Inventory of Supplemental Information for NEURON-D-09-00552R2:

Figure S1. Results of parametric voxel-wise analyses conducted on unsmoothed FA data. The two panels correspond to those in Figure 1 of the main paper and are discussed on pp. 7 and 13 of the main paper.

Figure S1. Results of non-parametric voxel-wise analyses conducted on unsmoothed FA data. The two panels correspond to those in Figure 1 of the main paper and are discussed on pp. 7 and 13 of the main paper.

Figure S3. Results of a parametric voxel-wise regression analysis showing a positive relationship between non-word decoding ability and FA and discussed on p. 9 of the main paper.

Figure S4. Results of multiple regression analyses assessing the relationship between non-word reading ability and FA, radial diffusivity, and axial diffusivity after controlling for age in the region showing a difference in FA between groups at the pre-remediation scan, and discussed on p. 9 of the main paper.

Table S1. Report of means, standard deviations, and tests for group differences for all behavioral measures collected at the time of the pre-remediation scan and discussed on p. 5 of the main paper.

Table S2. Report of means, standard deviations, and tests for differences across time for reading ability measures collected for the poor readers at the time of the pre-remediation and post-remediation scans and discussed on p. 6 of the main paper.

Supplemental Text. Supplemental Results and Discussion mentioned throughout the main paper supporting the conclusions that the treatment groups were equated on behavioral measures and on imaging measures prior to treatment and the changes in these measures were similar across different treatment groups. Additional results are also presented showing that the conclusions are supported by different analysis methods.

Supplemental References.

Supplemental Data for:

Altering cortical connectivity: Remediation-induced changes in the white matter of poor readers

Timothy A. Keller[1] & Marcel Adam Just[1]

[1]Center for Cognitive Brain Imaging, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

SUPPLEMENTAL FIGURES

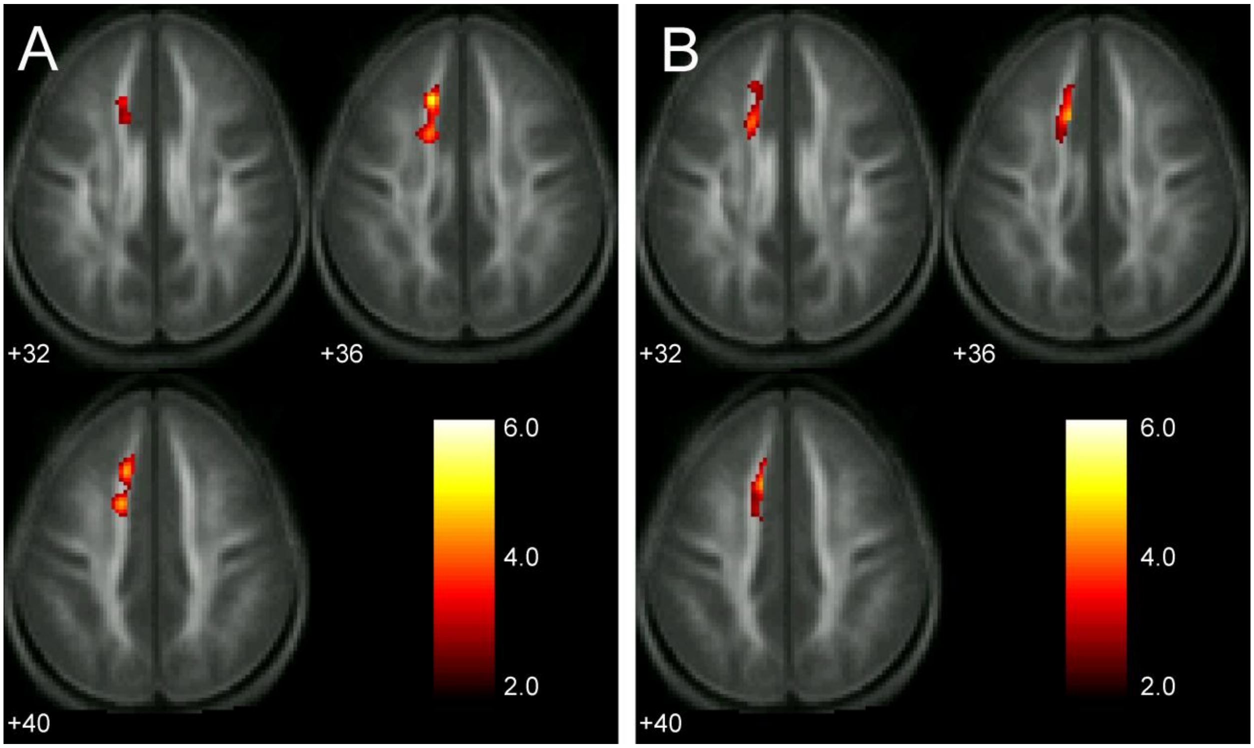

Figure S1. Results of parametric voxel-wise analyses conducted on unsmoothed FA data. (A) Regions where the treated poor reader group showed an increase in FA between the pre-remediation and post-remediation scans. For the unsmoothed data the remediation effect is found in two separate clusters in the same region as that reported for the smoothed data in Figure 1A (peak t(34) = 4.86, at Montreal Neurological Institute (MNI) coordinates -14 26 36, spatial extent = 18 voxels, p = .08, corrected, and peak t(34) = 4.04 at MNI coordinates -16 6 38, spatial extent = 23 voxels, $p < .05$, corrected). (B) Region showing a significant difference in FA between good readers and all poor readers at the initial scan (peak t(70) = 4.15, at MNI coordinates -12 18 38, spatial extent = 40 voxels, $p < .05$, corrected). Statistical maps are overlaid on normalized FA images averaged across all participants in both scans. The MNI z-coordinate is shown at the bottom left of each axial slice. Color scale represents t-values. Reported p values for t-tests are corrected for multiple comparisons across all of white matter based on cluster extent in the context of Gaussian Random Field theory.

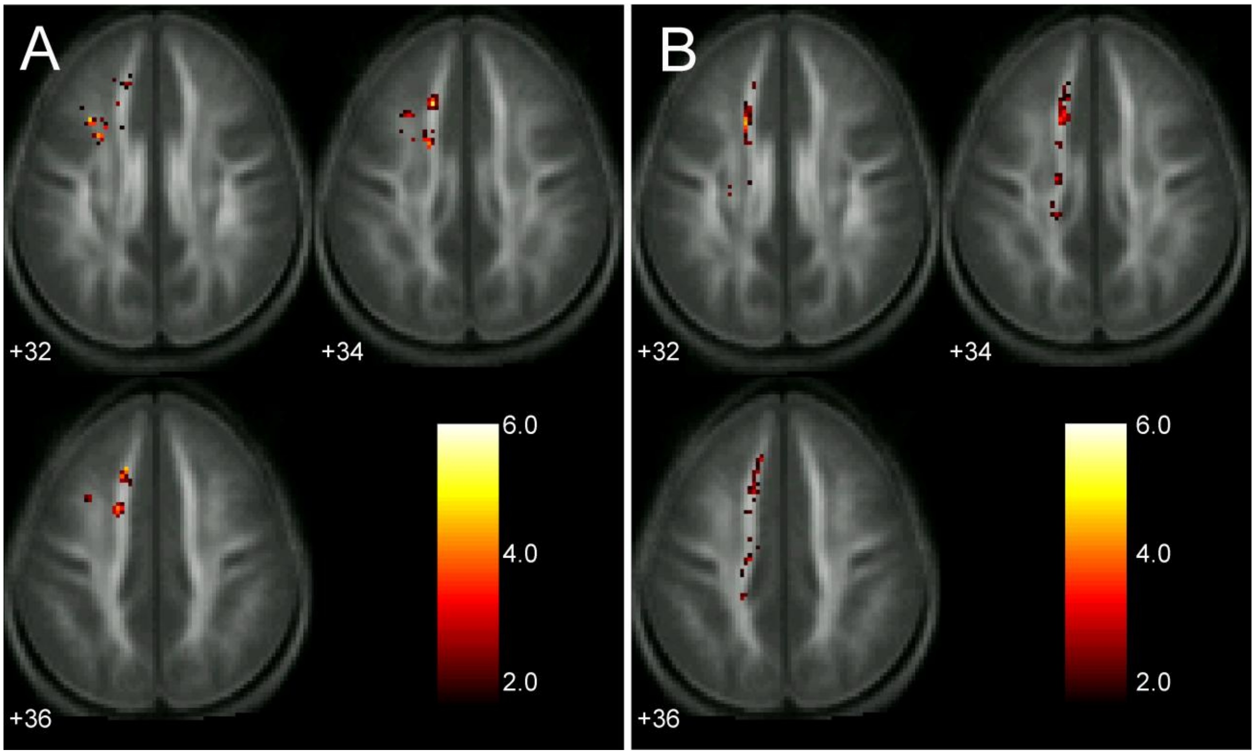

Figure S2. Results of non-parametric voxel-wise analyses conducted on unsmoothed FA data. (A) Voxels where the treated poor reader group showed an increase in FA between the pre-remediation and post-remediation scans. The largest increase is found in the same region as that reported for the smoothed data in Figure 1A (peak paired-t(34) = 4.59, at Montreal Neurological Institute (MNI) coordinates -12 26 36, p < .05, corrected. (B) Voxels showing a significant difference in FA between good readers and all poor readers at the initial scan (peak two-sample t(70) = 4.15, at MNI coordinates -12 18 38, p < .05, corrected). The MNI z-coordinate is shown at the bottom left of each axial slice. Color scale represents t-values. Reported p values for t-tests are corrected for multiple comparisons across all of white matter based on cluster extent using the Threshold-Free Cluster Enhancement method (Smith & Nichols, 2009).

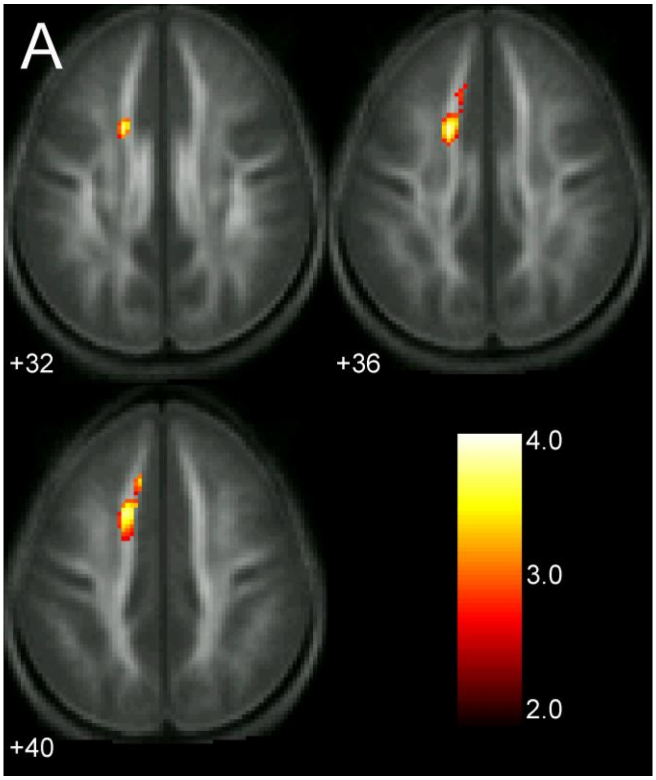

Figure S3. Area showing a positive relationship across the entire sample of chidren between raw WRMT Word Attack scores and FA (controlling for age) at the pre-remediation scan (peak t(69) = 3.89 at MNI coordinates -16 12 34, p < .005 corrected for multiple comparisons). The MNI z-coordinate is shown at the bottom left of each axial slice. Color scale represents t-values.

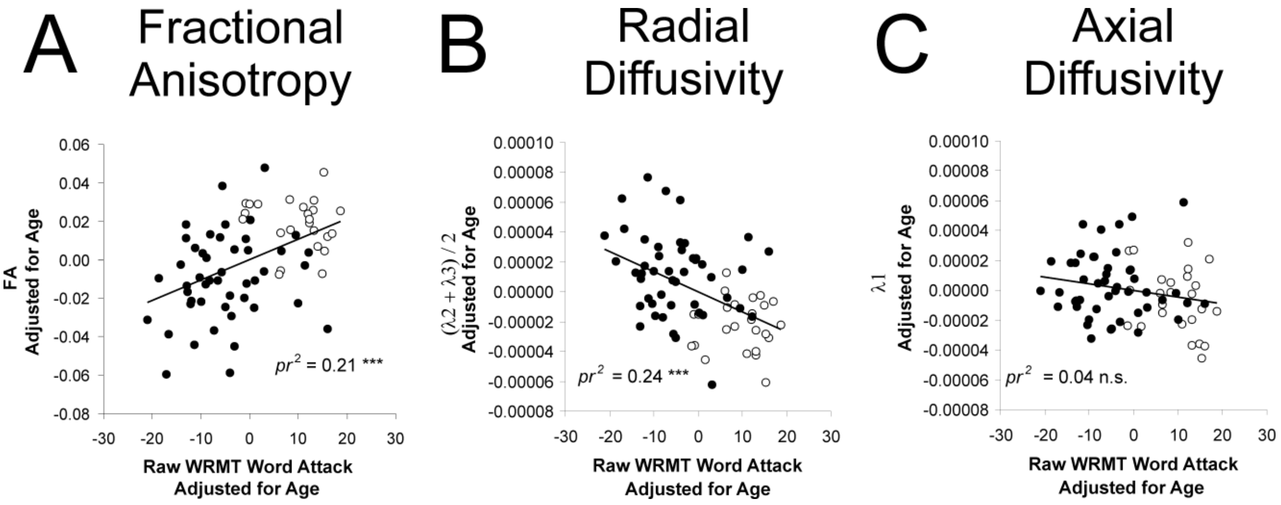

Figure S4. Continuous relationships between phonological decoding skill (age-standardized WRMT-R Word Attack) and three DTI measures at the pre-remediation scan across the region showing a reliable group difference in FA. Good readers are shown as open circles and poor readers as filled circles. pr^2-values are the squared partial correlation coefficients from a regression model including age as an additional covariate. (A) Fractional anisotropy is reliably positively related to phonological decoding ability. (B) Radial diffusivity is reliably negatively related to phonological decoding ability. (C) Axial diffusivity shows only a modest negative relationship to reading ability. *** $p < .0001$.

SUPPLEMENTAL TABLES

Table S1. Reading ability and other scores collected at the time of the pre-remediation scan.

	Group						Group Difference		
Measure	PR		PC		GC		PR - PC	PR - GC	PC - GC
N (n female)	35 (8)		12 (6)		25 (8)				
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*t(45)*	*t(58)*	*t(35)*
Age (years)	10.0	(1.1)	10.2	(1.2)	9.8	(1.0)	0.71	0.81	1.41*
Grade-standardized PPVT	96.5	(13.2)	97.5	(12.9)	110.8	(12.8)	0.04	4.43**	3.47**
Raw PPVT	120.9	(19.0)	128.2	(20.2)	140.5	(20.2)	1.00	4.17**	2.20**
Age-standardized WRMT-R Word Attack	92.2	(11.0)	95.7	(7.0)	114.7	(11.6)	-1.02	7.63***	5.23***
Raw WRMT-R Word Attack	18.0	(8.6)	22.5	(6.8)	33.9	(6.2)	-1.66	7.94***	5.11***
Age-standardized WRMT-R Word ID	90.5	(10.4)	88.8	(9.3)	108.4	(8.2)	0.53	7.13***	6.52***
Raw WRMT-R Word ID	54.2	(11.3)	53.0	(17.2)	73.4	(7.8)	0.28	7.28***	5.00***
Age-standardized WRMT-R Passage Comprehension	94.3	(11.2)	96.8	(10.5)	111.8	(10.4)	-0.68	6.15***	4.10**
Raw WRMT-R Passage Comprehension	31.7	(7.1)	34.3	(7.1)	42.2	(6.8)	-1.07	5.76***	3.30**
Age-standardized WRMT-R Basic Skills Cluster	90.6	(10.2)	89.7	(8.6)	114.3	(12.3)	0.29	8.14***	6.24***
Age-standardized WRMT-R Total Reading Cluster	91.4	(9.7)	90.3	(9.8)	111.8	(9.4)	0.35	8.14***	6.45***
Grade-standardized TOWRE Sight Word Efficiency	86.7	(9.1)	87.5	(5.7)	110.3	(11.1)	-0.30	9.08***	6.70***
Raw TOWRE Sight Word Efficiency	49.3	(9.9)	51.1	(6.0)	70.8	(8.0)	-0.58	8.96***	7.52***
Grade-standardized TOWRE Phonological Decoding Efficiency	83.0	(10.3)	84.0	(6.8)	107.9	(12.9)	-0.31	8.33***	6.00***
Raw TOWRE Phonological Decoding Efficiency	15.8	(8.7)	17.0	(7.6)	35.8	(9.4)	-0.42	8.47***	6.03***
Grade Standardized TOWRE Composite	81.7	(10.8)	82.9	(6.3)	111.0	(13.2)	-0.36	9.42***	6.94***
Raw TOWRE Composite	65.1	(17.1)	68.1	(11.9)	106.6	(15.5)	-0.55	9.60***	7.57***
Grade-standardized GRADE	96.3	(11.1)	99.2	(12.7)	114.6	(7.0)	-0.74	7.26***	4.79***
Raw GRADE	13.8	(4.8)	15.1	(5.7)	24.4	(3.1)	-0.78	9.85***	6.52***
Grade-standardized AIMS WEB	91.3	(7.2)	91.5	(6.5)	114.4	(11.6)	-0.08	9.49***	6.34***
Grade-standardized WJIII Spelling	90.9	(11.7)	92.4	(8.7)	113.7	(11.3)	-0.4	7.53***	5.72***
Raw WJIII Spelling	27.8	(5.7)	28.6	(3.7)	37.4	(5.1)	-0.44	6.76***	5.36***
Grade-standardized WJIII Calculation	97.4	(9.0)	96.5	(6.8)	118.6	(11.0)	0.33	8.21***	6.39***
Raw WJIII Calculation	15.6	(3.3)	15.7	(3.0)	21.4	(4.5)	-0.04	5.73***	3.99**

Notes. PR - poor readers who would receive remediation, PC - poor reader controls, GC - good reader controls. * *P* < .05, ** *P* < .005, *** *P* < .0001.

Table S2. Raw reading ability scores collected at the time of the post-remediation scan, and differences from pre-remediation scan.

	Group				**Change in Scores (Time 2 - Time 1)**			
Measure	PR		PC		PR		PC	
	Mean	*SD*	*Mean*	*SD*	*Diff*	*t(34)*	*Diff*	*t(11)*
WRMT-R Word Attack	25.1	(8.2)	24.3	(3.1)	7.1	6.39***	1.8	1.16
WRMT-R Word ID	61.0	(10.3)	59.5	(11.5)	6.8	4.35***	6.5	1.14
WRMT-R Passage Comprehension	36.8	(11.5)	34.3	(3.5)	5.1	2.46*	0.0	0
TOWRE Sight Word Efficiency	56.5	(10.7)	59.6	(8.7)	7.2	5.76***	8.5	4.13**
TOWRE Phonemic Decoding Efficiency	24.5	(17.1)	24.8	(5.6)	8.7	3.65**	7.8	4.15**
TOWRE Composite	80.9	(19.9)	84.4	(11.1)	15.8	7.71***	16.3	5.18**
AIMSWEB Total Errors	10.3	(7.0)	13.3	(8.9)	-6.4	-3.91**	-1.8	-0.77
AIMSWEB Total Correct	271.9	(82.6)	273.1	(55.5)	58.7	8.91***	54.2	5.54**
GRADE	18.1	(9.1)	16.0	(6.8)	4.3	3.18**	0.9	0.96
WJIII Spelling	30.8	(6.9)	31.1	(2.5)	3.0	3.70**	2.5	3.68**
WJIII Calculation	19.0	(3.8)	19.1	(3.5)	3.3	6.55***	3.4	5.40**

Notes. PR - poor readers who would receive remediation, PC - poor reader controls. * $p < .05$, ** $p < .005$, *** $p < .0001$. Data were not collected for good readers at this time.

SUPPLEMENTAL RESULTS AND DISCUSSION

Equated reading and other abilities among poor readers at the pre-remediation scan. The randomized assignment of poor readers to treatments succeeded in equating poor readers who would receive remediation (PR) with the poor reader control group (PC) who wouldn't, with respect to age and all test scores acquired prior to the initial scan and to the remediation, as shown by the two-sample t-tests in Table S1. Additional one-way ANOVAs on each measure comparing the five groups of poor readers (four instruction program groups and one group that received no remedial instruction) indicated that there were no reliable differences among any of the groups of poor readers for any of the measures prior to the treatment. The good reader control group (GC) was slightly younger than each of the two groups of poor readers, and scored significantly higher on every measure of vocabulary knowledge, reading ability, and spelling and calculation ability, collected at the initial scan.

Similarity of remediation effects on reading scores among poor readers who received slightly different treatments. Although it was not a focus of the neuroimaging study due to the small sample size per treatment, we also assessed whether there were differential effects of the four remedial programs on reading ability with two-factor ANOVAs (4 remedial treatment groups by 2 times of scan (pre-/post-remediation)) conducted on of the all measures in Table S1. Eighteen of the 22 ability measures showed a reliable main effect of time after correction for the number of tests considered, with an increase following the instruction. There were no measures that showed a main effect of instruction group and no measures that showed an interaction between instruction group and time of testing after correction for multiple tests. Thus, there was very substantial evidence that reading ability improved among poor readers who received some form of intensive remediation program, but no indication that the different instruction types were differentially effective in improving any specific reading skill.

Remediation effects on reading scores. Comparisons of the subscores of the reading measures of all remediated poor readers with control poor readers before and after the remediation period indicated that the remedial instruction improved the phonological decoding skills of the poor readers who received the instruction. Table S2 presents the raw scores on all reading ability measures collapsed across the four groups of poor readers that received some form of intensive remedial reading instruction, and compares them with the scores of the control group of poor readers who received no remedial instruction. Because the poor reader control group had continued to receive their normal reading curricula in the classroom during the course of the study, some improvement in raw scores among this group would be expected and indeed there were a number of measures on which the poor reader controls did improve. A 2 (treatment vs. control poor readers) by 2 (time) ANOVA conducted for raw scores on the WRMT-R Word Attack subtest (measuring the ability to decode pronounceable non-words) showed a reliable group by time interaction ($F_{1, 45} = 6.18$, $p < .05$), resulting from an improvement in scores between the two tests among poor readers who received remedial instruction, but no change in scores among poor readers who did not receive the remedial instruction (see Table S2). No other measure from any of the tests showed a reliable group by time interaction. Considered together, these results suggest that the primary effect of the remedial instruction was an improvement in phonological decoding as measured by the Word Attack subtest.

Equated FA among poor readers at the pre-remediation scan. A whole-brain, voxel-wise ANOVA confirmed that there were no significant differences in FA among the five groups of poor readers at the pre-remediation scan. This analysis revealed no clusters that reliably differed among the poor reader groups after correction for multiple comparisons. This was true even when the analysis was restricted to the cluster that showed a reliable difference in FA between good and poor readers at the pre-remediation scan. In addition, an analysis of the FA data averaged across voxels within this same volume of interest also failed to show any reliable differences among the five groups of poor readers at the pre-remediation scan (Failure Free

Mean = 0.35, SEM = 0.009; Spell Read Mean = 0.36; SEM = 0.007; Corrective Reading Mean = 0.36, SEM = 0.008; Wilson Reading Mean = 0.36, SEM = 0.008; Control Group Mean = 0.37; SEM = 0.007; $F_{1, 42} = 1.30$, $p = .28$).

Similarity in FA changes among poor readers who received slightly different treatments. To assess whether there were differences in the effects of the four instructional programs on the change in FA before and after remediation, the FA data were submitted to a 4 (group) by 2 (time) whole-brain, voxel-wise mixed ANOVA. The resulting F-maps for the test of the interaction were thresholded at $p < .005$ with a liberal cluster-size threshold of 50 voxels. No clusters survived this threshold. Additional voxel-wise, pair-wise contrasts conducted between each of the groups for the change in FA across the two scans found no clusters in any contrast that reliably differed between any of the pairs of groups at $p < .005$ for t-value and $p < .05$ using a cluster-size threshold to correct for multiple comparisons. Averaged FA data from each participant in each scan extracted from the volume of interest showing a reliable difference between good and poor readers at the pre-remediation scan were also submitted to a 4 (group) by 2 (time) ANOVA. Although there was a large main effect of time in this analysis ($F_{1, 31} = 12.48$, $p < .005$) resulting from an increase in FA across the groups, there was no main effect of remedial treatment group ($F_{3, 31} = 0.25$, $p = .86$) and no group by time interaction ($F_{3, 31} = 1.37$, $p = .27$). In sum, the analyses of both the behavioral data and the FA data provide no evidence of differences among the various remedial instruction programs. All the analyses presented in the main article and the subsequent analyses presented here therefore collapse across this factor.

Effect of spatial smoothing on the voxel-wise analyses of FA. The voxel-based methods used for comparing DTI data across groups have been questioned because of the dependence of the parametric statistical tests on the size of the spatial filtering kernel that is used to satisfy assumptions of Random Gaussian Fields Theory (Jones et al. 2005). To assess whether the choice of filter size used here influenced the ability to detect group FA differences and remediation effects among poor readers in left temporo-parietal white matter, the main voxel-wise analyses of the FA data reported in the main article were repeated without any spatial filtering of the data. Comparison of this outcome (shown in Figure S1) with Figure 1 in the main article indicates that although the sensitivity for detecting effects of group, time, and the interaction was enhanced by spatial filtering, the main conclusions of the study are not altered.

Non-parametric tests of changes and group differences in FA. The purpose of smoothing in the method used in our study is not only to enhance the signal to noise ratio in the data, but also to ensure that the assumptions underlying Gaussian random field theory, on which the correction for multiple comparisons are based, are met. It has been shown that as the smoothing filter width is decreased, these assumptions are increasingly violated for FA data (Jones et al. 2005). We therefore also explored alternative non-parametric, permutation-based methods for thresholding and correcting for multiple comparisons (Nichols and Holmes, 2002) in the analyses of the unsmoothed FA data. The two main contrasts of interest, change in FA among the PR group (paired t-test) and the good reader - poor reader group difference in FA at the pre-remediation scan (two-sample t-test) were tested in voxel-wise analyses by submitting the unsmoothed FA data to 5,000 permutations to generate null distributions of the statistics, and the threshold-free cluster enhancement method (TFCE, Smith & Nichols, 2009) was used to correct for multiple comparisons using the default neighborhood connectivity parameters. The resulting t-maps, presented in Figure S2, were thresholded at $p < .05$ after TFCE multiple comparison correction. A comparison between Figure S2 with Figure 1 in the main article indicates that our conclusions regarding both the longitudinal change in FA in the left anterior centrum semiovale among the poor readers that received intensive reading remediation, and the group difference in FA in this same region prior to the instruction phase, were both supported using these alternative, nonparametric methods.

Analyses restricted to an *a priori* left temporo-parietal region of interest. The voxel-wise analyses presented in the main article found no reliable differences in FA as a function of reading ability at the pre-remediation scan in left temporo-parietal white matter, in contrast to a number of previous studies (Beaulieu et al. 2005; Deutsch et al., 2005; Klingberg et al. 2000; Niogi & McCandliss, 2006; Odegard et al. 2009; Richards et al., 2008; Rollins et al. 2009). We also conducted voxel-wise analyses on both the smoothed and unsmoothed data restricted to a region of interest defined by a sphere of 1-cm radius centered at the mean of the MNI coordinates (-28 -22 26) reported as showing an FA relationship to reading ability in three previous studies (Beaulieu et al. 2005; Deutsch et al., 2005; Klingberg et al. 2000). No areas of reliably reduced FA among poor readers were found within this a priori ROI following a small volume correction for multiple comparisons, and no areas were reliably correlated with any reading ability measures in the area in voxel-wise analyses. However, an ROI-based analysis of the same region did suggest some consistency with earlier reports of reduced FA in this region. A one-tailed, two-sample t-test conducted for mean smoothed FA across the entire ROI indicated that poor readers had marginally lower FA in the ROI than good readers at the initial scan (Poor Reader Mean = 0.436, SEM = 0.003; Good Reader Mean = 0.442, SEM = 0.003; t(70) = 1.89, p = .06). A similar test on the unsmoothed data indicated reliably lower mean FA across the entire ROI for the poor readers at the initial scan (Poor Reader Mean = 0.449, SEM = 0.003; Good Reader Mean = 0.458, SEM = 0.003; t(70) = 2.27, p < .05), consistent with the previous findings. Additional analyses of mean radial and axial diffusivity in this a priori region of interest suggested non-significant trends toward higher radial diffusivity among poor readers at the initial scan in both the smoothed data (Poor Reader Mean = 6.24×10^{-4} mm^2/s, SEM = 4.01×10^{-6}; Good Reader Mean = 6.16×10^{-4} mm^2/s, SEM = 4.90×10^{-6}; t(70) = 1.56, p = .12) and the unsmoothed data (Poor Reader Mean = 5.82×10^{-4} mm^2/s, SEM = 3.40×10^{-6}; Good Reader Mean = 5.74×10^{-4} mm^2/s, SEM = 4.04×10^{-6}; t(70) = 1.82, p = .07), but there were no differences in axial diffusivity in either analysis (both p's > .25). Despite the indication of cross-sectional group differences in this area, additional analyses of longitudinal changes in these three DTI measures across the same a priori region of interest revealed no significant changes between scans among any of the groups (all p's > .15 for paired t-tests) and no reliable group x time interactions (all p's > .15 for interaction F-tests for FA, radial diffusivity, and axial diffusivity).

Thus, although there is some limited evidence that our sample of poor readers had reduced FA in the same area of left temporo-parietal white matter as reported in previous studies (Beaulieu et al. 2005, Klingberg et al. 2000; Deutsch et al. 2005), and that this lower FA was due to higher radial diffusivity in the area, there was no evidence of remediation-related changes in the microstructure of the white matter in this region.

Cross-sectional relationships at the pre-remediation scan among DTI measures, components of reading ability, and age. Good and poor readers were defined on the basis of total scores on the TOWRE prior to the remedial treatment, and thus the group difference in reading ability related to the reduction in FA involves measures of reading of both real words (which can be accessed via a direct orthographic route) and non-words (which can be produced only through a phonological route). Although these measures are highly correlated with each other and not independent (real words can also be read via a phonological decoding), it is nevertheless interesting to ask whether these different measures are differentially associated with FA across the entire range of reading ability. To investigate these relationships, the analyses used the scores from two subtests from the WRMT-R (Word Attack, measuring non-word reading, and Word ID, measuring real-word reading) as covariates, chosen because they were slightly less correlated with each other in our sample than other candidate measures and because they provide consistency with previous studies of the relationship between reading ability and FA (Beaulieu et al., 2005; Klingberg et al., 2000; Niogi and McCandliss, 2006). Pre-remediation FA data from all children were submitted to a voxel-wise multiple regression analysis with age, raw Word Attack score, and raw Word ID score, as independent variables.

The whole-brain, voxel-wise multiple regression analysis revealed no areas large enough to survive correction for multiple comparisons that were independently related to either Word Attack or Word ID scores (controlling for other variables in the model). Because the two reading ability measures were highly correlated across the entire sample of children ($r = .68$, $p < .0001$), we also tested separate models including only one or the other of the reading ability measures.

When only Word Attack scores and age were entered as independent variables, a large cluster showing a positive relationship between raw Word Attack scores and FA (controlling for age), was found in the same region of the anterior centrum semiovale that had shown a group difference in FA between good and poor readers (208 voxels with peak $t(69) = 3.89$ at MNI coordinates -16 12 34). This cluster is shown in Figure S3, and it survives correction for multiple comparisons when the region of interest is restricted to the area showing a group difference in FA at the pre-remediation scan ($p < .005$ corrected at the cluster level and $p < .009$ corrected at the voxel level). In contrast, when only Word ID scores and age were entered as independent variables, no areas of white matter were found to be reliably related to Word ID after controlling for age, even when the region of interest was restricted to the region showing a group difference in FA at the pre-remediation scan. These exploratory analyses suggest that the difference in FA between good and poor readers in the left anterior centrum semiovale at the first scan was due primarily to differences in phonological decoding ability rather than sight-word reading ability.

Consistent with previous studies of the development of white matter in this age range, there was a positive effect of age (controlling for both WRMT Word Attack and Word ID scores) on FA values in both hemispheres in this cross-sectional analysis. A large cluster of voxels with a peak near the right putamen (peak $t(68) = 5.39$, MNI coordinates 16 6 -8, spatial extent = 2894 voxels) extended inferiorly into the right anterior corona radiata and uncinate fasciculus and through the cerebral peduncle into the left hemisphere, eventually reaching the left putamen. Notably, this large cluster included areas consistent with the location of the right and left superior longitudinal fasciculi, indicating that FA in these important tracts was continuing to increase with development in this age range. There were no regions showing a reliable negative relationship between age and FA values after controlling for raw WRMT scores.

To further explore the relationships between phonological decoding and sight-word reading abilities with DTI measures, region of interest-based multiple regression analyses were carried out on the average FA, radial diffusivity, and axial diffusivity across all voxels in the cluster which showed a main effect of group at the pre-remediation scan. The full model regressing FA on age, raw WRMT-R Word ID scores, and raw WRMT-R Word Attack scores was significant ($F_{3,68} = 6.73$, adjusted multiple $R^2 = .19$, $p < .005$). There was no significant effect of age or raw Word ID scores, but there was a significant effect of raw Word Attack scores for this region (partial $r^2 = .10$, $t(68) = 2.75$, $p < .01$). When Word ID scores were dropped from the model, Word Attack scores accounted for 21% of the variance in FA, as shown in Figure S4A. When Word Attack scores were dropped from the model, Word ID scores explained 13% of the variance in FA. These results again indicate that the relationship between reading ability and FA in this region is primarily due to the phonological decoding aspects of reading ability.

A similar analysis assessed the relation between the two components of reading ability and radial diffusivity in the volume of interest showing a group difference at the pre-remediation scan. This analysis also indicated that the full model was significant ($F_{3,68} = 7.44$, adjusted multiple $R^2 = .21$, $P < .002$). As in the analysis of FA, only the partial regression coefficient for raw Word Attack scores was significant for the analysis of radial diffusivity in the region (partial $r^2 = .10$, $t(68) = -2.82$, $p < .01$). Dropping Word ID scores from the model resulted in raw Word Attack scores accounting for 24% of the variance in radial diffusivity after adjusting for age, as shown in Figure S4B. Dropping Word Attack scores from the model resulted in raw Word ID scores accounting for 15% of the variance in radial diffusivity after adjusting for age. With axial diffusivity as the dependent variable, however, the full regression model was not significant ($F_{3,68} = 1.76$, adjusted multiple $R^2 = .03$, $p = .16$), and none of the partial regression coefficients were reliable. For a model including only age and WRMT Word Attack, the latter variable accounted for only 4% of the variance in axial diffusivity (Figure S4C). For a model including only age and WRMT Word ID, the latter variable accounted for only 7% of the variance.

These cross-sectional analyses of diffusion measures acquired prior to remediation suggest that the region of the left anterior centrum semiovale showing a reduction in FA among poor readers may involve pathways of particular importance in phonological decoding and that the positive relationship between phonological decoding ability and FA is primarily due to a negative relationship between this ability and radial diffusivity.

SUPPLEMENTAL REFERENCES

Howe, K.B., & Shinn, M.M. (2002) Standard Reading Assessment Passages for Use in General Outcome Assessment: A Manual Describing Development and Technical Features. (Eden Prairie: MN: Edformation, Inc.)

S.M. Smith, M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Drobnjak, D.E. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, and P.M. Matthews. (2004) Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage *23(S1),* 208-219.

Smith, S.M. and Nichols, T.E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage *44*, 83-98.

Williams, K.T. (2001). Group Reading and Diagnostic Evaluation. (Circle Pines, MN: American Guidance Service).

Woodcock, R.W. (1998). Woodcock Reading Mastery Tests-Revised NU (WRMT-R/NU). (Circle Pines, MN: American Guidance Service).

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). Woodcock-Johnson III Tests of Achievement. (Itasca, IL: Woodcock Riverside Publishing).