Supplemental Material S4. Analysis of dimensionality for children with identified weakness in one or more areas of language.

As noted in the main article for this study, the full sample of children included a typically developing sample with average scores on standardized measures of language that were slightly below the normative mean for the measures, and a broad range of language abilities were represented in the sample—from two standard deviations below the normative mean to two standard deviations above the normative mean. It is possible that the dimensionality of language identified is affected by the generally average level of language skills exhibited by these children. To determine if more, fewer, or different dimensions would be identified with children who had substantial risk for language delay and impairment, secondary analyses were conducted on a sample of children selected on the basis of their low scores on the language measures.

Children were identified for these secondary analyses if they scored more than a standard deviation below the normative mean (a) on two or more measures of vocabulary, (b) on two or more measures of syntax, or (c) on one or more measures of listening comprehension. Of the identified children, 56% met the selection criterion in just one area (i.e., either vocabulary, syntax, or listening comprehension), 28% met the selection criterion in two areas (e.g., vocabulary and syntax), and 16% met the selection criterion in all three areas (i.e., vocabulary, syntax, and listening comprehension). A greater proportion of children from the full sample in the younger grades (e.g., 48% and 43% of preschool and kindergarten samples, respectively) were identified than were children in the older grades (e.g., 36% and 20% of fourth- and fifth-grade samples, respectively). Descriptive statistics for the children who met selection criterion in each grade are reported below in Table S4.1. As seen in the table, the selected sample was similar to the overall sample in terms of age and sex, but there were more black/African American children in the selected sample than in the full sample. In general, average standard scores for the standardized language measures within grade were half a standard deviation or more lower than the average within-grade scores for the full sample.

Because of the reduced sample size of the selected sample of children, confirmatory factor analyses (CFA) were conducted with the grades combined into younger (i.e., preschool through 2nd grade) and older (i.e., third- through fifth-grade) groups. As with analyses of the full sample, there were 18 theoretically plausible models of the dimensionality of language, ranging from a single-factor model (i.e., Language) to a six-factor model (i.e., Receptive Vocabulary, Expressive Vocabulary, Depth of Vocabulary Knowledge, Receptive Syntax, Expressive Syntax, Listening Comprehension). Results for the younger-grade group are shown below in Table S4.2. Of the 18 models, five yielded allowable solutions. Models with unallowable solutions included correlations between two or more factors that were ≥ 1.0. As with the full sample, the two-factor model with separate Vocabulary and Syntax factors provided the best fit to the data. This two-factor model fit the data significantly better than did the one-factor model ($p < .001$), and none of the other models provided a better fit to the data than did their comparison models. For the younger-grade group, the correlation between the Vocabulary and Syntax factor was .86.

Results for the older-grade group are shown below in Table S4.3. Of the 18 models, five yielded allowable solutions. Models with unallowable solutions included correlations between two or more factors that were ≥ 1.0. As with the full sample, the two-factor model with separate Vocabulary and Syntax factors provided the best fit to the data. This two-factor model fit the data significantly better than did the one-factor model ($p < .001$). One of the four-factor models—the model with separate Expressive Vocabulary, Receptive Vocabulary, Depth of

Vocabulary, and Syntax factors—had a significantly better fit to the data than did the two-factor, Vocabulary and Syntax model (p = .011); however, this difference was not significant after Benjamini–Hochberg correction. Additionally, the Akaike information criterion (AIC) values for Model 2 and Model 9 were very similar, supporting a conclusion of equivalent fit, with preference given to the more parsimonious two-factor model. None of the other models provided a better fit to the data than did their comparison models. For the older-grade group, the correlation between the Vocabulary and Syntax factor was .88.

Given the similarity of results between those obtained from the full sample of children and children selected because of significant weaknesses in one or more areas of language, it seems unlikely that the results of the primary analyses were a function of the population of children recruited for the study. Although the children in the sample selected for these secondary analyses were not children with an identified language impairment, they represented the subset of the full sample with substantial risk for language delay and impairment as well as the negative academic sequellae associated with below-average language skills. Additional work is required to address fully the question of the dimensionality of language in samples of children with identified language disabilities; however, the results of the secondary analyses suggest that the two-factor (i.e., Vocabulary and Syntax) dimensionality of language is likely to be consistent across a range of language abilities.

Table S4.1. Descriptive statistics for a representative sample of language measures by grade for children with identified weakness in one or more areas of language.

Construct	**Pre-K**	**K**	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**
n per Grade	179	100	100	62	118	83	46
Percent female	44	35	53	58	49	44	56
Percent white	44	70	70	61	55	54	39
Percent black	26	25	28	35	38	34	50
Measure	**Mean (*SD*)**	**Mean (*SD*)**	**Mean (*SD*)**	**Mean (*SD*)**	**Mean (*SD*)**	**Mean (*SD*)**	**Mean (*SD*)**
Age (months)	58.98 (5.12)	71.09 (6.58)	86.16 (6.67)	99.73 (8.53)	111.92 (7.55)	122.17 (6.35)	132.17 (8.02)
ROWPVT	86.21 (14.16)	88.20 (11.75)	95.93 (11.60)	93.85 (13/40)	95.47 (9.31)	97.68 (10.37)	99.14 (9.59)
EOWPVT	85.45 (12.80)	86.93 (15.81)	93.18 (11.19)	88.71 (13.00)	93.03 (11.66)	94.13 (10.92)	94.49 (12.52)
CASL-A	87.99 (14.44)	83.15 (14.07)	87.44 (12.11)	91.79 (13.45)	93.22 (12.12)	93.05 (9.98)	94.44 (10.07)
CASL-G	90.20 (8.24)	89.85 (12.91)	93.38 (13.35)	90.92 (10.86)	89.89 (12.66)	90.71 (12.11)	85.94 (10.69)
CASL-SC	81.57 (11.63)	80.72 (13.59	81.03 (13.87)	80.17 (12.72)	80.44 (15.09)	84.21 (15.14)	83.94 (13.46)
OWLS-LC	80.87 (10.57)	89.36 (10.59)	83.46 (10.81)	84.34 (12.29)	85.37 (12.20)	86.73 (8.17)	79.56 (8.01)
WJ-OC	90.50 (9.86)	90.01 (11.26)	94.83 (9.70)	90.86 (12.16)	93.01 (10.11)	94.94 (9.49)	92.27 (12.54)

Note. Standard score (normative mean = 100; *SD* = 15); CASL = Comprehensive Assessment of Spoken Language; ROWPVT = Receptive One-Word Picture Vocabulary Test; EOWPVT = Expressive One-Word Picture Vocabulary Test; CASL-A = CASL Antonyms subtest; CASL-G = CASL Grammaticality subtest; CASL-SC = CASL Syntax Construction subtest; OWLS-LC = Oral and Written Language Scales Listening Comprehension subtest; WJ-OC = Woodcock–Johnson III Oral Comprehension subtest.

Table S4.2. Model fit statistics for models of language dimensionality in preschool through second-grade group for children with identified weakness in one or more areas of language.

Model	Y-Bχ^2	*df*	CFI	TLI	RMSEA [90% CI]	AIC	BIC	Models Compared	Corrected $\Delta\chi^2$
Model 7	308.70	148	.91	.90	.05 [.04–.06]	16306.60	16556.03	2 v 7	5.00^{ns}
Model 6	310.28	148	.91	.90	.05 [.04–.06]	16307.53	16556.96	2 v 6	3.42^{ns}
Model 3	343.05	150	.90	.88	.05 [.05–.06]	16338.24	16579.50	1 v 3	1.97^{ns}
Model 2	313.70	150	.91	.90	.05 [.04–.06]	16308.37	16549.62	1 v 2	31.32***
Model 1	345.02	151	.90	.88	.05 [.05–.06]	16337.59	16574.75		

Note. Dimensions for each model are described in Table 2 in the main article. Only models with allowable solutions are shown. Y-Bχ^2 = Yuan–Bentler χ^2; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; CI = confidence interval; AIC = Akaike information criterion; BIC = Bayesian information criterion. All model χ^2s were significant at $p < .001$.
$^{ns}p > .05$. $^{***}p < .001$.

Table S4.3. Model fit statistics for models of language dimensionality in third- through fifth-grade group for children with identified weakness in one or more areas of language.

Model	Y-Bχ^2	*df*	CFI	TLI	RMSEA [90% CI]	AIC	BIC	Models Compared	Corrected $\Delta\chi^2$
Model 9	303.99	163	.92	.91	.06 [.05–.07]	9241.60	9476.72	2 v 9	14.88*
Model 8	318.40	166	.91	.90	.06 [.05–.07]	9248.78	9473.38	2 v 8	1.43^{ns}
Model 3	351.57	168	.89	.88	.07 [.06–.08]	9276.86	9494.44	1 v 3	2.88^{ns}
Model 2	319.50	168	.91	.90	.06 [.05–.07]	9245.91	9463.49	1 v 2	41.75***
Model 1	354.31	169	.89	.88	.07 [.06–.08]	9277.29	9491.36		

Note. Dimensions for each model are described in Table 2 in the main article. Only models with allowable solutions are shown. Y-Bχ^2 = Yuan–Bentler χ^2; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; CI = confidence interval; AIC = Akaike information criterion; BIC = Bayesian information criterion. All model χ^2s were significant at $p < .001$.
$^{ns}p > .05$. $^{*}p < .05$. $^{***}p < .001$.