative strategies

Even though our approach to prediction is grounded in an extensive literature, we have to acknowledge the possibility that the prediction accuracy will not generalize beyond the specifics of previous samples. **We contend that a rigorous test of biomarkers proposed in the literature is important and we will, therefore, preregister our study such that the results are published irrespective of prediction accuracy[168].**

Machine learning: Aim 3.A confronts specific hypotheses that are grounded in previous work but, as an alternative approach, we will capitalize on new machine learning algorithms to explore combinations of neuroanatomical features that might be used to predict RTI. This alternative approach will be supported through collaboration with Data Scientist Ariel Rokem at the UW eScience Institute. Our approach to machine learning

will start with feature extraction that is informed by the scientific literature and regularized regression: data from **3.A** and **3.B** will be organized into a predictor matrix and (1) linear combinations of features that predict reading improvements (e.g., using Elastic Net[169,170]), as well as non-linear combinations (e.g., using Random Forest Regression[171]) will be learned from the data. We can incrementally increase the size of the predictor set by adding (a) data from voxels rather than average ROI responses (akin to MVPA[38]), and (b) data from all nodes in the tracts rather than averaged for regions of tracts. PCA will be used for dimensionality reduction[165,172]. Cross-validation will be used to control for over-fitting and ensure out of sample generalization at each stage.

LITERATURE CITED

1. McKenzie, I. A. *et al.* Motor skill learning requires active central myelination. *Science (80-.).* **346,** 318–322 (2014).
2. Gibson, E. *et al.* Neuronal Activity Promotes Oligodendrogenesis and Adaptive Myelination in the Mammalian Brain. *Science (80-.).* **344,** 480–481 (2014).
3. Fields, R. D. A new mechanism of nervous system plasticity: activity-dependent myelination. *Nat. Rev. Neurosci.* **16,** 756–67 (2015).
4. Yeatman, J. D., Wandell, B. A. & Mezer, A. A. Lifespan maturation and degeneration of human brain white matter. *Nat. Commun.* **5,** 4932 (2014).
5. Mezer, A. *et al.* Quantifying the local tissue volume and composition in individual brains with magnetic resonance imaging. *Nat. Med.* **19,** 1667–72 (2013).
6. Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A. & Feldman, H. M. M. Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS One* **7,** e49790 (2012).
7. Pestilli, F., Yeatman, J. D., Rokem, A., Kay, K. N. & Wandell, B. A. Evaluation and statistical inference for human connectomes. *Nat. Methods* **11,** 1058–63 (2014).
8. Rokem, A. *et al.* Evaluating the accuracy of diffusion MRI models in white matter. *PLoS One* **10,** e0123272. (2015).
9. Donnelly, P. M., Huber, E. & Yeatman, J. D. Intensive summer intervention drives linear growth of reading skill in children with dyslexia. *PsyArXiv psyarxiv.com/jfpe3* (2017).
10. Joo, S. J., Donnelly, P. M. & Yeatman, J. D. The causal relationship between dyslexia and motion perception reconsidered. *Sci. Rep.* **7,** (2017).
11. Huber, E., Donnelly, P. M. P. M., Rokem, A. & Yeatman, J. D. Rapid and widespread white matter plasticity during an intensive reading intervention. *Nat. Commun.* **9,** 2260 (2018).
12. Cohen, L. *et al.* The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123,** 291–307 (2000).
13. Dehaene, S., Le Clec, H. G., Poline, J. B., Le Bihan, D. & Cohen, L. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport* **13,** 321–325 (2002).
14. Yeatman, J. D., Rauschecker, A. M. A. M. & Wandell, B. A. Anatomy of the visual word form area: Adjacent cortical circuits and long-range white matter connections. *Brain Lang.* **125,** 146–155 (2013).
15. Price, C. J. & Devlin, J. T. The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn Sci* **15,** 246–253 (2011).
16. Richlan, F., Kronbichler, M. & Wimmer, H. Meta-analyzing brain dysfunctions in dyslexic children and adults. *Neuroimage* **56,** 1735–1742 (2011).
17. Kay, K. N. N. & Yeatman, J. D. Bottom-up and top-down computations in word- and face-selective cortex. *Elife* **6,** (2017).
18. Snow, C. E., Burns, M. S. & Griffin, P. *Preventing Reading Difficulties in Young Children*. (National Academies Press, 1998). doi:10.17226/6023
19. Morris, R. R. D. *et al.* Multiple-Component Remediation for Developmental Reading Disabilities: IQ, Socioeconomic Status, and Race as Factors in Remedial Outcome. *J. Learn. Disabil.* **45,** 99–127 (2012).
20. Shaywitz, S. E., Morris, R. & Shaywitz, B. A. The education of dyslexic children from childhood to young adulthood. *Annu. Rev. Psychol.* **59,** 451–475 (2008).
21. Berninger, V. W. Understanding the 'Lexia' in dyslexia: A multidisciplinary team approach to learning disabilities. *Ann. Dyslexia* **51,** 21–48 (2001).
22. Torgesen, J. K. Catch them before they fall: Identification and assessment to prevent reading failure in young children. *Am. Educ.* **22,** 32–39 (1998).
23. Torgesen, J. K. Preventing Early Reading Failure. *Am. Educ.* **28,** (2004).

24. Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J. & Lindamood, P. Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Ann. Dyslexia* **60,** 40–56 (2010).

25. Torgesen, J. K. *et al.* Preventing Reading Failure in Young Children wtih Phonological Processing Disabilities: Group and Individual Responses to Instruction. *J. Educ. Pyschology* **91,** 579–593 (1999).

26. Wolf, M. & Katzir-Cohen, T. Reading Fluency and Its Intervention. *Sci. Stud. Read.* **5,** 211–239 (2001).

27. Wolf, M., Miller, L. & Donnelly, K. Retrieval, Automaticity, Vocabulary Elaboration, Orthography (RAVE-O) A Comprehensive, Fluency-Based Reading Intervention Program. *J. Learn. Disabil.* (2000).

28. Lovett, M. W. *et al.* Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. *J. Educ. Psychol.* **92,** 263–283 (2000).

29. Lovett, M., Borden, S. & DeLuca, T. Treating the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically-and strategy-based reading training programs. *Developmental* (1994).

30. Torgesen, J. Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learn. Disabil. Res. Pract.* (2000).

31. Tran, L., Sanchez, T., Arellano, B. & Swanson, H. L. A meta-analysis of the RTI literature for children at risk for reading disabilities. *J. Learn. Disabil.* **44,** 283–295 (2011).

32. Al Otaiba, S. & Fuchs, D. Characteristics of Children Who Are Unresponsive to Early Literacy Intervention: A Review of the Literature. *Remedial Spec. Educ.* **23,** 300–316 (2002).

33. Langer, N. *et al.* White Matter Alterations in Infants at Risk for Developmental Dyslexia. *Cereb. Cortex* bhv281 (2015). doi:10.1093/cercor/bhv281

34. Ozernov-Palchik, O. & Gaab, N. Tackling the 'dyslexia paradox': Reading brain and behavior for early markers of developmental dyslexiax. *Wiley Interdiscip. Rev. Cogn. Sci.* **7,** 156–176 (2016).

35. Wandell, B. A. & Yeatman, J. D. Biological development of reading circuits. *Curr. Opin. Neurobiol.* **23,** 261–268 (2013).

36. Wang, Y. *et al.* Development of Tract-Specific White Matter Pathways During Early Reading Development in At-Risk Children and Typical Controls. *Cereb. Cortex* bhw095 (2016). doi:10.1093/cercor/bhw095

37. Yeatman, J. D., Dougherty, R. F., Ben-Shachar, M. & Wandell, B. A. Development of white matter and reading skills. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E3045-53 (2012).

38. Hoeft, F. *et al.* Neural systems predicting long-term outcome in dyslexia. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 361–366 (2011).

39. Gabrieli, J. D. E., Ghosh, S. S. & Whitfield-Gabrieli, S. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* **85,** 11–26 (2015).

40. Gabrieli, J. D. E. Dyslexia: a new synergy between education and cognitive neuroscience. *Science* **325,** 280–3 (2009).

41. Saygin, Z. M. *et al.* Tracking the roots of reading ability: white matter volume and integrity correlate with phonological awareness in prereading and early-reading kindergarten children. *J. Neurosci.* **33,** 13251–8 (2013).

42. Gullick, M. M. & Booth, J. R. The direct segment of the arcuate fasciculus is predictive of longitudinal reading change. *Dev. Cogn. Neurosci.* **13,** 68–74 (2015).

43. McNorgan, C., Alvarez, A., Bhullar, A., Gayda, J. & Booth, J. R. Prediction of reading skill several years later depends on age and brain region: implications for developmental models of reading. *J. Neurosci.* **31,** 9641–9648

44. Otaiba, S. Al, Wagner, R. & Miller, B. "Waiting to Fail" Redux Understanding Inadequate Response to Intervention. *Learn. Disabil. Q.* **37,** 129–133 (2014).

45. Klingberg, T. *et al.* Microstructure of temporo-parietal white matter as a basis for reading ability: evidence

from diffusion tensor magnetic resonance imaging. *Neuron* **25,** 493–500 (2000).

46. Dougherty, R. F. *et al.* Temporal-callosal pathway diffusivity predicts phonological skills in children. *Proc Natl Acad Sci U S A* **104,** 8556–8561 (2007).

47. Vanderauwera, J., Wouters, J. & Vandermosten, M. Neural connectivity as a pre-reading marker of dyslexia.

48. Vandermosten, M. *et al.* A tractography study in dyslexia: neuroanatomic correlates of orthographic, phonological and speech processing. *Brain* **135,** 935–48 (2012).

49. Saygin, Z. M. *et al.* Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* **19,** 1250–5 (2016).

50. Frye, R. E. *et al.* Diffusion tensor quantification of the relations between microstructural and macrostructural indices of white matter and reading. *Hum. Brain Mapp.* doi:10.1002/hbm.21103

51. Myers, C. A. *et al.* White matter morphometric changes uniquely predict children's reading acquisition. *Psychol. Sci.* **25,** 1870–83 (2014).

52. Gullick, M. M. & Booth, J. R. Individual differences in crossmodal brain activity predict arcuate fasciculus connectivity in developing readers. *J. Cogn. Neurosci.* **26,** 1331–46 (2014).

53. Beaulieu, C. *et al.* Imaging brain connectivity in children with diverse reading ability. *Neuroimage* **25,** 1266–1271 (2005).

54. Niogi, S. N. & McCandliss, B. D. Left lateralized white matter microstructure accounts for individual differences in reading ability and disability. *Neuropsychologia* **44,** 2178–2188 (2006).

55. Yeatman, J. D. *et al.* Anatomical Properties of the Arcuate Fasciculus Predict Phonological and Reading Skills in Children. *J. Cogn. Neurosci.* **23,** 3304–17 (2011).

56. Deutsch, G. K. *et al.* Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. *Cortex* **41,** 354–363 (2005).

57. Beaulieu, C. & Christian, B. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed.* **15,** 435–455 (2002).

58. O'Donnell, L. J. & Pasternak, O. Does diffusion MRI tell us anything about the white matter? An overview of methods and pitfalls. *Schizophr. Res.* **161,** 133–141 (2015).

59. Assaf, Y. & Pasternak, O. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *J. Mol. Neurosci.* **34,** 51–61 (2008).

60. Jones, D. K., Knösche, T. R. & Turner, R. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *Neuroimage* **73,** 239–54 (2013).

61. Yeatman, J. D., Richie-Halford, A., Smith, J. K. J. K., Keshavan, A. & Rokem, A. A browser-based tool for visualization and analysis of diffusion MRI data. *Nat. Commun.* **9,** 940 (2018).

62. Kennedy, D. N., Haselgrove, C., Riehl, J., Preuss, N. & Buccigrossi, R. The NITRC image repository. *Neuroimage* **124,** 1069–1073 (2016).

63. Stikov, N. *et al.* Bound pool fractions complement diffusion measures to describe white matter micro and macrostructure. *Neuroimage* **54,** 1112–1121 (2011).

64. Bell, N. *Seeing Stars: Symbol Imagery for Phonological and Orthographic Processing in Reading and Spelling*. (Gander Publishing; 2nd edition, 2013).

65. Byiers, B. J., Reichle, J. & Symons, F. J. Single-subject experimental design for evidence-based practice. *Am. J. Speech-Language Pathol.* **21,** 397–414 (2012).

66. O'Brien, G. E., McCloy, D. R., Kubota, E. C. & Yeatman, J. D. Reading ability and phoneme categorization. *Sci. Rep.* (2018).

67. Al Otaiba, S., Lewis, S., Whalon, K., Dyrlund, A. & McKenzie, A. R. Home Literacy Environments of Young Children With Down Syndrome. *Remedial Spec. Educ.* **30,** 96–107 (2009).

68. Brites, C., Salgado-Azoni, C. A., Ferreira, T. L., Lima, R. F. & Ciasca, S. M. Development and applications

of the SWAN rating scale for assessment of attention deficit hyperactivity disorder: a literature review. *Brazilian J. Med. Biol. Res. = Rev. Bras. Pesqui. medicas e Biol.* **48,** 965–72 (2015).

69. Swanson, J. M. *et al.* Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. *Int. J. Educ. Psychol. Assess.* **10,** 51–70 (2012).

70. Romeo, R. R. *et al.* Socioeconomic Status and Reading Disability: Neuroanatomy and Plasticity in Response to Intervention. *Cereb. Cortex* **91,** 1–16 (2017).

71. Neath, A. A. & Cavanaugh, J. E. The Bayesian information criterion: background, derivation, and applications. *Wiley Interdiscip. Rev. Comput. Stat.* **4,** 199–203 (2012).

72. Schwarz, G. Estimating the Dimension of a Model. *The Annals of Statistics* **6,** 461–464

73. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models using lme4. (2014).

74. Lebel, C., Treit, S. & Beaulieu, C. Diffusion MRI of typical white matter development from childhood to adulthood. *NMR Biomed* 1–23 (2017). doi:10.1002/nbm.3778

75. Fields, R. D. White matter in learning, cognition and psychiatric disorders. *Trends Neurosci.* **31,** 361–70 (2008).

76. Freud, E. *et al.* Three-Dimensional Representations of Objects in Dorsal Cortex are Dissociable from Those in Ventral Cortex. *Cereb. Cortex* 1–13 (2015). doi:10.1093/cercor/bhv229

77. Thomason, M. E. & Thompson, P. M. Diffusion imaging, white matter, and psychopathology. *Annu. Rev. Clin. Psychol.* **7,** 63–85 (2011).

78. Ben-Shachar, M., Dougherty, R. F. & Wandell, B. A. White matter pathways in reading. *Curr Opin Neurobiol* **17,** 258–270 (2007).

79. Vandermosten, M., Boets, B., Wouters, J. & Ghesquière, P. A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. *Neurosci. Biobehav. Rev.* **36,** 1532–52 (2012).

80. Keller, T. a & Just, M. A. Altering cortical connectivity: remediation-induced changes in the white matter of poor readers. *Neuron* **64,** 624–31 (2009).

81. Andersson, J. L. R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *Neuroimage* **20,** 870–888 (2003).

82. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125,** 1063–1078 (2016).

83. Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61,** 1000–16 (2012).

84. Tournier, J.-D. *et al.* Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage* **42,** 617–25 (2008).

85. Tournier, J.-D., Calamante, F. & Connelly, A. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* **35,** 1459–72 (2007).

86. Garyfallidis, E. *et al.* Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* **8,** 8 (2014).

87. Maier-Hein, K. *et al.* Tractography-based connectomes are dominated by false-positive connections. *doi.org* 084137 (2016). doi:10.1101/084137

88. Stüber, C. *et al.* Myelin and iron concentration in the human brain: A quantitative study of {MRI} contrast. *Neuroimage* (2014). doi:10.1016/j.neuroimage.2014.02.026

89. Sagi, Y. *et al.* Learning in the Fast Lane: New Insights into Neuroplasticity. *Neuron* **73,** 1195–1203 (2012).

90. Purger, D., Gibson, E. M. & Monje, M. Myelin plasticity in the central nervous system. *Neuropharmacology* **110,** 563–573 (2016).

91. Barres, B. A. The mystery and magic of glia: a perspective on their roles in health and disease. *Neuron* **60,** 430–440 (2008).

92. Barres, B. A. & Raff, M. C. Proliferation of oligodendrocyte precursor cells depends on electrical activity in axons. *Nature* **361,** 258–260 (1993).

93. Zatorre, R. J., Fields, R. D. & Johansen-berg, H. Plasticity in gray and white : neuroimaging changes in brain structure during learning. *Nat. Publ. Gr.* **15,** 528–536 (2012).

94. Ishibashi, T. *et al.* Astrocytes promote myelination in response to electrical impulses. *Neuron* **49,** 823–32 (2006).

95. Fields, R. D. White matter matters. *Sci Am* **298,** 42–49 (2008).

96. Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B. & Resnick, S. M. Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Struct. Equ. Model. A Multidiscip. J.* **19,** 268–292 (2012).

97. Quinn, J. M., Wagner, R. K., Petscher, Y. & Lopez, D. Developmental Relations Between Vocabulary Knowledge and Reading Comprehension: A Latent Change Score Modeling Study. *Child Dev.* **86,** 159–175 (2015).

98. Petscher, Y., Quinn, J. M. & Wagner, R. K. Modeling the Co-Development of Correlated Processes With Longitudinal and Cross-Construct Effects. *Dev. Psychol.* **52,** 1690–1704 (2016).

99. Berman, S., West, K. L., Does, M. D., Yeatman, J. D. & Mezer, A. A. Evaluating g-ratio weighted changes in the corpus callosum as a function of age and sex. *Neuroimage* (2017). doi:10.1016/j.neuroimage.2017.06.076

100. Mezer, A., Rokem, A., Berman, S., Hastie, T. & Wandell, B. A. Evaluating quantitative proton-density-mapping methods. *Hum. Brain Mapp.* **37,** 3623–3635 (2016).

101. Jelescu, I. O. *et al.* One diffusion acquisition and different white matter models: How does microstructure change in human early development based on WMTI and NODDI? *Neuroimage* **107,** 242–256 (2015).

102. Jelescu, I. O. *et al.* In vivo quantification of demyelination and recovery using compartment-specific diffusion MRI metrics validated by electron microscopy. *Neuroimage* **132,** 104–114 (2016).

103. Jensen, J. H. & Helpern, J. A. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed.* **23,** 698–710 (2010).

104. Kelm, N. D. *et al.* Evaluation of diffusion kurtosis imaging in ex vivo hypomyelinated mouse brains. *Neuroimage* **124,** 612–626 (2016).

105. Fieremans, E., Jensen, J. H. & Helpern, J. A. White matter characterization with diffusional kurtosis imaging. *Neuroimage* **58,** 177–188 (2011).

106. Lambert, C. *et al.* Characterizing aging in the human brainstem using quantitative multimodal MRI analysis. *Front. Hum. Neurosci.* **7,** 462 (2013).

107. Yeatman, J. D. *et al.* The vertical occipital fasciculus: A century of controversy resolved by in vivo measurements. *Proc. Natl. Acad. Sci.* **111,** E5214–E5223 (2014).

108. De Santis, S., Drakesmith, M., Bells, S., Assaf, Y. & Jones, D. K. Why diffusion tensor MRI does well only some of the time: Variance and covariance of white matter tissue microstructure attributes in the living human brain. *Neuroimage* **89,** 35–44 (2014).

109. Weiskopf, N., Mohammadi, S., Lutti, A. & Callaghan, M. F. Advances in MRI-based computational neuroanatomy: from morphometry to in-vivo histology. *Curr. Opin. Neurol.* **28,** 313–22 (2015).

110. Callaghan, M. F. *et al.* Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. *Neurobiol. Aging* **35,** 1862–1872 (2014).

111. Lutti, A., Dick, F., Sereno, M. I. & Weiskopf, N. Using high-resolution quantitative mapping of {R1} as an index of cortical myelination. *Neuroimage* 1–13 (2013). doi:10.1016/j.neuroimage.2013.06.005

112. Dick, F. *et al.* In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *J. Neurosci.* **32,** 16095–16105 (2012).

113. Lutti, A., Hutton, C., Finsterbusch, J., Helms, G. & Weiskopf, N. Optimization and validation of methods

for mapping of the radiofrequency transmit field at 3T. *Magn. Reson. Med.* **64,** 229–38 (2010).

114. Sereno, M. I., Lutti, A., Weiskopf, N. & Dick, F. Mapping the human cortical surface by combining quantitative t1 with retinotopy. *Cereb. Cortex* **23,** 2261–8 (2013).

115. Nilsson, M., Van Westen, D., Ståhlberg, F., Sundgren, P. C. & Lätt, J. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. *Magn. Reson. Mater. Physics, Biol. Med.* **26,** 345–370 (2013).

116. Landman, B. A. *et al.* Complex geometric models of diffusion and relaxation in healthy and damaged white matter. *NMR Biomed.* **23,** n/a-n/a (2009).

117. Henriques, R. N., Correia, M. M., Nunes, R. G. & Ferreira, H. A. Exploring the 3D geometry of the diffusion kurtosis tensor-Impact on the development of robust tractography procedures and novel biomarkers. *Neuroimage* **111,** 85–99 (2015).

118. Hall, M. G. & Alexander, D. C. Convergence and Parameter Choice for Monte-Carlo Simulations of Diffusion MRI. *IEEE Trans. Med. Imaging* **28,** 1354–1364 (2009).

119. Schlaggar, B. L. & McCandliss, B. D. Development of neural systems for reading. *Annu. Rev. Neurosci.* **30,** 475–503 (2007).

120. Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A. & Eden, G. F. Development of neural mechanisms for reading. *Nat. Neurosci.* **6,** 767–73 (2003).

121. Jobard, G., Crivello, F. & Tzourio-Mazoyer, N. Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *Neuroimage* **20,** 693–712 (2003).

122. Wandell, B. A. The neurobiological basis of seeing words. *Ann. N. Y. Acad. Sci.* **1224,** 63–80 (2011).

123. Wandell, B. A., Rauschecker, A. M. A. M. M. & Yeatman, J. D. Learning to see words. *Annu. Rev. Psychol.* **63,** 31–53 (2012).

124. Maisog, J. M., Einbinder, E. R., Flowers, D. L., Turkeltaub, P. E. & Eden, G. F. A meta-analysis of functional neuroimaging studies of dyslexia. *Ann. N. Y. Acad. Sci.* **1145,** 237–259 (2008).

125. Hoeft, F. *et al.* Functional and morphometric brain dissociation between dyslexia and reading ability. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 4234–4239 (2007).

126. Pugh, K. R. *et al.* Neurobiological studies of reading and reading disability. *J. Commun. Disord.* **34,** 479–92 (2001).

127. Shaywitz, B. A. *et al.* Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biol. Psychiatry* **52,** 101–10 (2002).

128. Wimmer, H., Ludersdorfer, P., Richlan, F. & Kronbichler, M. Visual Experience Shapes Orthographic Representations in the Visual Word Form Area. *Psychol. Sci.* (2016). doi:10.1177/0956797616657319

129. Richlan, F. Developmental dyslexia: dysfunction of a left hemisphere reading network. *Front. Hum. Neurosci.* **6,** 1–5 (2012).

130. Olulade, O. A., Flowers, D. L., Napoliello, E. M. & Eden, G. F. Dyslexic children lack word selectivity gradients in occipito-temporal and inferior frontal cortex. *NeuroImage Clin.* **7,** 742–754 (2015).

131. Dehaene, S. & Cohen, L. The unique role of the visual word form area in reading. *Trends Cogn Sci* **15,** 254–262 (2011).

132. Dehaene, S. & Cohen, L. Cultural recycling of cortical maps. *Neuron* **56,** 384–98 (2007).

133. Dehaene, S., Cohen, L., Sigman, M. & Vinckier, F. The neural code for written words: a proposal. *Trends Cogn. Sci.* **9,** 335–41 (2005).

134. Bitan, T., Cheon, J., Lu, D., Burman, D. D. & Booth, J. R. Developmental increase in top-down and bottom-up processing in a phonological task: an effective connectivity, fMRI study. *J. Cogn. Neurosci.* **21,** 1135–1145 (2009).

135. Twomey, T., Kawabata Duncan, K. J., Price, C. J. & Devlin, J. T. Top-down modulation of ventral occipito-temporal responses during visual word recognition. *Neuroimage* **55,** 1242–1251 (2011).

136. Temple, E. *et al.* Neural deficits in children with dyslexia ameliorated by behavioral remediation: evidence from functional MRI. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 2860–2865 (2003).

137. Yamada, Y. *et al.* Emergence of the neural network for reading in five-year-old beginning readers of different levels of pre-literacy abilities: An fMRI study. *Neuroimage* **57,** 704–713 (2011).

138. Shaywitz, B. A. *et al.* Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. *Biol. Psychiatry* **55,** 685–691 (2004).

139. Simos, P. G. *et al.* Magnetic source imaging studies of dyslexia interventions. *Dev Neuropsychol* **30,** 591–611 (2006).

140. Aylward, E. H. *et al.* Instructional treatment associated with changes in brain activation in children with dyslexia. *Neurology* **61,** 212–9 (2003).

141. Gebauer, D. *et al.* Differences in brain function and changes with intervention in children with poor spelling and reading abilities. *PLoS One* **7,** e38201 (2012).

142. Barquero, L. A., Davis, N. & Cutting, L. E. Neuroimaging of Reading Intervention: A Systematic Review and Activation Likelihood Estimate Meta-Analysis. *PLoS One* **9,** e83668 (2014).

143. Pugh, K. R. *et al.* Mapping the word reading circuitry in skilled and disabled readers. in *The Neural Basis of Reading* 281–305 (2012).

144. Kay, K. N., Winawer, J., Mezer, A. & Wandell, B. A. Compressive spatial summation in human visual cortex. *J. Neurophysiol.* **110,** 481–94 (2013).

145. Kay, K. N., Winawer, J., Rokem, A., Mezer, A. & Wandell, B. A. A two-stage cascade model of BOLD responses in human visual cortex. *PLoS Comput. Biol.* **9,** e1003079 (2013).

146. Fischl, B. FreeSurfer. *Neuroimage* **62,** 774–81 (2012).

147. Kay, K. N., Rokem, A., Winawer, J., Dougherty, R. F. & Wandell, B. A. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Front. Neurosci.* **7,** 247 (2013).

148. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15,** 536–548 (2014).

149. Weiner, K. S. & Grill-Spector, K. Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychol. Res.* **77,** 74–97 (2013).

150. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17,** 4302–11 (1997).

151. Kanwisher, N. Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 11163–11170 (2010).

152. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. *J. Neurosci.* **35,** 12412–12424 (2015).

153. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Differential rate of temporal processing across category-selective regions in human high-level visual cortex. *J. Vis.* **14,** 187–187 (2014).

154. Cohen, L. *et al.* Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain* **125,** 1054–1069 (2002).

155. Price, C. J. & Devlin, J. T. The myth of the visual word form area. *Neuroimage* **19,** 473–481 (2003).

156. Shaywitz, S. & M.D. *Overcoming Dyslexia: A New and Complete Science-Based Program for Reading Problems at Any Level*. (Knopf Doubleday Publishing Group, 2008).

157. Shaywitz, B. A. *et al.* Development of left occipitotemporal systems for skilled reading in children after a phonologically- based intervention. *Biol. Psychiatry* **55,** 926–933 (2004).

158. Pugh, K. R. *et al.* The relationship between phonological and auditory processing and brain organization in beginning readers. *Brain Lang.* **125,** 173–183 (2013).

159. Weiner, K. S. S., Yeatman, J. D. & Wandell, B. A. The posterior arcuate fasciculus and the vertical occipital fasciculus. *Cortex* **97,** 274–276 (2017).

160. Saxe, R., Brett, M. & Kanwisher, N. Divide and conquer: a defense of functional localizers. *Neuroimage* **30,** 1088–1089 (2006).

161. Glezer, L. S. & Riesenhuber, M. Individual variability in location impacts orthographic selectivity in the "visual word form area". *J. Neurosci.* **33,** 11221–6 (2013).

162. Avants, B. & Gee, J. C. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *Neuroimage* **23,** S139-50 (2004).

163. Vandermosten, M. *et al.* A DTI tractography study in pre-readers at risk for dyslexia. *Dev. Cogn. Neurosci.* **14,** 8–15 (2015).

164. Peterson, R. L. & Pennington, B. F. Developmental Dyslexia. *Annu. Rev. Clin. Psychol.* (2015). doi:10.1016/j.pcl.2007.02.015

165. Yeatman, J. D., Richie-Halford, A., Smith, J. K. & Rokem, A. AFQ-Browser: Supporting reproducible human neuroscience research through browser-based visualization tools. *bioRxiv* (2017).

166. Rezaie, R. *et al.* Temporo-parietal Brain Activity as a Longitudinal Predictor of Response to Educational Interventions among Middle School Struggling Readers. *J. Int. Neuropsychol. Soc.* **17,** 875–885 (2011).

167. Rezaie, R. *et al.* Engagement of Temporal Lobe Regions Predicts Response to Educational Interventions in Adolescent Struggling Readers. *Dev. Neuropsychol.* **36,** 869–888 (2011).

168. Munafò, M. R. *et al.* A manifesto for reproducible science. *Nat. Hum. Behav.* **1,** 1–9 (2017).

169. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning*. **1,** (2009).

170. Zou, H., Zou, H. & Hastie, T. Regularization and variable selection via the Elastic Net. *J. R. Stat. Soc. Ser. B* **67,** 301--320 (2005).

171. Breiman, L. Statistical Modeling: The Two Cultures. *Stat. Sci.* **16,** 199–215 (2001).

172. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12,** 2825–2830 (2011).

Recruitment and Retention Plan

We will recruit 80 (40 intervention, 40 control) children between 9 and 11 years of age from the University of Washington Reading and Dyslexia Research Database (UW-RDRD). The UW-RDRD is a subject pool that is maintained by the PI's laboratory. It includes hundreds of children in the Seattle area who are interested in participating in reading research and have undergone extensive behavioral testing and questionnaires. Currently about 1,000 children are enrolled in the database and there is an enrollment rate of 200-300 subjects per year. This resource supports the timely recruitment of specific samples of subjects including children with dyslexia and as proposed in the present study. We don't anticipate any issues with subject recruitment: due to the popularity of our previous NSF-funded intervention study there are dozens of that have already requested to be put on a wait-list for future intervention studies after the previous study was fully enrolled.

The Yeatman Lab has been very successful in retaining subjects for longitudinal studies. For example, the previous intervention study (see Pilot data in Research Strategy) involved 4 scans collected longitudinally and **there was no subject attrition**. This success has been achieved by hiring laboratory personnel who are good with children and their families. All of our research assistants have years of experience working with children and are exceptional at making sure families feel that their commitment to research is valued, and that children enjoy coming in for their visits. Research assistants are knowledgeable about the scientific studies they work on and engage children and their families in the process of scientific research. In our experience, by instilling a genuine interest in the research, families feel engaged, committed, and excited to continue participating in our studies.

Protection of Human Subjects

1 - Risks to the Subjects

1a - Human Subjects Involvement, Characteristics and Design We will recruit 80 children (9-11years of age) from the University of Washington Reading and Dyslexia Research Database (UW-RDRD). The UW-RDRD is a subject pool that is maintained by the PI's laboratory. It includes hundreds of children in the Seattle area who are interested in participating in reading research and have undergone extensive behavioral testing and questionnaires. This resource supports the timely recruitment of specific samples of subjects.

For all subjects, inclusion criteria include having no major contraindication for MRI (braces, metal implants, pacemakers, vascular stents, or metallic ear tubes). Because the study involves measurements of reading and language ability, new recruits will be native English speakers. Subjects have no history of neurological disorder, significant psychiatric problems or ADHD diagnosis. We also exclude claustrophobic subjects since an MRI might be uncomfortable for them.

40 subjects will be recruited for the intervention study and 40 subjects will be recruited as control subjects. As described in detail in the Research Strategy Intervention subjects will be children between nine and eleven years of age who are below the 25th percentile in timed (TOWRE composite) and untimed (WJ Basic Reading Skills) single word decoding abilities, and within the normal range (+/- 1 SD) on measures of general cognitive abilities (WASI). Of the control subjects, half will be matched in terms of reading skills and half will have typical reading skills (see Research Strategy).

1b – Study procedures, materials and potential risks

All data obtained in this study is obtained for research purposes. Sources of research material include: 1) screening information, 2) behavioral tests, and 3) MRI data. All of the data obtained is confidential and research reports never use data from a named individual. When data are coded for computer analysis, all participants' names are kept separately by research staff and findings made available only to legitimate agents of the participant (e.g., personal physician, etc.) with the permission of the participants or legal guardian.

The experience gained through previous research projects on reading skills in children has helped the PI develop secure, efficient, and useful techniques for data handling. For each subject that enters the database, their data is split between two separate, password-protected data stores connected by a unique identifier All identifying information is stored in a registry that is used for maintaining contact with subjects and coordinating consent procedures. Data, including online questionnaires and reading scores from lab visits, is stored in a de-identified repository. Moreover, a folder is created that contains all cognitive tests and

imaging data. This folder and the repository serve as a permanent archive of original subject data. Extensive precautions are taken to insure the privacy of subjects and the confidentiality of data. Specifically, subject identity is numerically coded on all pages within subject folders and in the database. All subject folders are kept in confidential, locked filing cabinets at the Institute for Learning and Brain Sciences. The database is password-protected and is hosted on a University of Washington server. Only personnel directly associated with the grant have access to subject information. Subject background information including name, gender, ethnicity, and relevant medical and personal information is kept in the registry, which represents a centralized, restricted access location that is separate from the behavioral and imaging results. This data handling system is efficient, guarantees subject confidentiality, and serves research needs expeditiously.

There are no risks to individuals from the interviewing, testing (paper-and- pencil or computer). There is minimal non-significant risk (i.e., "research not involving greater than minimal risk" as defined by DHHS) associated with the MRI scanning procedure. The MRI machine uses a strong magnet and radio frequency magnetic fields to make images of the brain. The magnetism and radio frequency magnetic fields do not cause harmful effects at the levels used in the MRI machine.

2 - Adequacy of Protection Against Risks

2a - Recruitment and informed consent

All child subjects are recruited from King County, and surrounding area. Flyers, ads in local newspapers, and notices placed in school, community, electronic bulletin boards and reading clinics serve as the primary means of recruitment into the UW-RDRD. We then invite subjects to participate in specific studies based on meeting the specific study enrollment criteria. The study is explained in detail to all potential participants. Parental guardians of all human subjects included in this study sign a written, informed consent. For children, informed consent is obtained from the subject and the subject's legally authorized representative (parent or legal guardian). Consent forms are written in language understandable to the subjects and their representatives, and subjects are allowed sufficient opportunity to consider whether or not to participate so as to minimize the possibility of coercion or undue influence. An investigator or research assistant explains the study to the child and a parent or guardian who is asked whether they have concerns with participation in the study. If any hesitation is noted, the subject is not included in the study.

All subjects are told that their participation is voluntary, and that during the study they may withdraw from the research at any time. Some children may want to withdraw from the study while their parents feel that they have incentives to participate in the study. During the prescreening as well as the scanning sessions, we query children both in the absence and presence of their parents. If they show less enthusiasm or they are reluctant when their parents are not present, we discontinue participation. Subjects are also informed that reluctance to participate does not in any way compromise the availability of their health care or their eligibility to participate in other studies. Also, once the study has begun, subjects can change their mind at any time about whether they wish to continue in the project. This does not affect their medical care.

Written consent will be documented with University of Washington IRB-approved consent and assent procedures.. Remuneration for expenses related to participation in the study will be offered at $60 for a session involving neuroimaging and $20 for behavioral sessions. All subjects in this study will be informed if any medically important information is learned from their research participation that may significantly affect their current diagnosis or treatment, or influence their willingness to continue participation in this study.

2b - Protection against risks

The consent forms list any possible risks to the MRI scanning procedure. The operators of the MRI at the MRI center are well trained to use protocols and procedures to ensure that the MRI equipment is used correctly to minimize any risks. The imaging coils and software are also tested for safety. National and UW guidelines have been developed for these machines, and these recommendations will be strictly followed.

Rigorous attempts will be made to desensitize participants to the MRI scan acquisition procedure as described in Research Design and Methods. Participants who have not experienced an MRI scan previously will undergo training on the UW MRI simulator, which acclimate them to the actual MRI scanning procedure. However, some subjects may still feel anxious before or during the MRI scan. We will continually assess

for any sign of discomfort during implementation of the research protocol. Subjects can alert the console operator that they need help or need to ask a question by squeezing a hand-held device during the scan. Scans will be terminated at a subject's request or if the console operator or individuals in the room detect significant subject discomfort. To date, the PI has had significant success in scanning children and adults.

3 - Potential benefits of proposed research to human subjects and others

Individuals participating in this study as intervention subjects will receive a high-quality reading intervention program free of charge. Based on our previous research, this intervention is likely to lead to substantial improvements in reading skills. Individuals participating in this study as control subjects may not receive any direct benefits. The general benefit to medical science will be in furthering our understanding of the neurobiological mechanisms underlying the developmental of skilled reading and the development of personalized reading intervention programs. It is our belief that the potential benefits from this study to the advancement of scientific knowledge (and, therefore, indirectly to participants and their families), substantially outweigh the minimal risk to human subjects. Children may learn important reading skills over the course of the study.

4 - Importance of the knowledge to be gained

This research offers promise in furthering our understanding of both brain function and structure in children. The studies proposed will provide new opportunities to explore fundamental issues of cognitive neuroscience that relate cognitive functions to brain organization. It will provide the foundation for treatment and remediation options for individuals with dyslexia, and elucidate the mechanisms of learning.

Data and Safety Monitoring Plan

The study involves behavioral testing (paper and pencil, computer and interview) and MRI scanning and all procedures involve minimal non-significant risk. The intervention also involves no risk to the subjects: participants work one-on-one with a trained instructor to practice reading related skills much like a typical educational setting. Our safety monitoring is commensurate with this minimal level of risk and, as outlined in the Protection of Human Subjects, involves ensuring the safety and confidentiality of each of the procedures (e.g., screening for metal before the MRI). The aspect of this study that makes it a clinical trial is the reading instruction program, and we have taken care to design a procedure to ensure the fidelity of the intervention approach. This will fidelity monitoring will be carried out by co-Investigator Roxanne Hudson and PI Jason Yeatman. This fidelity monitoring plan is laid out in the Research Strategy as it is a critical component of the research methodology:

Intervention protocol: Intervention will be administered to participants in Lindamood-Bell learning centers in the Seattle area four hours a day, five days a week, for eight consecutive weeks during summer vacation. Trained tutors will work one-on-one with each child. The tutors will be trained employees of Lindamood-Bell, who have extensive experience with the intervention protocol and have completed a 3-week training program (see more below). Research subjects will be provided with the *Seeing Stars* intervention free of charge (see Lindamood-Bell letter of support) and the curriculum will be administered in Lindamood-Bell learning centers but with extensive and independent monitoring by trained personnel on the research team (see Fidelity Monitoring below).

Importantly, the scientific research will be completely independent of the administration of the intervention to eliminate any potential conflicts of interest. Lindamood-Bell will not have access to any information the PI's laboratory collects on the subjects, including subject attrition, subject testing results which are conducted in the PI's laboratory or MRI data collection.

The Lindamood-Bell "*Seeing Stars*" instruction program teaches phonological and orthographic processing through a combination of mental imagery and sensory-motor learning. It is laid out in detail in the *Seeing Stars* manual which is publicly available and has been extensively described in other publications. Children practice visualizing the orthography of words, starting from simple consonant-vowel syllables and systematically working into more complex consonant-vowel-consonant pairings. The instruction focuses on segmenting and blending strings of phonemes into words and visualizing the relationship between the articulatory elements, and their representation as visual symbols. Phonological awareness is systematically built through articulatory exercises where children are taught to attend to the relationship between motor movements of the mouth and tongue, and speech sounds in words, and the corresponding letters. As children build a stronger foundation of phonological and orthographic knowledge, additional

practice with increasingly complex sight-word identification and phonological decoding is layered into the intervention.

Intervention fidelity and monitoring: The intervention curriculum is detailed in the publicly available *Seeing Stars* Teacher's Manual allowing the procedure to be precisely reproduced. Fidelity to the published curriculum will be ensured in two ways. First, Lindamood-Bell has established a rigorous methodology for ensuring the quality and fidelity of instruction: (1) Each instructor completes a 3-week training program; (2) At each learning center a "coach" with multiple years of experience monitors each instructor's progress and provides feedback, ensuring that the instructor is properly progressing through the steps of the curriculum with their student; and (3) Each coach provides weekly reports to the regional director to ensure consistency of implementation across learning centers. This fidelity monitoring will be recorded and databased for each research study participant. Second, one session a week (for each child) will be recorded with a digital video camera and a member of the research team (independent of Lindamood-Bell) will score the fidelity of implementation for each session based on: (1) adherence to the procedures of the intervention using a 5-point implementation behavior scale and (2) engagement of the child. Child engagement will be measured using a 5-second partial interval time sampling procedure on the same videotaped session used for procedural fidelity. Reliability procedures will require research staff to rate a sample of training videos and for their ratings to correlate at $r \geq .80$ prior to data collection. The researcher will provide regular feedback to the Lindamood-Bell regional director based on the independent assessment of fidelity so that instructors can receive coaching in the case of deviations from the published curriculum. Finally, scoring will be databased so that details of the intervention delivery can be examined as covariates in statistical models of intervention effects. Roxanne Hudson, a member of the research team with extensive experience monitoring and ensuring the fidelity of educational interventions will be primarily in charge of the monitoring procedures and will work closely with Yeatman and the rest of the research team to ensure the procedures are carried out faithfully.

Statistical Design and Power

Aim 1: Aim 1 targets the relationship between white matter plasticity and learning and seeks to quantify changes in the white matter that occur with intervention. There will be 40 intervention subjects and 40 control subjects. Linear mixed effects models will be used to compare white matter changes during the intervention period to (a) the control period and (b) the control group. Simulations based on pilot date in 26 subjects indicate exceptional statistical power: 0.98 for the arcuate fasciculus (primary tract of interest) and 0.95 for the inferior longitudinal fasciculus (secondary tract of interest) at alpha=0.05. This power analysis takes into account an expected 10% attrition rate. In the pilot study there was no subject attrition. Additional details are provided on page 6 (see Figure 2 caption) of the Research Strategy.

Aim 2: Aim 2 targets changes in the computations performed by the visual word form area, a region that is critical for rapid and automatic word recognition. Functional magnetic resonance imaging data will be used to measure the selectivity for words compared to other stimuli in this region. This aim uses the same subjects as Aim 1. Linear mixed effects models will be used to compare white matter changes during the intervention period to (a) the control period and (b) the control group. Based on pilot data, a power analysis confirms excellent statistical power: mean % BOLD increase = 0.15, SD = 0.034, bias corrected SD = 0.036, power > 0.95 at α=0.05. Additional details are provided on page 10 of the Research Strategy.

Aim 3: Aim 3 develops a predictive model of individual differences in learning within the intervention group (40 subjects) and is, therefore, not listed as a clinical trial outcome measure. There is still excellent statistical power for accomplishing this aim: regression analysis will be used to predict individual differences in reading skill improvement and pilot data demonstrates statistical power of 0.96 (at $\alpha = 0.05$) for our primary predictor (left arcuate fasciculus diffusivity). Additional details are provided on page 11 of the Research Strategy and in Figure 4. The analysis for Aim 3 will also be pre-registered to avoid publication bias. Machine learning will then be used to test the hypothesis that prediction accuracy can be improved (compared to the aforementioned regression analysis) based on learning linear and non-linear combinations of features. This is considered an exploratory sub-aim and we have defined a detailed plan to use cross-validation to control for over-fitting and ensure out of sample generalization at each stage.