Supplemental Material S5. Analysis of dimensionality of language skills using alternative classification of measures.

One of the anonymous reviewers for this manuscript suggested a different classification of measures than the one used in the main manuscript, and, in this section, we evaluate the fit of models using this alternative classification. Specifically, the reviewer suggested that (a) the Clinical Evaluation of Language Fundamentals (CELF) Word Class Receptive and CELF Word Class Expressive should both be considered measures of depth of vocabulary, (b) all of the measures classified as receptive syntax measures in the main article—except the CELF Sentence Structure subtest—should be classified as metalinguistic tasks, (c) all of the measures classified as expressive syntax measures in the main article should be combined with the CELF Sentence Structure subtest as syntax measures, and (d) all of the measures identified as indicators of the Listening Comprehension factor should be considered syntax measures.

We evaluated eight models in each grade (i.e., preschool through fifth grade) using this classification, which comprises up to six possible dimensions.

1. The *receptive vocabulary* dimension was indexed by two measures: Receptive One-Word Picture Vocabulary Test (ROWPVT) and Test of Language Development–Primary: Fifth Edition (TOLD-PV).
2. The *expressive vocabulary* dimension was indexed by two measures: Expressive One-Word Picture Vocabulary Test–Third Edition (EOWPVT-3) and CELF Expressive Vocabulary subtest (CELF-EV).
3. The *depth of vocabulary* dimension was indexed by five measures: Comprehensive Assessment of Spoken Language Antonyms subtest (CASL-A), CELF Word Definitions subtest (CELF-WD; grades 3–5 only), Test of Language Development–Primary: Fourth Edition (TOLD-P) and Test of Language Development–Intermediate: Fourth Edition (TOLD-I) Relational Vocabulary subtests (TOLD-RV), CELF Word Classes–Receptive I and II subtests (CELF-WCR), and CELF Word Classes–Expressive I and II subtests (CELF-WCE).
4. The *syntax dimension* was indexed by five measures: CASL Syntax Construction subtest (CASL-SC), CELF Formulated Sentences subtest (CELF-FS), Spoken Morphological Awareness Task (SMA), TOLD-P Morphological Completion subtest (TOLD-MCmpl; grades pre-K–2) or TOLD-I Sentence Combining subtest (TOLD-SC; grades 3–5), and CELF Sentence Structure subtest (CELF-SS).
5. The *metalinguistic* dimension was indexed by four measures: CASL Grammaticality subtest (CASL-G), CELF-SS, TOLD-P Syntactic Understanding subtest (TOLD-SU; grades pre-K–2] or TOLD-I Morphological Comprehension subtest (TOLD-MComp; grades 3–5), and Morphological Syntax Awareness task (MSA).
6. The *listening comprehension* dimension was indexed by three measures: CELF Concepts and Following Directions subtest (CELF-CFD), Oral and Written Language Scales Listening Comprehension subtest (OWLS-LC), and Woodcock–Johnson III Tests of Achievement Oral Comprehension subtest (WJ-OC).

Because the alternative classification did not involve individual measures originally classified as listening comprehension measures, we left these measures as a possible distinct dimension. Table S5.1 below shows the eight models that were evaluated.

Results

Preschool to Second-Grade Groups

In the Pre-K-to-Grade 2 sample, Model 6 yielded a solution that was not allowable (i.e., correlations between factors => 1.0). Similar to the model tests reported in the article, the different vocabulary factors were correlated => 1.0. Once all of these indicators were included on the same "Vocabulary" factor (Model 5b), the Syntax, Listening Comprehension, and Metalinguistic-Task factors were correlated at => 1.0 (for preschool, first grade, and second grade), indicating that the best fitting model was the two-factor Vocabulary/Syntax model identified as the best fitting model in the main article. In the kindergarten group, the Listening Comprehension and Metalinguistic-Task factors were correlated at => 1.0. A three-factor model with separate Vocabulary and Syntax factors, as well as a factor that combined the Listening Comprehension and Metalinguistic-Task factors, yielded an allowable solution. However, this model did not provide a better fit to the data than did the one-factor model, corrected-$\Delta\chi^2 = 7.85$, $df = 3$, $p = .05$ (not statistically significant following Benjamini–Hochberg correction), suggesting that the alternative classification of measures to dimensions did not partition the common variance between measures as well as the classification used in the main analyses did, which resulted in the two-factor (Vocabulary/Syntax) model providing a better fit to the data than did the one-factor model.

Table S5.1. Description of structure of models evaluated using alternative classification of measures to dimensions.

Model	Factors	Constructs included in each factor
1	1	(E-Vocab + R-Vocab + Depth-Vocab + Syntax + L-Comp + Metalinguistic)
2	2	(E-Vocab + R-Vocab + Depth-Vocab); (Syntax + L-Comp + Metalinguistic)
3a	3	(E-Vocab + R-Vocab); Depth; (Syntax + L-Comp + Metalinguistic)
3b	3	E-Vocab; (R-Vocab + Depth); (Syntax + L-Comp + Metalinguistic)
3c	3	(E-Vocab + Depth); R-Vocab; (Syntax + L-Comp + Metalinguistic)
4	4	E-Vocab; R-Vocab; Depth-Vocab; (Syntax + L-Comp + Metalinguistic)
5a	5	E-Vocab; R-Vocab; Depth-Vocab; (Syntax + L-Comp); Metalinguistic
5b	5	(E-Vocab + R-Vocab + Depth-Vocab); Syntax; L-Comp; Metalinguistic
6	6	E-Vocab; R-Vocab; Depth-Vocab; Syntax; L-Comp; Metalinguistic

Note. Constructs enclosed in parentheses and connected by a plus (+) sign were combined to form a factor; constructs separated by a semicolon (;) were represented as one factor. E-Vocab = expressive vocabulary; R-Vocab = receptive vocabulary; Depth-Vocab = depth of vocabulary; Syntax = syntax; L-Comp = listening comprehension; Meta-Linguistic = metalinguistic tasks. Model 5a was the model evaluated in the third- to fifth-grade group, and Model 5b was the model evaluated in the preschool to second-grade group, based on results of Model 6.

Third- to Fifth-Grade Groups

Results of models and model comparisons for the third- to fifth-grade groups are shown in Tables S5.2, S5.3, and S5.4, below. As with the results from the analyses described in the main article, the two-factor (Vocabulary/Syntax) model fit the data significantly better than did the one-factor model. None of the three-factor models provided a better fit to the data than did the two-factor model for the third- or fourth-grade groups (following Benjamini–Hochberg correction for multiple comparisons). As can be seen in Tables S5.2 and S5.3, fit indices were similar for all two- through five-factor models in these two groups. In these grade groups, the six-factor model produced a solution in which the Listening Comprehension and Syntax factors were correlated at 1.0 (i.e., Model 5). For the fifth-grade group, in contrast, all of the three-factor models fit the data better than did the two-factor model (see Table S5.4), and the four-factor model fit the data better than did the best fitting three-factor model (all following Benjamini-Hochberg correction for multiple comparisons). Neither the five- nor the six-factor model resulted in significantly than did the best fitting three-factor model (all following Benjamini-Hochberg correction for multiple comparisons).

Table S5.2. Fit statistics for models of language dimensionality for third-grade group with alternative classification of measures to language dimensions.

Model	Y-Bχ^2	*df*	CFI	TLI	RMSEA [90% CI]	AIC	BIC	Models Compared	Corrected $\Delta\chi^2$
Model 5a	242.38	159	.96	.96	.04 [.03–.05]	10302.44	10562.76	5 v 4	8.44^{ns}
Model 4	250.82	163	.96	.95	.04 [.03–.05]	10305.53	10551.18	4 v 3a	7.27^{ns}
Model 3a	258.09	166	.96	.95	.04 [.03–.05]	10307.03	10541.69	3a v 2	7.10^{*b}
Model 3b	259.04	166	.96	.95	.04 [.03–.05]	10308.10	10542.75	3b v 2	6.15^{*b}
Model 3c	259.35	166	.96	.95	.04 [.03–.05]	10308.17	10542.82	3c v 2	5.84^{ns}
Model 2	265.19	168	.96	.95	.05 [.03–.06]	10310.40	10537.72	2 v 1	45.17^{***a}
Model 1	310.36	169	.94	.93	.05 [.04–.06]	10354.54	10578.19		

Note. Dimensions for each model are described in Table S5.1. 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.

[a]$\Delta\chi^2$ significant following Benjamini–Hochberg correction for multiple comparisons. [b]$\Delta\chi^2$ not significant following Benjamini–Hochberg correction for multiple comparisons. All model χ^2s were significant at $p < .001$.

[ns]$p > .05$. [*]$p < .05$. [**]$p < .01$. [***]$p < .001$ (for uncorrected p values).

Table S5.3. Fit statistics for models of language dimensionality for fourth-grade group with alternative classification of measures to language dimensions.

Model	Y-Bχ^2	*df*	CFI	TLI	RMSEA [90% CI]	AIC	BIC	Models Compared	Corrected $\Delta\chi^2$
Model 5a	269.58	159	.94	.93	.06 [.04–.07]	8656.69	8900.48	5 v 4	4.42^{ns}
Model 4	272.87	163	.94	.93	.05 [.04–.06]	8655.17	8885.23	4 v 3c	8.16^{*b}
Model 3a	281.03	166	.94	.94	.06 [.04–.06]	8657.21	8876.96	3a v 2	7.77^{*b}
Model 3b	283.60	166	.94	.93	.06 [.04–.07]	8659.67	8879.43	3b v 2	4.14^{ns}
Model 3c	280.67	166	.94	.93	.06 [.04–.07]	8656.68	8876.44	3c v 2	7.77^{*b}
Model 2	288.44	168	.94	.93	.06 [.05–.07]	8660.33	8873.22	2 v 1	34.31^{***a}
Model 1	328.23	169	.92	.91	.06 [.05–.07]	8699.01	8908.47		

Note. Dimensions for each model are described in Table S5.1. 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.

[a]$\Delta\chi^2$ significant following Benjamini–Hochberg correction for multiple comparisons. [b]$\Delta\chi^2$ not significant following Benjamini–Hochberg correction for multiple comparisons. All model χ^2s were significant at $p < .001$.

[ns]$p > .05$. [*]$p < .05$. [**]$p < .01$. [***]$p < .001$ (for uncorrected p values).

Neither the five- nor the six-factor model resulted in significantly improved model fit. Correlations between factors from the four-factor model for the fifth-grade group are shown in Table S5.5, below. As with the two-factor model in the analyses reported in the main article, correlations between these factors were high and indicated that the dimensions of language represented by these factors shared between 61%–76% of their variance.

Summary

For six of the seven grade groups, the alternative classification of measures to language dimensions did not result in the identification of a different best-fitting model than was reported in the main manuscript. For preschool, kindergarten, first grade, second grade, third grade, and fourth grade, the two-factor model with separate Vocabulary and Syntax factors provided the best fit to the data. Results for Grade 5 indicated that a four-factor model in which the three dimensions of vocabulary were separate factors provided the best fit using the alternative classification of measures. Notably, however, the measures were not selected to represent these dimensions, and as a consequence, both the Expressive Vocabulary and Receptive Vocabulary factors were under-identified (i.e., each factor comprised of only two measures) and the missing-by-design assessment strategy was not constructed to account for this grouping of measures (i.e., the patterns of missingness for measures were determined for groups of measures within the

Table S5.4. Fit statistics for models of language dimensionality for fifth-grade group with alternative classification of measures to language dimensions.

Model	Y-Bχ^2	*df*	CFI	TLI	RMSEA [90% CI]	AIC	BIC	Models Compared	Corrected $\Delta\chi^2$
Model 6	247.53	154	.95	.94	.05 [.04–.06]	8709.49	8972.09	6 v 5	4.38^{ns}
Model 5a	251.68	159	.95	.94	.05 [.04–.06]	8704.04	8949.37	5 v 4	6.91^{ns}
Model 4	258.71	163	.95	.94	.05 [.04–.06]	8704.39	8935.89	4 v 3a	12.20^{**a}
Model 3a	270.50	166	.94	.94	.05 [.04–.06]	8709.90	8931.04	3a v 2	22.67^{***a}
Model 3b	285.80	166	.94	.93	.06 [.04–.07]	8725.31	8946.45	3b v 2	10.83^{**a}
Model 3c	279.81	166	.94	.93	.05 [.04–.07]	8719.28	8940.42	3c v 2	14.98^{***a}
Model 2	294.79	168	.93	.92	.06 [.05–.07]	8730.24	8944.47	2 v 1	52.35^{***a}
Model 1	338.60	169	.91	.90	.07 [.06–.08]	8771.57	8982.35		

Note. Dimensions for each model are described in Table S5.1. 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.

[a]$\Delta\chi^2$ significant following Benjamini-Hochberg correction for multiple comparisons. All model χ^2s were significant at $p < .001$.

[ns]$p > .05$. [*]$p < .05$. [**]$p < .01$. [***]$p < .001$ (for uncorrected p values).

Table S5.5. Correlations between factors for fifth-grade group using alternative classification of measures to language dimensions.

	Receptive Vocabulary	Expressive Vocabulary	Depth of Vocabulary	Syntax
Receptive Vocabulary	—			
Expressive Vocabulary	.86	—		
Depth of Vocabulary	.84	.83	—	
Syntax	.81	.78	.87	—

Note. All correlations significant at $p < .001$.

classification scheme used in the main article). In the case of factor under-identification, measure-specific variance (as opposed to construct-specific variance) cannot be ruled out as a reason that the two measures within each factor had more in common with each other than with the other vocabulary measures. The potential effect of a different distribution of missingness across measures within each dimension is unknown. Regardless, these results suggest that it is possible that dimensions of vocabulary differentiate at around the fifth grade; however, additional study using sufficient measures for each dimension is needed. Moreover, the pattern of correlations between factors indicates that, although potentially separable, the dimensions represented have more variance in common than is distinctly associated with one dimension or the other.