Supplemental Materials

Supplemental Phenotypic Results

We reran the phenotypic analyses from Christopher et al. (2012) in our expanded sample. First we conducted phenotypic confirmatory factor analyses (CFAs) to ensure our variables loaded onto their respective latent constructs and to test for possible age effects between the two age groups (ages 8-10 and ages 11-16). Then we tested whether the cognitive and reading latent variables fit the data equally well for both age groups. Finally we tested whether the latent variables were measuring separable constructs by comparing the full models with four cognitive latent variables and three reading variables with more parsimonious models that combined together various combinations of the latent variables.

To alleviate issues related to nonindependence, one twin from each pair was selected at random and used for all phenotypic analyses. The CFA analyses were conducted using AMOS (Version 22.0.0; Arbuckle, 2013). Four different fit indices are reported to assess how well the models fit the raw data: chi-square (χ^2), chi-square difference ($\Delta\ \chi^2$), comparative fit index (CFI), and root-mean-square error of approximation (RMSEA). Because chi-square values are sensitive to sample size, the CFI and RMSEA are good alternative fit indices. In line with Hu and Bentler (1999), models with CFI values between .95 and 1.00 are considered to be good fits to the original data. Loehlin (1998) suggests that RMSEA values less than .10 show that a model has good fit.

Cognitive CFA. The hypothesized cognitive model with the ten cognitive variables loading onto four latent variables (working memory, inhibition, processing speed, and naming speed) fit both age groups well, with a significant improvement in model fit found when the residuals for the two CPT tasks were correlated (ages 8-10: $\Delta\ \chi^2(1) = 15.04$, $p < .01$ compared to model with no CPT residual correlation; ages 11-16: $\Delta\ \chi^2(1) = 5.86$, $p = .02$ compared to model with no CPT residual correlation): ages 8-10, $\chi^2(28) = 28.39$, $p = .44$, CFI = 1.00, RMSEA = .006, RMSEA 90% confidence interval (CI) = [.00, .041]; ages 11-16, $\chi^2(28) = 37.72$, $p = .10$, CFI = .986, RMSEA = .034, RMSEA 90% CI = [.00, .059]. These CFAs are shown in Supplemental Figure 1 with the results for ages 8-10 shown before the slash and ages 11-16 shown after the slash. For this figure, as well as for Supplemental Figure 2, numbers occluding single-headed arrows are standardized regression coefficients, numbers on double-headed pathways are correlation coefficients, and numbers above the individual measures (shown in rectangles) are squared multiple correlation coefficients that show how much of the variance in each measure is captured by its respective latent variable. Short single-headed arrows pointing to the individual measures represent the residual variance in each measure and can be calculated by subtracting the square multiple correlation coefficients from one.

In line with recommendations from Tabachnick and Fidell (2007) and Byrne (2001), we tested a series of models that placed increasing equality constraints across the two age groups. The results of these analyses found that the four-factor cognitive models were invariant across the two age groups: Model 1 (fully unconstrained), $\chi^2\ (56)= 66.12$, $p = .167$, CFI = .993, RMSEA = .016, RMSEA 90% CI = [.00, .030]; Model 2 (factor weights and intercepts constrained), $\Delta\ \chi^2(16)$ compared to Model 1 = 6.930, $p = .975$; Model 3 (fully constrained), $\chi^2(93) = 95.441$, $p = .411$, CFI = .998, RMSEA = .006, RMSEA 90% CI = [.00, .022], $\Delta\ \chi^2(21)$ compared to Model 2 = 22.396, $p = .377$. Therefore, we concluded that our working memory, inhibition, processing speed, and naming speed constructs were developmentally stable and, for all additional analyses, the two age groups were combined together.

The final step for the cognitive CFA analyses was to test whether the four-factor model was the best representation of our data, or if a more parsimonious model would work as well. We tested all permutations of two- and three-factor models, as well as a model one-factor model that had the cognitive measures load onto one general factor. In the interest of space, only a subset of the models tested are shown: three-factor (working memory and inhibition combined, naming speed, processing speed), $\Delta\chi^2(3)$ compared to four-factor = 28.737, $p < .001$; three-factor (processing speed and naming speed combined, working memory, inhibition), $\Delta\chi^2(3)$ compared to four-factor = 76.731, $p < .001$; two-factor (processing speed and naming speed combined, working memory and inhibition combined), $\Delta\chi^2(4)$ compared to four-factor = 102.299, $p < .001$; one-factor, $\Delta\chi^2(6)$ compared to four-factor = 272.979, $p < .001$. The four-factor model that allowed for four separate cognitive constructs was the best fit to our data.

Reading CFA. The hypothesized three-factor reading model (listening comprehension, reading comprehension, and word reading) fit both age groups well: ages 8-10, $\chi^2(29) = 65.89$, $p < .01$, CFI = .982, RMSEA = .059, RMSEA 90% CI = [.040, .078]; ages 11-16, $\chi^2(29) = 43.08$, $p = .045$, CFI = .990, RMSEA = .040, RMSEA 90% CI = [.01, .06]. Significant improvement in model fit was found when the residuals for the two QRI tasks were allowed to correlate (ages 8-10: $\Delta\chi^2(1) = 23.53$, $p < .01$; ages 11-16: $\Delta\chi^2(1) = 19.06$, $p < .01$) and when the Woodcock passage comprehension and the PIAT reading comprehension tests cross-loaded onto both the reading comprehension and word reading latent variables (ages 8-10: $\Delta\chi^2(2) = 40.19$, $p < .01$; ages 11-16: $\Delta\chi^2(2) = 35.39$, $p < .01$). The CFA results are shown in Figure 2 with ages 8-10 shown before the slash and ages 11-16 shown after the slash.

The need for cross-loadings was discussed at length in Christopher et al. (2012), but, briefly, they indicate that the Woodcock passage comprehension and PIAT reading comprehension tests use relatively short passages that rely as much on word reading ability as comprehension ability (see also Keenan et al., 2008). This is most true for the younger age group (see Supplemental Figure 2).

We next tested for age invariance in the reading CFA. Unlike the cognitive CFA, there was significant change in fit between the unconstrained and fully constrained model: Model 1 (fully unconstrained), $\chi^2(58) = 108.97$, $p < .01$, CFI = .985, RMSEA = .036, RMSEA 90% CI = [.025, .046]; Model 2 (factor weights and intercepts constrained), $\Delta\chi^2(19)$ compared to Model 1 = 23.19, $p = .229$; Model 3 (weights, intercepts, and variances constrained), $\Delta\chi^2(13)$ compared to Model 2 = 28.986, $p = .01$; Model 4 (fully constrained), $\chi^2(94) = 185.39$, $p < .01$, CFI = .973, RMSEA = .038, RMSEA 90% CI = [.030, .046], $\Delta\chi^2(36)$ compared to Model 1 = 76.42, $p = .001$. The largest age difference appeared to be for the correlation between reading comprehension and word reading. Ages 8-10 had a correlation of $r = .72$ while ages 11-16 had a correlation of $r = .58$. It is important to note, however, that the fully constrained model still had good fit to the data. Because of this, as well as to maximize power for the behavioral genetic analyses, we use a fully constrained reading model for all additional analyses.

Finally, we tested whether the three-factor model with separate word reading, listening comprehension, and reading comprehension factors was the best model. Given the very high correlation between listening comprehension and reading comprehension ($r = .94$), there was evidence to suggest that a two-factor model with the two comprehension latent variables collapsed might be as good, or better, than the three-factor model. The results of these comparisons follow: two-factor (reading and listening comprehension combined, word reading), $\Delta\chi^2(2)$ compared to three-factor = 8.69, $p = .013$; two-factor (word reading and reading comprehension combined, listening comprehension), $\Delta \chi^2(4)$ compared to three-factor = 333.80, $p < .01$; one-factor, $\Delta \chi^2(5)$ compared to three-factor = 405.33, $p < .01$.

Combining listening and reading comprehension into one factor significantly decreased model fit ($p = .013$). This contrasts with Christopher et al. (2012) where the two comprehension factors could be collapsed into one factor. The difference is likely due to the larger sample (n = 676 compared to n = 483) and because the present model is fully constrained across the two age groups (in Christopher et al. the two QRI tasks were left to freely vary). While the high correlation ($r = .94$) suggest that modeling the comprehension tasks on one latent variable may be justified, they are kept separate for all analyses. This allows us to directly test whether listening and reading comprehension have different genetic and environmental relations with the cognitive factors.

Supplemental References

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281-300. doi: 10.1080/10888430802132279

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn and Bacon.

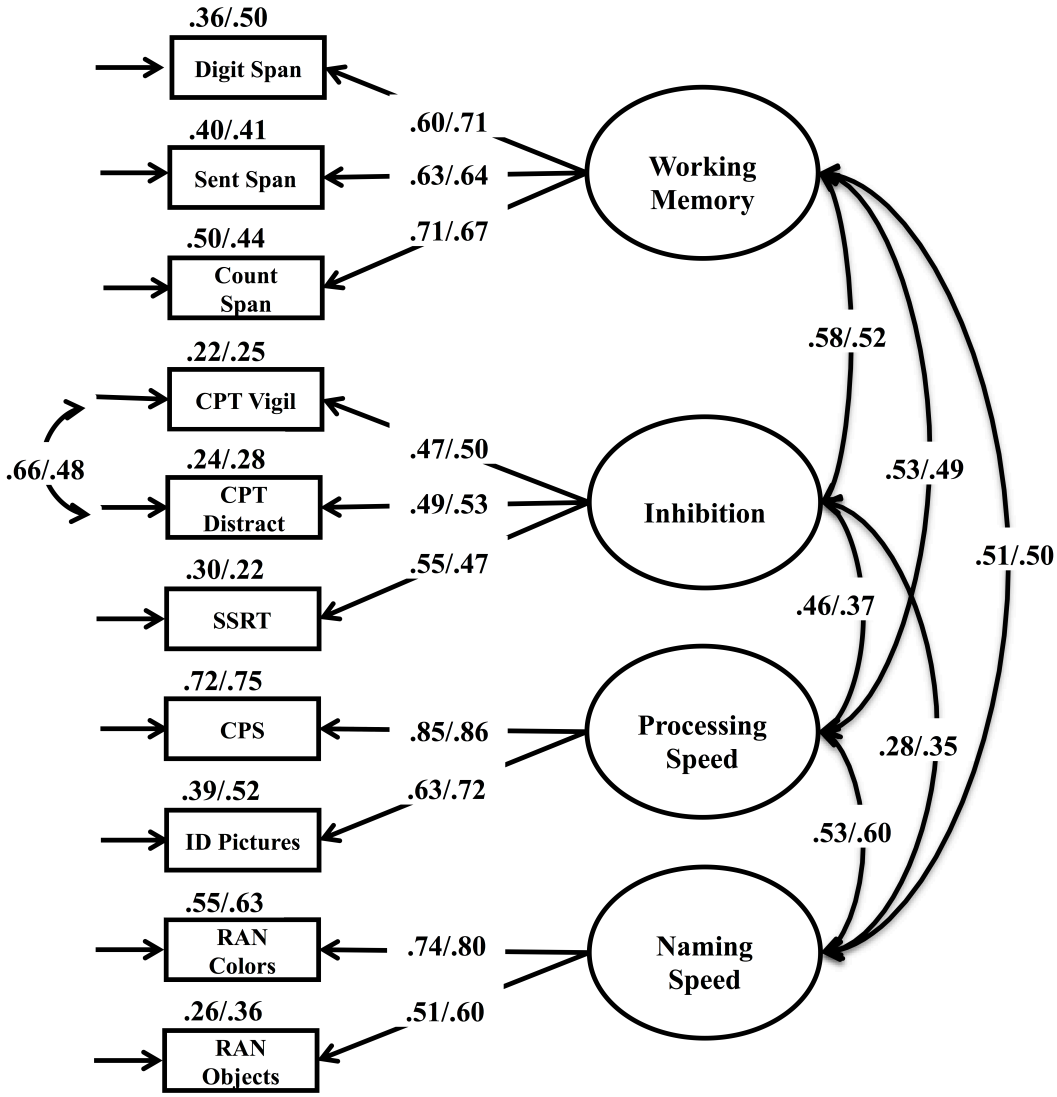
Supplemental Table 1.
Phenotypic Correlations.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Digit Span	-	.41	.42	.20	.21	.15	.28	.12	.21	.11	.23	.24	.23	.29	.30	.43	.42	.47	.42	.38
2. Sent Span	.45	-	.44	.14	.20	.17	.30	.20	.22	.17	.34	.25	.28	.34	.35	.38	.41	.39	.34	.31
3. Count Span	.46	.45	-	.24	.26	.21	.34	.18	.31	.16	.24	.24	.23	.30	.29	.39	.40	.35	.39	.30
4. CPT Vigilance	.21	.18	.15	-	.75	.26	.14	.12	*.09*	*.04*	.22	.20	.21	.22	.23	.25	.21	.25	.15	.16
5. CPT Distract	.23	.17	.17	.62	-	.28	.16	.11	.12	*.00*	.22	.23	.25	.21	.25	.25	.24	.25	.17	.16
6. SSRT	.17	.12	.19	.23	.24	-	.26	.19	.11	.13	*.06*	*.10*	.11	*.10*	.14	.17	.15	.15	*.10*	.11
7. CPS	.37	.28	.28	.15	.14	.16	-	.53	.30	.24	.29	.23	.18	.31	.24	.44	.43	.51	.45	.53
8. ID Pictures	.19	.17	.19	.16	.20	.12	.62	-	.29	.22	.16	.16	*.10*	.24	.17	.21	.25	.28	.19	.32
9. RAN Color	.32	.24	.26	.16	*.11*	.12	.40	.36	-	.38	.18	.21	.17	.25	.19	.34	.27	.38	.27	.24
10. RAN Object	.15	.19	.20	*.07*	*.10*	.17	.29	.35	.48	-	.21	.16	*.09*	.18	*.10*	.21	.18	.23	.16	.15
11. WJOC	.29	.34	.19	*.11*	.17	*.10*	.26	.29	.17	.24	-	.43	.48	.53	.40	.53	.51	.49	.42	.36
12. QRI Listening	.23	.32	.23	.16	.20	.15	.31	.20	.12	.21	.45	-	.38	.44	.52	.37	.33	.31	.27	.19
13. Barnes KNOW-IT	.15	.20	*.08*	*-.03*	*-.01*	*.05*	*.06*	*.09*	*.04*	*.03*	.32	.35	-	.40	.42	.37	.34	.28	.24	.18
14. GORT	.24	.32	.18	.15	.24	*.02*	.20	.17	*.06*	.15	.46	.43	.29	-	.44	.56	.49	.50	.40	.39
15. QRI Reading	.18	.29	.18	*.11*	*.09*	*.05*	.15	.16	*.11*	.16	.43	.58	.28	.47	-	.54	.47	.41	.45	.38
16. WJPC	.36	.38	.29	*.10*	.15	*.10*	.33	.26	.21	*.11*	.58	.50	.31	.51	.45	-	.68	.73	.64	.59
17. PIAT Compre.	.40	.39	.27	.20	.23	.13	.37	.26	.18	.13	.52	.49	.27	.50	.40	.55	-	.69	.66	.64
18. Word Recog.	.43	.34	.33	.17	.17	.17	.55	.25	.29	.12	.36	.41	.14	.35	.29	.55	.58	-	.80	.70
19. PIAT Recog.	.46	.39	.34	.22	.21	.14	.53	.27	.27	.19	.39	.43	.17	.40	.30	.51	.66	.78	-	.72
20. PIAT Spelling	.37	.29	.28	*.11*	.17	.18	.56	.31	.25	.17	.33	.37	.15	.36	.25	.45	.53	.65	.68	-

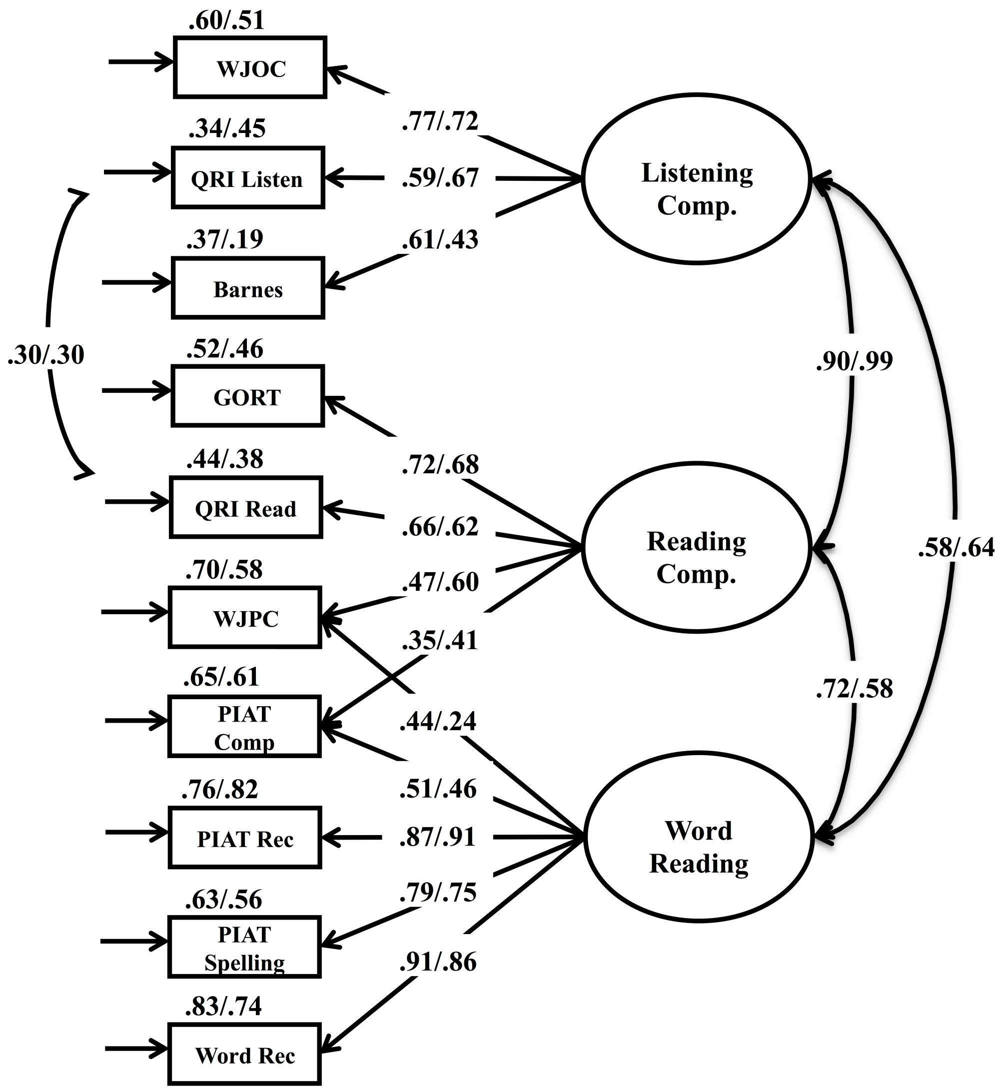
Note: One twin from each pair selected at random; ages 11-16 are shown below diagonal and ages 8-10 are shown above diagonal; *italics* equal not significant at $p < .05$, correlations greater than .15 are significant at $p < .01$.

Supplemental Table 2.
Twin Correlations by Sex

	Monozygotic (MZ) Correlations		**Dizygotic (DZ) Correlations**		
Latent Variable	**Female**	**Male**	**Female**	**Male**	**Opposite Sex**
Working Memory	.87	1.00	.59	.52	.52
Inhibition	1.00	1.00	.96	.16	.23
Processing Speed	.86	.93	.47	.58	.48
Naming Speed	.93	.90	.32	.43	.37
Listening Comprehension	1.00	1.00	.46	.60	.63
Reading Comprehension	1.00	.98	.64	.56	.58
Word Reading	.96	.98	.52	.51	.55



Supplemental Figure 1. Cognitive CFA for Both Age Groups (Ages 8-10/Ages 11-16). The amount of variance in the manifest variables explained by the respective latent variables is shown above each manifest variable.



Supplemental Figure 2. Reading Ability CFA for Both Age Groups (Ages 8-10/Ages 11-16). The amount of variance in the manifest variables explained by the respective latent variables is shown above each manifest variable.