12-10-2018

Longitudinal Structure of Expressive and Receptive Language Among Young African American Children: An Examination of the Preschool Language Scale-5

Leslie E. Hodges
Georgia State University

LONGITUDINAL STRUCTURE OF EXPRESSIVE AND RECEPTIVE LANGUAGE AMONG YOUNG AFRICAN AMERICAN CHILDREN: AN EXAMINATION OF THE PRESCHOOL LANGUAGE SCALE-5

by

LESLIE E. HODGES

Under the Direction of Lee Branum-Martin, PhD

ABSTRACT

Existing language assessments and theories of language development assume a clear division into receptive and expressive processes. However, measurement studies provide only mixed support for this structure of language–some studies support the division of language into dual processes of expressive and receptive language while other studies conclude that language is a single process. The Preschool Language Scale – 5 (PLS-5, Zimmerman, Steiner, & Pond, 2011) is a commonly administered assessment used for diagnostic and research purposes and thus, the psychometric properties should be well-established. The PLS-5 is one of only a few assessments to cover the age range of birth to seven years old, meaning it is useful for research on longitudinal language development in addition to the recommended clinical use. This study uses PLS-5 data collected from 2014 to 2017 at urban childcare centers serving primarily low SES African American children to address two research questions: 1) to what extent does a confirmatory factor model support the division of language into receptive and expressive language components? 2) what does this structure of language say about how students grow in preschool? The results of the longitudinal confirmatory factor analyses suggest that language is a general construct rather than divided by modality, and thus the total language score on the PLS-5 may be preferable to interpreting the individual subscale scores of Expressive Communication and Auditory Comprehension.

INDEX WORDS: Language structure, growth, Expressive language, Receptive language, PLS-5, PPVT, Get Ready to Read, Multi-level modeling framework, SEM

LONGITUDINAL STRUCTURE OF EXPRESSIVE AND RECEPTIVE LANGUAGE AMONG YOUNG AFRICAN AMERICAN CHILDREN: AN EXAMINATION OF THE PRESCHOOL LANGUAGE SCALES-5

by

LESLIE E. HODGES

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2018

LONGITUDINAL STRUCTURE OF EXPRESSIVE AND RECEPTIVE LANGUAGE AMONG YOUNG AFRICAN AMERICAN CHILDREN: AN EXAMINATION OF THE PRESCHOOL LANGUAGE SCALES-5

by

LESLIE E. HODGES

Committee Chair: Lee Branum-Martin

Committee: Julie Washington

Gary Bingham

Elizabeth Tighe

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2018

ACKNOWLEDGEMENTS

Dr. Lee Branum-Martin has my eternal gratitude for being an outstanding mentor. He took me on for my last two years as a graduate student and suffered through a full six years of questions packed into that time. I am thankful for my other committee members, Dr. Elizabeth Tighe, Dr. Julie Washington, and Dr. Gary Bingham, for their support, encouragement, and time. I appreciate the support of the Urban Child Studies Center and their willingness to share their data and answer my questions. Finally, I am thankful to my husband for his support, and to my daughter for her patience and tolerance of being the last kid at daycare for so many days.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS IV

LIST OF TABLES VIII

LIST OF FIGURES IX

1 INTRODUCTION 1

1.1 At-risk Population 1

1.2 Language: Two Modality Constructs or One General Construct? 3

1.2.1 Arguments for two constructs. 4

1.2.2 Arguments for one general construct. 6

1.3 Developmental Differences in Language Structure 15

1.4 Longitudinal Structural Equation Modeling 17

1.5 Current Study 18

1.5.1 Preschool Language Scale-Fifth Edition 19

1.5.2 Other Measures 20

1.5.3 Research questions 21

2 METHOD 22

2.1 Participants 23

2.2 Measures 24

2.2.1 Preschool Language Scales-5th Edition. 24

2.2.2 Peabody Picture Vocabulary Test-Fourth Edition (PPVT). 25

2.2.3 Get Ready to Read Screening Tool-Revised (GRTR-R). 26

2.3 Data Analysis Plan 28

3 RESULTS 31

3.1 Descriptive statistics 32

3.2 Models 36

3.2.1 Trivariate Growth Model 37

3.2.2 Method Factor Model 39

3.2.3 Second-Order Language Factor Model 39

3.3 Full Model 43

4 DISCUSSION 47

4.1 Validation of the PLS-5 47

4.2 Language Modality and the Structure of Language 48

4.3 Language Growth 49

4.4 Relation to reading readiness 51

4.5 Limitations 51

4.6 Future Directions 53

4.7 Conclusions 54

REFERENCES 55

APPENDICES 64

Appendix A 64

Appendix B 65

LIST OF TABLES

Table 1. Completed sessions at each data collection period. 23

Table 2. Observed Correlations and Descriptive Statistics 32

Table 3. Correlations Among Latent Variables in the Trivariate Growth Model 39

Table 4. Model fit 41

Table 5. Parameter estimates for the Unified Language Model 43

Table 6. Parameter estimates for Unified Language Model with GRTR-R and covariates 46

Table 7. Parameter Estimates for the Trivariate Growth Model 64

Table 8. Parameter estimates for the Method Factor Model 65

LIST OF FIGURES

Figure 1. Conceptual representation of the language structure models. 29

Figure 2. Longitudinal plots of the Preschool Language Scale-5 34

Figure 3. Longitudinal plots of the PPVT-IV and GRTR-R 35

Figure 4. Trivariate Growth Model 38

Figure 5. Unified Language Model with general language intercept and slope 40

Figure 6. Full model with covariates. 45

1 INTRODUCTION

Considering the strong associations found between early oral language skill and emergent literacy skills in kindergarten, and between emergent literacy skills in kindergarten and reading success into high school (Arnold & Doctoroff, 2003; Shanahan & Lonigan, 2010), identification and intervention of language difficulties at the earliest possible point is key to academic success. For this reason, it is important to assess language early using high-quality (valid and reliable) assessments. The Preschool Language Scales-Fifth edition (PLS-5; Zimmerman, Steiner, & Pond, 2011) is a play-based language assessment normed from birth to seven years. The assessment includes two subscales, one each for receptive and expressive modalities of language, and the manual includes instructions to use the difference between the subscale scores to identify children as having a language disorder or needing language support. However, measurement studies of language assessments do not support the division of language into expressive and receptive language (e.g., Tomblin & Zhang, 2006). To date, there are no studies that have factor-analyzed longitudinal assessments of expressive and receptive language for children younger than pre-kindergarten. Longitudinal studies allow the simultaneous examination of language structure and trajectories of language growth. The current study will address the question of language structure, language change over time, and its relation to reading readiness using a longitudinal confirmatory factor analysis approach on a sample of low SES African-American preschoolers (1 to 5 years old) who completed the PLS-5, the Peabody Picture Vocabulary Test – IV (Dunn & Dunn, 2007)—a receptive vocabulary measure, and the Get Ready to Read-Revised (Whitehurst & Lonigan, 2001) —a reading readiness screener.

1.1 At-risk Population

The majority of the recommended uses of the PLS-5 in the technical manual focus on determining if a language delay or disorder is present, however, there are compelling reasons to use the assessment for children at higher risk of struggling academically. There are known academic achievement gaps between children from families with low socioeconomic status (SES) and those from middle/high SES families (Arnold & Doctoroff, 2003; García, 2015). These achievement gaps include delays in acquiring language skills, delays in early literacy skills such as letter recognition and phonological awareness, and a high prevalence of reading difficulties (Arnold, 2003). Young children from low SES families also have more limited oral language skills (García, 2015), which place them at particular risk of later reading disability (Catts, Fey, Tomblin, & Zhang, 2002). Explanations for the language skill difference include an array of home and wider environmental factors from maternal education (Rindermann & Baumeister, 2015), to parental emphasis on education, shared reading experiences, and even early teachers' attitudes (Arnold & Doctoroff, 2003). Early oral language skills—which encompass expressive and receptive language modalities and can be considered as part of the same construct as listening comprehension (LARRC, 2017)—have been established as important and reliable predictors of later reading ability (Shanahan & Lonigan, 2010). Identifying low language performance in pre-kindergarten (or younger) allows early oral language intervention, and subsequently increased chances of reading success years down the road.

In a longitudinal study of children followed from three to ten years old, structural equation modeling showed that language abilities at four years of age had a direct effect on reading comprehension at 10 years of age (Dietrich, Assel, Swank, Smith & Landry, 2006). While children with impaired oral language skills when entering kindergarten are at a much higher risk of reading failure in later grades, the subset of children who showed improvement of language skills from kindergarten to early elementary grades were closer to the reading comprehension level of children with typical oral language than children who did not improve in oral language (Catts, Fey, Tomblin, & Zhang, 2002). Since the PLS-5 assessment is designed for children who are prelinguistic up to early elementary school years, it provides an opportunity to examine early language growth, and with the Get Ready to Read-Revised measure, the relation to pre-reading skills. In order to model the relation of language and reading readiness, the structure of language needs to be better understood. The PLS-5 divides language into expressive language (spoken modality) and receptive language (listening modality), but language might better be represented as a general construct as it pertains to modality (other possible divisions of language into grammar, vocabulary, or syntax constructs cannot be addressed with the design of the PLS-5).

1.2 Language: Two Modality Constructs or One General Construct?

Are speaking and understanding two separate processes, or a single process manifested in two modalities? Bloom and Lahey define language as “a code whereby ideas about the world are represented through a conventional system of arbitrary symbols for communication” (1978, p. 4) and as “knowledge of the integration of content/form/use…such knowledge underlies the behaviors of speaking and understanding” (1978, p. 22). Though these statements suggest a structure of language as one cohesive construct used in two modalities (e.g., speaking and understanding), the authors later suggest that expressive and receptive language are separate, but related processes. However, measurement evidence of school-aged children indicates the opposite: language is language, regardless of whether it is spoken or heard (e.g., Tomblin & Zhang, 2006; Lonigan & Milburn, 2017). It is unclear, though, if this unity across modalities remains true across development. Perhaps for young language learners, receptive and expressive language are two constructs that consolidate with increasing competence. The evidence to support each of these the two language structure options will be reviewed in turn: 1) that language is divided by modality into expressive and receptive language versus 2) that language is a unitary construct in regards to modality.

1.2.1 Arguments for two constructs.

In the field of psycholinguistics, expressive and receptive language are claimed to be separable constructs because each relies on different processes. For example, only expressive language requires the motor process of articulation while receptive language involves cognitively parsing the sound stream into meaningful units. Levelt (1993) describes an architecture of spoken language with two streams of processes working in parallel—one set of processes for producing speech and one set for understanding speech—with a 'conceptualizer' linking the processes. Straight wrote extensively on this process-based division, referring to it as "processualism" (1976). Processualism was proposed in response to theories of language that are neutral to input/output (speaking and listening), meaning that language is language regardless of the modality used: structuralism is an example of a modality-neutral language theory. According to de Saussure, language can be described as a set of linked sounds and meanings and the grammar of a language describes the pattern of how the set is used (Straight, 1976). Straight and others stressed that the differences between producing speech and comprehending speech are too great to ignore, making a modality-neutral theory of language, such as de Saussure's, inadequate.

Existing language assessments also suggest that language is meaningfully divisible into separate expressive and receptive language constructs. The design of a language measure is based on the creators' assumptions about the underlying structure of language. Several commonly used language assessments, including the Preschool Language Scale-5 (Zimmerman, Steiner, & Pond, 2011), have been designed to include subscales of receptive and expressive language—resting on an assumption that these are meaningful and separable constructs. Alternatively, the Peabody Picture Vocabulary Test-Fourth Edition uses only the receptive modality and only assesses vocabulary knowledge. The PLS-5 includes items assessing a variety of language aspects including morphology, syntax, and pragmatic language use in each subscale, but scores are not provided for these—whereas in other tests there are subscales of morphology, syntax, or vocabulary but not expressive and receptive language. If a test-creator does not support a theory of expressive and receptive language, there would be no reason to design an assessment in this fashion and provide instructions on calculating subscale scores and comparing the difference between the sub-scale scores.

Separate language disorder diagnoses exist for a delay in expressive and/or receptive language, but a single diagnosis of language disorder may be more appropriate (see Leonard, 2009). Though the DSM-5 condensed these diagnoses into one general language disorder in 2013, the ICD-10—the manual of all diagnostic codes from the World Health Organization—still recognizes both Expressive Language disorder and Receptive Language disorder. There may be practical value in the field of speech language pathology for these separate diagnoses, but that is beyond the scope and focus of this project.

The traditional argument of children’s language development is that language comprehension (receptive language) develops before language production (expressive language) (Benedict, 1979). Thus, particularly for children, comprehension is expected to exceed production—if comprehension develops first in the unfolding of language mastery, it would logically follow that comprehension levels would be higher than production. Parent report diaries of English-speaking middle-class children from 10 to 24 months of age indicated word comprehension started at a younger age and increased more rapidly than word production such that for the 8 children studied, receptive vocabulary was 5-10 times larger than productive vocabulary during their second year of life (Benedict, 1979). Fenson et al. (1994) reported a fan-shaped pattern of growth for 16- to 30- month-old children; children at higher percentile ranks demonstrated increases in both expressive and receptive language, but children at lower percentile ranks demonstrated minimal increases in expressive language despite increases in receptive language, which might suggest that the developmental trajectories of expressive and receptive language are different for younger children with lower ability levels. A comparison of parent-reported receptive and expressive language (based on one question each from UNICEF's Multiple Indicator Cluster Survey) for over 100,000 international children, aged two to ten years old, indicated higher levels of comprehension than production with only a small to moderate correlation between the two (r = .21; Bornstein & Hendricks, 2012), indicating only minimal overlap of expressive and receptive language.

1.2.2 Arguments for one general construct.

The apparent gap in the language comprehension and production abilities of children has led to divergent conclusions: either these are different constructs of expressive and receptive language (e.g., Bloom & Lahey, 1972) or children have one underlying grammar that is still developing in how strictly items are represented (Smolensky, 1996). There are several areas of research that either undermine the receptive expressive language divide or provide support for for language as a single general ability. First, there are clear violations of the temporal expectations of language comprehension developing first followed by language production later. For example, the first words children indicate comprehending are not the same words children first produce, nor is there a developmental progression on the easiness scale of comprehension to later production (Bloom & Lahey, 1972; Hendriks & Koster, 2010). Second, the majority of factor analytic studies addressing language structure support the more parsimonious model of a single construct of general language over separate constructs of expressive and receptive language. These two bodies of evidence supporting language as a single process will be reviewed in turn.

1.2.2.1 Temporal relation between receptive and expressive language.

If comprehension develops first and leads to production then language comprehension at earlier ages should predict later language production, since having higher levels of receptive language earlier should lead to higher levels of expressive vocabulary later. A vocabulary comprehension and production study of 26 children assessed from 8 months old until 21 months old provides evidence against this (Bauer, Goldfield, & Reznick, 2002). Parents completed the MacArthur Communicative Development Inventories: Words and Gestures (CDI; a parent-report checklist of words their child understands and says) monthly from child age 8 months until 11 months, and at child age of 21 months parents completed the CDI: Words & Sentences (parent report of child's spoken words). While the number of words children comprehended greatly exceeded the number of words children produced, receptive vocabulary scores from any earlier timepoint were not a significant predictor of later expressive vocabulary at 21 months. Children were identified as having a fast or slow trajectory for each expressive and receptive vocabulary: some children had slow growth in both, some children had fast growth in both, and some children had slow growth in one and fast growth in the other. Interestingly, those identified as having a fast trajectory in receptive vocabulary but slow trajectory for expressive vocabulary did not have higher expressive vocabulary scores at 21 months compared to children with slow growth trajectories in both, and both the slow expressive vocabulary trajectory groups had significantly lower expressive vocabulary than children with a fast trajectory for expressive vocabulary (regardless of those children's receptive vocabulary trajectory; Bauer, Goldfield, & Reznick, 2002). Furthermore, the oft-referenced childhood vocabulary *production* spurt (e.g., Goldfield & Reznick, 1990) may temporally coincide with a vocabulary *comprehension* spurt, rather than following. Reznick & Goldfield (1992) used a visual preference task to better estimate English-speaking children's receptive vocabulary development from 14 to 22 months of age rather than relying on parent reports. Comparing the bi-monthly lab measures of receptive vocabulary to parent report diaries of expressive vocabulary, the researchers showed that not only is there a spurt in children's receptive vocabulary but it occurs in the same two-month period as a spurt in expressive vocabulary (if such a spurt was seen in the child; Reznick & Goldfield, 1992).

There are also cases where language production seems to exceed comprehension, which would be unexpected from the traditional view of language development. Focusing on a particular English grammar construction, researchers demonstrated a case where young children's comprehension exceeds production as well as a case where production exceeds comprehension. Van Hout, Harrigan, & de Villiers (2010) conducted studies of comprehension and production of definite and indefinite noun phrases with English speaking children ranging in age from 3-5 years old. In study one, children overused 'the' in definite noun phrases (the car vs. a car) in production but interpreted definite noun phrases correctly, providing evidence that in this case receptive language abilities exceed expressive language. However, the second study with indefinite noun phrases showed that children produced these without error but were overly liberal with their interpretations in the comprehension task, showing that for indefinite noun phrases expressive abilities exceed receptive abilities.

There are several significant limitations of the research on the timing of children's language production and comprehension discussed here. The sample sizes are typically small, and limited measures of language assessment are administered. The research conducted by Van Hout et al. (2010) focused on such a specific grammatical construction that generalization to other features of language is limited. The studies reviewed in the next section utilized much larger sample sizes, multiple measures of language, and factor analytic statistical modeling approaches, which together eliminate the limitations noted here.

1.2.2.2 Statistical models of language structure.

Factor analysis is an approach researchers have used to directly assess questions about the structure of language. Using batteries of language assessments to test such things as knowledge of vocabulary, grammar, or ability to follow directions using both expressive and receptive modalities, the fit of models with one general language factor can be compared to other possible language structures. In studies including children from pre-kindergarten up to eighth grade with typical development and diagnosed language delays, the conclusions support a single general language factor over a two-factor structure of expressive and receptive language, at least for English-speaking children.

To illustrate that language assessments—purported to independently test specific language skills such as syntactic knowledge or expressive language—actually all test general language, Sommers, Erdige, and Peterson (1978) factor analyzed a battery of language tests administered to children diagnosed with developmental language disorders. The first of their two studies included measures of receptive language—the PPVT, the Test for Auditory Comprehension of Language (Carrow, 1973), and Northwestern Screening Syntax Task-Receptive subtest (Lee, 1969)— and expressive language—the Northwestern Screening Syntax Task-Expressive subtest administered to 122 children ranging from 3 to 9 years old. This study, with mostly receptive language measures, indicated the possibility of two factors—receptive and expressive language. In a second study, the measures were more heavily focused on expressive language: the Manyuk Sentence Repetition Task, Developmental Sentence Analysis (Lee, 1974) and Mean Length of Utterance (McCarthy, 1930) assessed expressive language and 3 subtests of the Test for Auditory Comprehension of Language assessed receptive language. In this study, a single factor (language) explained most of the variance. Sommers et al. (1978) concluded the first study's solution was likely due to shared variance as a result of the task modality, since there was an unequal weighting of receptive to expressive language measures.

More recent factor analyses of the structure of language in school-aged children also indicate that language is not structured into receptive and expressive components. Tomblin and Zhang (2006) followed a group of children, both typically developing and children diagnosed with language impairment from kindergarten (n = 1929) to second (n = 604), fourth (n = 570), and eighth grade (n = 527). The students completed a battery of language tests at the word level and sentence level, with differences in the measures administered at each grade. For kindergarten four subtests of the Test of Language Development-2:Primary (TOLD-2:P, Newcomer & Hammill, 1988) were used—for receptive language: the Picture Identification and Grammatical Understanding and Oral Vocabulary and for expressive language: the Grammatical Completion and Sentence Imitation. For second grade receptive language, the PPVT-R (Dunn & Dunn, 1981) and the Concepts and Directions and Sentence Structure subtests of the Clinical Evaluation of Language Fundamentals-III (CELF-III; Semel, Wiig, & Secord, 1995) were used, and for expressive language, the Expressive Vocabulary subtest of the Comprehensive Receptive and Expressive Vocabulary Test (CREVT, Wallace & Hammill, 1994) and the Word Structure and Recalling Sentences subtests of the CELF-III were used. For fourth and eighth grade receptive language, the PPVT-R and the Concepts and Directions subtest of the CELF-III were used and for expressive language, the Recalling Sentences and Formulating Sentences (only fourth grade) subtests of the CELF-III and the Expressive Vocabulary subtest of the CREVT were used. Using a confirmatory factor analysis by grade, the researchers showed that a two-factor structure of receptive language and expressive language did not fit the data better than a single factor structure (Tomblin & Zhang, 2006).

A principal components analysis—including six subtests of the Clinical Evaluation of Language Fundamentals test (CELF-4), the Children's Test of Non Word Repetition (Gathercole & Baddeley, 1996) and the PPVT-III administered to typically developing Australian children aged 5 and 7—indicated one overall language component explained 57% of the variance in the measures used (Matov, Mensah, Cook, & Reilly, 2018). The expressive and receptive language tasks did not load onto the same subsequent components—instead, after the first component, the remaining components were task-specific. In addition, the subtests were so highly related that only two subtests, the Concepts and Following Directions and Recalling Sentences, were necessary to identify children with low language ability almost as well as the much longer full test battery.

In a cross-sectional study of predominantly English-speaking middle class children from pre-kindergarten to third grade, multiple measures of grammar, vocabulary, and listening comprehension were administered including the PPVT, the Expressive Vocabulary Test-2 and subtests of the CELF-4 (LARRC, 2016). The researchers used confirmatory factor analyses to compare the fit of two-model of language and listening comprehension to a single factor model—though the fit indices were slightly better for the two-factor model, the correlation between the latent factors ranged from .87-.91 supporting the selection of the single factor model instead (LARRC, 2016).

Lonigan and Milburn (2017) used confirmatory factor analyses for a sample of 1,895 pre-kindergarten through fifth grade students—with a sizable portion of their sample coming from schools with the majority of children eligible for free and reduced lunch (an indicator of lower socio-economic status which is correlated with greater risk of educational difficulty)—who completed a battery of language assessments. The measures included subtests of the TOLD-I, CELF-4, CASL, Woodcock-Johnson III, and Oral and Written Language Scales, as well as the Receptive One-Word Picture Vocabulary Test, and the Expressive One-Word Picture Vocabulary Test. Models were tested separately by grade—for kindergarten through fifth grades, a single language factor fit the data better than two factors of expressive and receptive language (though a two-factor model of syntax and vocabulary was the best-fitting model). However, for pre-kindergarten, the two-factor model of expressive and receptive language had a better fit than the single factor model, indicating possible developmental differences in language structure. The difference in the model fit for language structure for children in prekindergarten versus older children warrants further exploration.

1.2.2.3 General language + expressive and receptive language.

As indicated in the psycholinguistics field, there are differences between listening and speaking, but these are not incompatible with a structure of one underlying general language ability. A statistical approach that would allow explicitly modeling the task differences inherent between speaking and listening yet also include a general construct of language, is the bifactor model. In this type of model, performance on an indicator is explained by more than one factor—one factor explains the variance common across the indicators, and specific factors explain the remaining variance among subsets of indicators (Little, 2013; Maul, 2013). Several cross-sectional studies have tested the structure of language as a bifactor model for children, with a general factor of language and different specific factors. The majority of these prior studies included school-aged children rather than the much younger sample used in the current study, leaving the question of language structure insufficiently examined in preschool children.

In a cross-sectional study of 1792 fourth through tenth grade students, Foorman and colleagues used confirmatory factor analysis to test different models, including a bifactor model, of oral language—which encompassed both speaking and understanding (Foorman, Koon, Petscher, Mitchell, & Truckenmiller, 2015). Oral language measures using a receptive modality included the PPVT-4, the Study Aid and Reading Assistant Vocabulary and Morphological Awareness subtests. Measures using the expressive modality included the CELF-4 Recalling Sentences subtest and the CASL Grammaticality subtest. The researchers found that a bi-factor model of general language ability with specific factors of vocabulary and syntax best fit the data for each of the seven grade levels — however, these specific factors could have been named expressive and receptive language since all of the vocabulary tasks were in the receptive modality and the syntax tasks were all in the expressive modality.

In two studies with elementary students, Kieffer, Petscher, Proctor, & Silverman (2016) showed a bifactor model was the best fitting model for a set of language assessments, though neither study tested a model with specific factors for expressive and receptive language, the test battery of the second study included both receptive (R) and expressive (E) tasks. In the first study 144 sixth grade students, primarily Latino and economically disadvantaged native English speakers, completed researcher-created receptive (multiple-choice) measures of morphological awareness and vocabulary knowledge to assess language (no expressive measures). Using an item-level analysis for the two measures, the bifactor model with a general language factor and specific factors for each test was shown to be the best-fitting model.

In their second study, Kieffer et al. (2016) used a wider age range of students and more language measures, incorporating standardized measures of language, and using both modalities. In a cross-sectional design, 311 third through fifth graders completed the Woodcock-Muñoz Language Survey Picture Vocabulary subtest (E), CELF Formulated Sentences (E) and Word classes (R) subtests, the Extract the Base test of morphological awareness (E), as well as a researcher-designed measure of morphological awareness (R). Measures intended to test morphological awareness and vocabulary included both expressive and receptive items. The students were racially diverse, with about half of the sample reporting as Latino and a third as Black, the majority of the sample economically disadvantaged, and approximately a quarter of the sample were non-native English speakers. Again, a bi-factor model—with a general language factor and specific factors for vocabulary, morphological awareness, and syntax—best fit the data, though a model with specific factors of expressive and receptive language was not compared to the chosen model. Consequently, though the Kieffer et al. (2016) study provides support for a general factor of language, it provides limited insight into the question of expressive and receptive language constructs as first-order factors or even specific factors.

In a study with younger children, 286 four-year-old English-Spanish bilingual children completed expressive and receptive language measures in Spanish, including subtests of the CELF-P2, the Expressive One Word Picture Vocabulary test, Receptive One Word Picture Vocabulary, several comprehension monitoring tasks, inference tasks, and a language impairment screener (LARRC, 2015a). The researchers tested a two-factor model of expressive and receptive language: this model did not improve fit over a unidimensional model, and the latent factors were perfectly correlated, meaning there was no discrimination between the constructs. Modality differences were not probed further, and the researchers concluded that a bifactor model with a general language factor and two specific factors of ‘word knowledge’ (including measures of vocabulary and comprehension monitoring) and ‘integrative language knowledge’ (including measures of grammar and inference tasks) best fit the data (LARRC, 2015a).

Together, the bifactor models of language structure for school-age children do not support the differentiation of language by modality, but instead support a modality-independent language structure. Though different models are more appropriate with the longitudinal data used in the present study, the bifactor modeling approach used with cross-sectional data of children at different ages provides strong support for a general language factor. Since multiple underlying growth trajectories are possible from the same set of data, longitudinal datasets are necessary to reveal patterns of growth in language (McArdle 1988; 2009).

1.3 Developmental Differences in Language Structure

For students in kindergarten and above, research consistently indicates that a one-factor model of language fits better than two factors of receptive and expressive language, but for pre-kindergarten the evidence is less conclusive—perhaps receptive and expressive language are initially separate constructs that later consolidate into one. A study of native English speaking pre-kindergarten students, primarily low SES and African-American, completed receptive and expressive tests of grammar (CELF-Preschool-2 sentence structure and word structure subtests), semantics (one word picture vocabulary tasks), and phonology (tests of articulation judgment and production; Anthony, Davis, Williams, & Anthony, 2014). Using multi-level exploratory factor analysis to account for classroom nesting, the researchers showed these tests loaded onto one general language factor (with separate factors for articulation and speech perception)—the expressive and receptive modalities were not sufficient to differentiate language skills at either the student level or the classroom level.

The Language and Reading Research Consortium (LARRC; 2015b) reached a similar conclusion of a single general language factor for pre-kindergartners. A cross-sectional study of children from pre-kindergarten to third grade completed measures of vocabulary, grammar, and discourse—though only the vocabulary measures (PPVT, Expressive Vocabulary Test, and the CELF-4 word Classes Receptive and Word Classes Expressive subtests) included both receptive and expressive modalities. The LAARC tested models of language dimensionality in terms of vocabulary, grammar, and discourse by grade: the authors did settle on a one-factor solution for the 420 pre-kindergartners, but was also evidence to support a (highly correlated) two-factor solution of grammar/vocabulary and discourse (2015b). Unfortunately, the lack of discourse and grammar measures administered in the receptive modality means this study can only support the general unidimensionality of language in pre-kindergartners without specifically addressing the structure of expressive and receptive language.

Klem, Gustavfsson, and Hagtvet (2015) used latent variable modeling to provide evidence of validity for a Norwegian preschool language screener (Language4) administered to 600 children. The Language4 screener was established as unidimensional—however, nearly all of the items were assessed in the expressive language modality. To test concurrent validity, the researchers created a criterion latent language factor composed of standardized language assessments including the British Picture Vocabulary Scale II, Weschler Preschool and Primary scale of Intelligence-III Picture Naming subtest, the Illinois Test of Psycholinguistic Abilities Grammatic Closure subtest, and the Test for Reception of Grammar-2, such that there were two receptive and two expressive measures; all four measures loaded onto a single factor of language with strong factor loadings (.69 - .78). The Language4 screener's general language factor explained >60% of the variance in the criterion general language factor, providing evidence for concurrent validity of the screener, but also pointing to the strong relations among language measures and providing support for language as a unidimensional construct.

Lonigan and Milburn (2017) found that for pre-kindergarten, but not for kindergarten through third grade, a two-factor structure of expressive and receptive language was a better fit to the data than a single factor of language. Several other studies analyzed language structure and stability over time in children ranging from 5 months to 14 years of age, and found a one-factor solution to best fit the data; however these studies included measures that were composites of expressive and receptive subscale scores and so could not address the question of expressive and receptive language constructs (Bornstein & Putnick, 2012; Bornstein, Hahn, Putnick, & Suwalksy, 2014; Bornstein, Hahn, & Putnick, 2016; Putnick, Bornstein, Eryigit-Madzwamuse, & Wolke, 2017). It is possible that the structure of language is different for children in pre-kindergarten and younger than for older children, but there are no studies that have factor-analyzed longitudinal assessments of expressive and receptive language for children younger than pre-kindergarten. Longitudinal studies would allow the examination of children's language growth as well as language structure.

1.4 Longitudinal Structural Equation Modeling

The majority of the language structure studies reviewed thus far have been cross-sectional in design. While cross-sectional data is useful in addressing questions of language structure, it does not allow for probing into the nature of change within individuals or differences in the nature of change across individuals. As illustrated by McArdle (1988; 2009), beneath the same set of cross-sectional scores, there are a multitude of possible patterns that can be revealed with longitudinal data and appropriate analyses. Using longitudinal data, structural modeling approaches can simultaneously address questions of language structure and change over time (McArdle, 1988; McArdle, 2009; Little, 2013). Little, Card, Preacher & McConnell (2009) discuss design considerations when collecting longitudinal data in order to accurately address the research questions. However, the current study utilizes an archival dataset, meaning that decisions such as cohort design and data collection intervals have already been made.

Structural equation modeling (SEM) is well-suited for modeling longitudinal data for several reasons including, but not limited to, the following: latent constructs are free of measurement error since they are modeled by multiple indicators, SEM can deal with violations of normality in several ways, and correlated errors between observed variables can be explicitly modeled rather than treated as a violated assumption (Little, Bovaird, & Slegers, 2006). Within SEM, there are a variety of options of how to represent the relation of previous scores on measures to later scores, depending on the research questions, theory, and the available data. The research questions of the current study are focused on the structure of language in addition to the change in language over time. These questions are interdependent and require a modeling approach that allows structure and change to be examined simultaneously: growth curve modeling (Bollen & Curran, 2006; Little, 2013; McArdle, 2009). The change in a measure over time, the relations between measures—including the relations of change—can be incorporated into a single model.

1.5 Current Study

The current study asks whether the structure of language is better represented as two constructs of expressive and receptive language or as a single construct of language and how that construct (or constructs) may change over time. To test this question, an assessment with expressive and receptive sub-test is analyzed (Preschool Language Scale – Fifth Edition), along with another measure of receptive language (Peabody Picture Vocabulary Test – Fourth Edition) to test for differential relations among receptive and expressive measures. The sample used here is a group considered at-risk of reading difficulty due to lower family income/education, so the relation of the language variables to a measure of reading readiness, the Get Ready to Read-Revised test, is also modeled. In addition, person-level covariates of special needs services and gender are considered. Special needs services included the child receiving an individual education plan or family services plan at any point during the course of participation. Gender was chosen as a person-level covariate because of the conflicting conclusions of research into sex differences in language skill; some research has concluded that girls have an advantage in language around ages 1-6 (Bornstein, Hahn, & Haynes, 2004) though a meta-analysis of gender differences argued that the difference is so small as to be negligible (Hyde & Linn, 1988). Considering that this study primarily involves children from low SES families, it is also important to note that gender differences were found to be more extreme between low SES girls and boys (Barbu et al., 2015).

1.5.1 Preschool Language Scale-Fifth Edition

The PLS-5 (Zimmerman, Steiner, & Pond, 2011) is built on the assumption of two language constructs: expressive and receptive language, as evidenced by the two-subtest construction of the assessment. The test is normed for children from birth to 7 years old and is widely used for diagnosis and research—published research has cited the various versions of the PLS hundreds of times. Unfortunately, there is limited evidence for the construct validity of the PLS-5 (Miles, Fulbrook, & Mainwaring-Mägi, 2016). To date, there is no factor analysis of the

PLS-5 (Denman et al., 2017) meaning that construct validity questions of underlying theoretical structure or relations to other assessments are unanswered (Whiteley, 1983; Messick 1989; 1995).

The PLS-5 manual reports a rather high correlation of .75 between the Auditory Comprehension (PLS-AC, receptive language) and Expressive Communication (PLS-EC, expressive language) subscales for all ages. This is alarming, as the instructions for the use of the PLS-5 as a screening measure instructs the use of a comparison between the AC and EC subscale scores to assess if a deficit is present in one or both scales. This seems of limited utility since approximately 50% of the norming sample had a higher expressive than receptive subscale score, and the other 50% showed the opposite pattern (Zimmerman, Steiner, & Pond, 2011). The even split of higher score patterns provides another layer of evidence against the proposal that receptive language develops earlier and exceeds expressive language (see discussion above).

1.5.2 Other Measures

The use of an SEM framework for analysis allows other measures to be added into the growth model, meaning that concurrent and structural validity can be evaluated. By including another language measure that only uses a single modality, such as the PPVT which is a receptive vocabulary measure, the relation between the PPVT and each of the PLS-5 subtests can be compared. If language is better represented as separate constructs of receptive and expressive language, a stronger relation between the PPVT and the PLS-AC would be expected compared to the relation between the PPVT and PLS-EC, establishing convergent validity for the receptive measures as relatively distinct from the expressive measure. If language is better represented as a single construct, then the relations of the PPVT to the PLS-5 subtests should be the same.

A reading readiness test can also be incorporated into the model. The Get Ready to Read – Revised Test (GRTR-R; Whitehurst & Lonigan, 2001) is a short screener of pre-literacy skills, appropriate for preschool children. It is reasonably predictive of reading success in second grade. The use of the screener before the introduction of formal reading instruction provides an opportunity for enforcing foundational pre-literacy skills before the child struggles to learn to read. This study provides a connection of early oral language to the GRTR-R.

To summarize, The PLS-5 is a standardized normed language assessment for prelinguistic to early elementary children which provides subscale scores for receptive and expressive language. The PLS-5 instructs decisions of language delay or support services to be determined based on the difference between the subscale scores. However, measurement evidence of studies with children in pre-kindergarten and older supports a modality-neutral structure of language over separate constructs of expressive and receptive language. To date, no factor analysis exists for the PLS-5 to support the structure of the test, nor are there longitudinal studies of the PLS-5.

1.5.3 Research questions

The mixed evidence of language structure for pre-kindergarten students, differing theories on the existence of receptive and expressive language as separable constructs, and the need for construct validation of the PLS-5 particularly for a predominantly typically-developing low-SES African American population all lead to the research questions for this study: a) to what extent is language development better characterized as one process or two processes? b) how do children's language skills co-develop over time? c) how do language skills relate to early reading? Due to the nature of the models tested—which incorporate both structure and growth—research questions a and b will be answered simultaneously rather than sequentially.

2 METHOD

Data for this study come from the Urban Child Studies Center at Georgia State University and were collected as part of an external evaluation of urban childcare centers from the spring of 2014 to the spring of 2017. Assessments were administered in fall—from September to November— and spring—from March to May, for a total of seven collection periods. Since the data were collected as part of a childcare center evaluation, children joined throughout the three years of data collection as they enrolled at the facility. Table 1 shows the data collection time and the number of children who completed a session (their first, second etc.) at that time, as well as the count of the completed measures of interest at each data collection period. For example, the table shows that 21 children completed their seventh wave of observation at the last collection time.

Table 1. *Completed sessions at each data collection period.*

	Data Collection Period						
Completed Session	1	2	3	4	5	6	7
1st	176	68	0	115	11	100	5
2nd		103	68	1	104	8	92
3rd			103	40	14	55	8
4th				40	42	12	51
5th					39	13	12
6th						24	12
7th							21
TOTAL	176	171	171	196	210	212	201
Measure							
PLS-AC	134	125	105	165	164	172	173
PLS-EC	134	125	105	163	163	173	174
PPVT	133	118	101	157	163	158	158
GRTR-R	0	101	101	161	151	162	152

2.1 Participants

Due to the rolling enrollment of data collection, descriptive statistics reported here are for the first session the child completed. 475 children (224 females, M_{age} = 41.9 months, SD = 15.8) completed at least one data collection session. The majority of the children were reported as African American (n = 469), and a much smaller number of children were reported as White (n = 2) or Hispanic/Latino (n = 4). A subset of children (16%, n = 77) were identified as 'needs special services' by having an Individualized Family Service Plan (IFSP; developed for children younger than 3) or an Individualized Education Plan (IEP; for children 3 and older) documented for at least one data collection session. Since no information was available about specific diagnoses (such as a diagnosed language delay) the 'needs special services' variable was used to test if model fit differed by group (those identified as needing special services compared to the other students).

2.2 Measures

Classroom, teacher, and family level assessments were administered as well as student level measures of language, vocabulary, school readiness, math concepts, and executive function. Only measures intended for analysis are described in detail below. The PLS-5 includes both expressive and receptive scales of language, and the PPVT is a receptive measure of language. The relation between the PLS-5 and the PPVT will indicate convergent validity. The GRTR-R is included as a measure of predictive validity.

2.2.1 Preschool Language Scales-5th Edition.

The PLS-5 is a play-based language assessment normed from birth to 7;11. Items incorporate manipulatives, everyday routines, gesture comprehension, and even parent report to assess an array of prelinguistic indicators of language, morphology, syntax, semantics, and even predictors of academic success like knowing numbers and colors. Raw scores range from 0 to 67.

2.2.1.1 PLS-5 reliability

Test-retest reliability coefficients range from .86 to .95. Inter-rater reliability coefficients are high for both subscales (all $\kappa > .95$), and inter-scorer agreement for items requiring subjective scoring were all high as well (> 90% agreement). Internal consistency—as measured by split-half reliability using the Spearman-Brown formula—is reported as .91 for the Auditory Comprehension sub-scale for the entire age range of the norming sample, and .93 for the Expressive Communication sub-scale. The correlations between the PLS-4 and PLS-5 are reported as .80 for each subtest.

2.2.1.2 PLS-5 validity

The correlations between the PLS-5 and the Clinical Evaluation of Language Fundamentals-Fourth Edition were .70 between the receptive language subscales and .82 for the expressive language subscales. Performance differences are reported between those diagnosed with a language disorder, including Receptive Language Disorder, Expressive Language Disorder and Expressive-Receptive Language Disorder and a nonclinical sample. Children diagnosed with Receptive Language disorder performed significantly worse than children without a diagnosis on both the auditory comprehension subscale (AC, d = 1.89) and the expressive communication subscale (EC, d = 1.76), with a larger effect size for the receptive language subscale. Similarly, children diagnosed with Expressive Language disorder performed worse than children without a diagnosis, on both the AC subscale (d = 1.87) and the EC subscale (d = 2.10), with a larger effect size reported for the expressive language subscale. Children diagnosed with Expressive-Receptive Language Disorder also performed worse on AC (d = 1.93) and EC (d = 1.91). The sensitivity of the PLS-5 to diagnose a language disorder—defined as a Total Language score of 1 SD below the mean or 85—is .83, meaning 83% of children with a language disorder diagnosis were identified as such by the PLS-5. The specificity is reported as .80, meaning 80% of children without a language disorder diagnosis were also not identified as disordered by the PLS-5.

2.2.2 Peabody Picture Vocabulary Test-Fourth Edition (PPVT).

The PPVT (Dunn & Dunn, 2007) is used here as an additional receptive language measure. It is a normed measure of receptive vocabulary intended for ages 2;6 to 90 years old and commonly used to assess language. The examinee is asked to point to one of four pictures that matches the word provided by the assessment administrator.   Raw scores range from 0 to 228.

2.2.2.1 PPVT reliability

All reliability and validity values reported here are the weighted means for the entire age range of the norming sample unless otherwise stated. Internal consistency is very high for the PPVT, with split-half reliability =.94 and Cronbach's alpha = .97 for the norming sample. Alternate-form reliability between form A and B is also high, r = .88. Test-retest reliability—which indicates measurement stability—is high as well, r = .92.

2.2.2.2 PPVT validity

Correlations with several other standardized tests of language are provided as evidence of convergent validity. The overall mean correlation with the Expressive Vocabulary Test-second edition is .82 for the norming sample. For the Comprehensive Assessment of Spoken Language (CASL) subtests of Basic Concepts, Antonyms and Sentence Completion for ages 3-5 correlations with the PPVT-IV range from .37 to .54. For the CELF-4 core language, receptive language, and expressive language index scores for children ages 5-8, correlations with the PPVT-IV range from .67 to .73. For fall kindergartners, the Group Reading Assessment and Diagnostic Evaluation (GRADE) total test score and the PPVT-IV correlation = .66. For the PPVT-III and PPVT-IV, r = .84. The PPVT-IV also discriminates between children with and without a language delay: children aged 3-7 diagnosed with a language delay had a mean score almost 1 standard deviation below children without a language delay.

2.2.3 Get Ready to Read Screening Tool-Revised (GRTR-R).

The revised version of the Get Ready to Read screener (Whitehurst & Lonigan, 2001) was administered to children 36 months and older starting at the second data collection session. It is intended for children aged 3;0-5;11—adding two years to the range of the original GRTR—and measures students' literacy skills as precursors to learning to read and write in kindergarten.

Items assess print knowledge (e.g., differentiating print and pictures), emergent writing skills (e.g., printing name), and linguistic awareness (e.g., rhyming) where the child chooses an answer from four picture options per item. The GRTR-R includes 25 items, 19 of which are the same as the original version of the screener. Score ranges are assigned to categories: 0-4 = limited understanding of beginning literacy skills, 5-13 = basic understanding, 14-20 = beyond basic understanding, and 21-25 = solid understanding of beginning literacy skills. The developers tentatively proposed that children with a score of 12-13 on the screener in the Fall of pre-kindergarten are likely to be successful readers in second grade (Whitehurst, 2001).

Since the revised version of the screener and the original screener share most items (19/20 of the original screener and 19/25 of the revised version), the development of the original screener and its relation to other measures is relevant. The original GRTR instrument was developed with a sample of children ranging from 48-59 months, with disproportionate representation of children from low-income families. The sample had a mean score of 9.14 correct out of 20 (*SD* = 4.31), with 68% percent of the sample scoring between 5-13 correct answers. When divided into groups, significant differences between the mean scores of White (*M* = 11.91) and African American children were observed (*M* = 9.03), as well as between children from middle-income households (*M* = 12.52) and children in Head Start (*M* = 8.52), though means for mutually exclusive groups of race and household income level were not provided.

2.2.3.1 GRTR-R reliability

Internal consistency is good, (α =.88), for the entire age range of the norming sample (3 to 5 years old) and was still acceptable when broken down into ethnic groups by age (α > .7) except for 3-year-old Latino children (α =.66). The items are intended to assess print knowledge and phonological awareness, though an exploratory factor analysis only supported a single factor for the screener, meaning that subscores are not appropriate (Lonigan & Wilson, 2008). Raw scores, standard scores, and percentiles are provided for the revised version.

2.2.3.2 GRTR-R validity

Fourteen measures of the Developing Skills Checklist focused on emergent literacy were chosen as a standard measure of comparison with the screener, as these scores were previously shown to be highly predictive of reading success: the correlation with the GRTR was .69 for the development sample. The GRTR also correlates reasonably well with the Peabody Picture Vocabulary Test, $r = .58$ (Whitehurst, 2001).

2.3 Data Analysis Plan

Using a multilevel SEM framework (Kaplan & Elliott, 1997; Mehta & Neale, 2005; Muthen, 1991, 1994), alternate models of language structure using the PLS-EC, PLS-AC, and PPVT were fit and compared. The planned models are listed below:

1. Trivariate Growth Model with three parallel processes
2. Modality Model with expressive and receptive language factors
3. Method Model with PLS and PPVT test factors
4. Unified Language Model with second order latent factors of language level and growth
5. Final model + prediction of reading, with gender and special services covariates

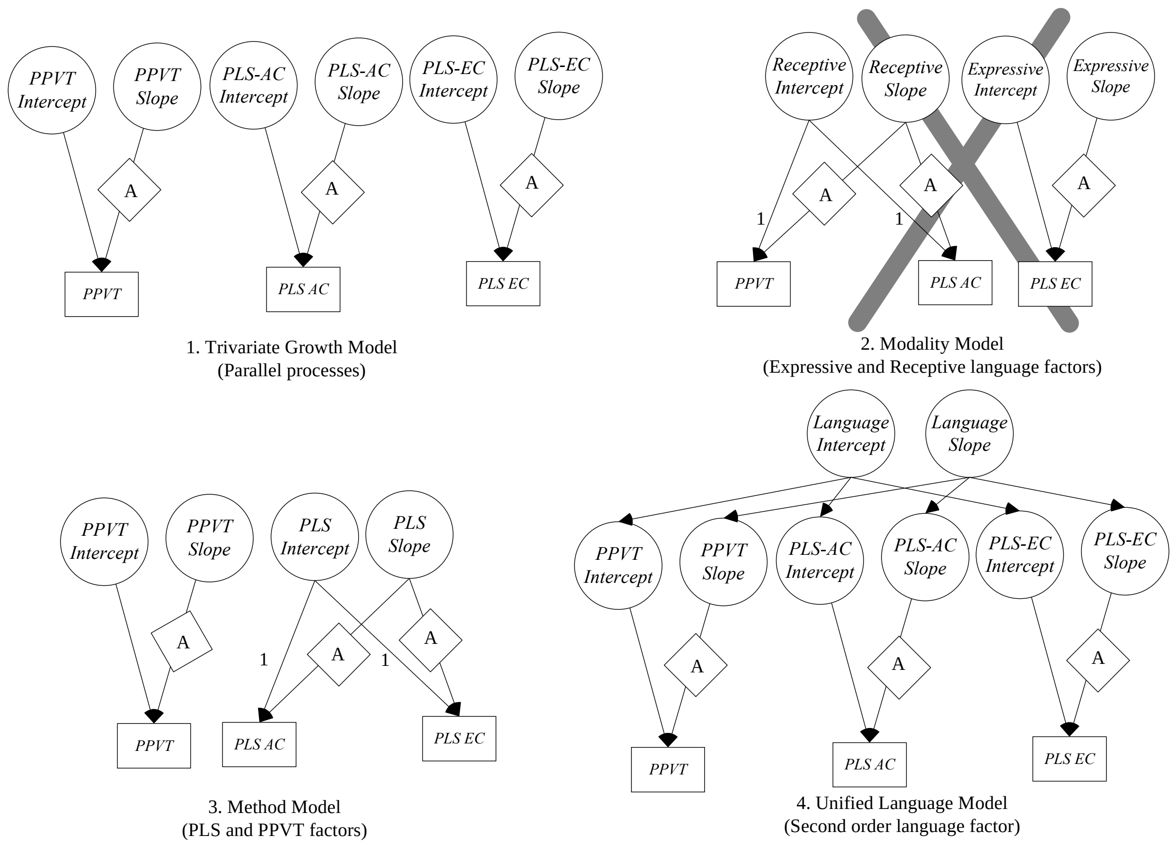

Figure 1. Conceptual representation of the language structure models.
Figure 1 provides a conceptual representation of the structural models tested (1-4). Since these are conceptual models, for purposes of visual clarity all variances, covariances, and mean structure are implied but not shown.

Figure 1 provides a visual representation of the planned structural models (Models 1-4). The Modality Model (Model 2) in Figure 1, where receptive language is measured by both the PLS-AC and the PPVT and expressive language is measured by the PLS-EC, tests the hypothesis that language is better represented by modality. However, model identification is a problem. With only two measures, the factor loadings for the receptive intercept factor would need to be equal, which is not justifiable across the PLS-AC and the PPVT. The PLS-AC uses a variety of task types to assess an array of language and language-related skills, while the PPVT only assesses receptive vocabulary using a picture identification task and the scale range of the PPVT is more than three times the range of the PLS-5. There is little reason to believe these would be perfectly equal indicators of the same latent variable. For this reason, the modality model was not tested (however, the modality question can be addressed in model 4).

The Method Model (Model 3) in Figure 1 tests whether there are method effects, meaning differences exist between the PLS-5 subscales and PPVT as a result of different item types between assessments and shared items on the PLS-5. The Method Model (PLS-5 versus PPVT) is identifiable (unlike the Modality Model). Since the subscales of the PLS come from the same assessment, have similar item types, and are scaled approximately the same, it is justifiable to equate the latent factor loadings. In the Method Model, shown in Model 3 of Figure 1, the PPVT was modeled with its own intercept and slope, and the PLS-AC and the PLS-EC were used to estimate the intercept and slope for the PLS-5 Test.

To test whether language is better represented as a unified structure, second order latent intercept and growth constructs were added to the model. The Unified Language model is shown as model 4 in Figure 1.The second order intercept and slope, representing general language intercept and growth, were estimated by the three first order latent intercepts and slopes. Due to the scaling issues mentioned, the approach of estimating one intercept and slope as first order latent constructs by all three measures was eliminated in favor of using second order constructs.

Considering the interest in the early relation between oral language and reading readiness, the GRTR-R measure was incorporated into the final language structure model. Since there were multiple administrations of the GRTR-R, a separate growth model was estimated for the screener. The relations between the latent intercepts and slope for GRTR-R growth model and the final language structure model were then examined. To this same model, covariates were added at the person level for gender and special services status (having an IFSP/IEP at any session) to examine relations with these potentially important background variables.

3 RESULTS

Linear growth curve models produce five parameters of interest: the mean latent intercept, the variance around the mean intercept, the mean latent slope (average growth rate), the variance in growth rate, and the covariance between intercept and growth rate. In the current study, in which age was centered at 48 months and the time scale is in months, the latent intercept is the true score on the measure at age 4 and the slope is the latent growth per month. Variance around the latent intercept and slope represents individual deviance from the mean values. The covariance between intercept and slope represents the nature of the relation between intercept and slope; if children with higher starting values grow at faster rate then the covariance would be positive and if those children grow at a slower rate the covariance would be negative. To test questions of structure, models will be compared with separate slopes and intercepts for each measure/subtest to alternative structures.

In the current study, children's language and other knowledge were assessed at 6-month intervals over several years. The multiple responses over time can be thought of as nested within persons. Because we have a priori hypotheses about the structure of these multilevel data, we use a multilevel SEM framework.

Descriptive statistics and graphs for the observed variables are presented first. Then the series of language models (shown in Figure 1) are discussed in turn, accompanied by figures. Finally, the selected language model is presented with the addition of the GRTR-R growth model and the covariates of gender and special services status.

3.1 Descriptive statistics

Correlations between the four observed variables PLS-AC, PLS-EC, PPVT, and GRTR, as well as age, are shown in the top part of Table 2. The lower section of Table 2 displays descriptive statistics for the observed variables.

Table 2. *Observed Correlations and Descriptive Statistics*

	1.	2.	3.	4.	5.	
1. PLS_EC	1.00					
2. PLS_AC	.87	1.00				
3. PPVT	.77	.77	1.00			
4. GRTR-R	.73	.73	.65	1.00		
5. Age	.73	.76	.64	.58	1.00	
	N	Mean	SD	Skew	Kurtosis	Range
PLS_EC	1037	43.77	8.66	0.00	0.17	8 - 64
PLS_AC	1038	44.99	7.97	-0.31	-0.18	18 - 62
PPVT	988	58.57	22.96	0.18	-0.38	0 - 122
GRTR-R	828	13.83	5.55	0.02	-1.07	2 - 25
Age	1299	44.87	14.80	-0.63	0.39	4 - 71

Figure 1 shows spaghetti plots for a random selection of 100 participants for each measure. Each dot is a measured time point for an individual child, with lines connecting trajectories for a given child. Each plot also has a dark line, representing a loess regression (i.e., a moving average) for the overall, potentially nonlinear, trend. These plots allow a visual inspection of growth patterns to justify the linear growth trajectory estimated in the models. For all of the measures, there appears to be a clear pattern of linear growth, marked by the loess fit line (the dark line in each graph). Growth trajectory patterns were visually assessed in more detail using the entire data set in groups of 50, and coincide with the patterns shown here. However, for the sake of space and visual clarity, only one group of 100 is presented in Figure 2 and Figure 3 for each measure. Since no indication of non-linear growth was detected, and there was no other reason to assume a different pattern of growth, only linear growth was estimated and no other growth patterns. Importantly, these plots show the uneven spacing of age at assessment. Though the assessments occurred in six-month intervals, the children were different ages at their first assessment, meaning clearly defined cohorts were not available. The mean age of entry into the study was approximately 42 months, or around 3.5 years of age.

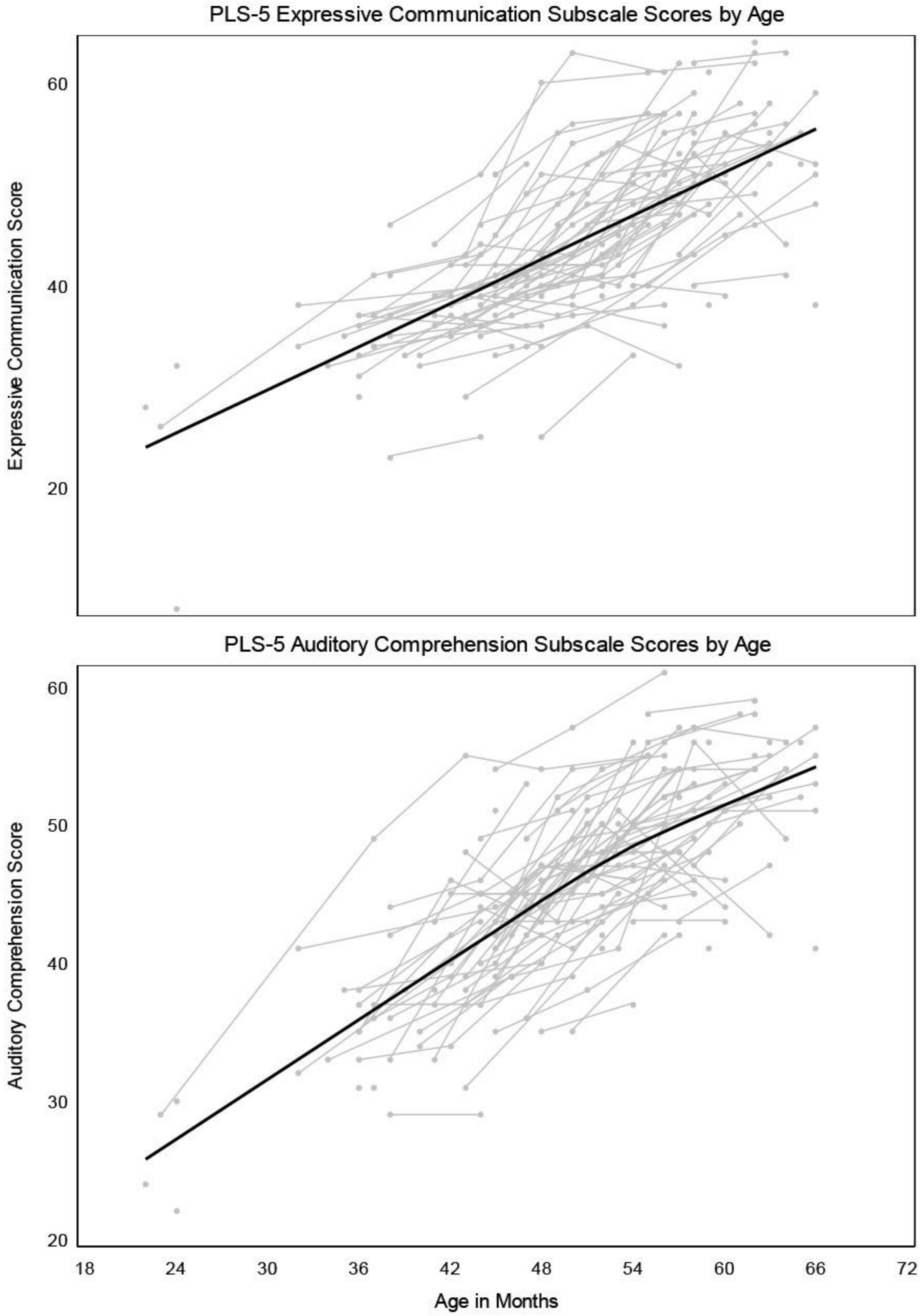

Figure 2. Longitudinal plots of the Preschool Language Scale-5
Note. Observed growth of the PLS-EC shown at the top of the figure and growth on the PLS-AC shown on the bottom of the figure. The dark line is a loess line of fit.

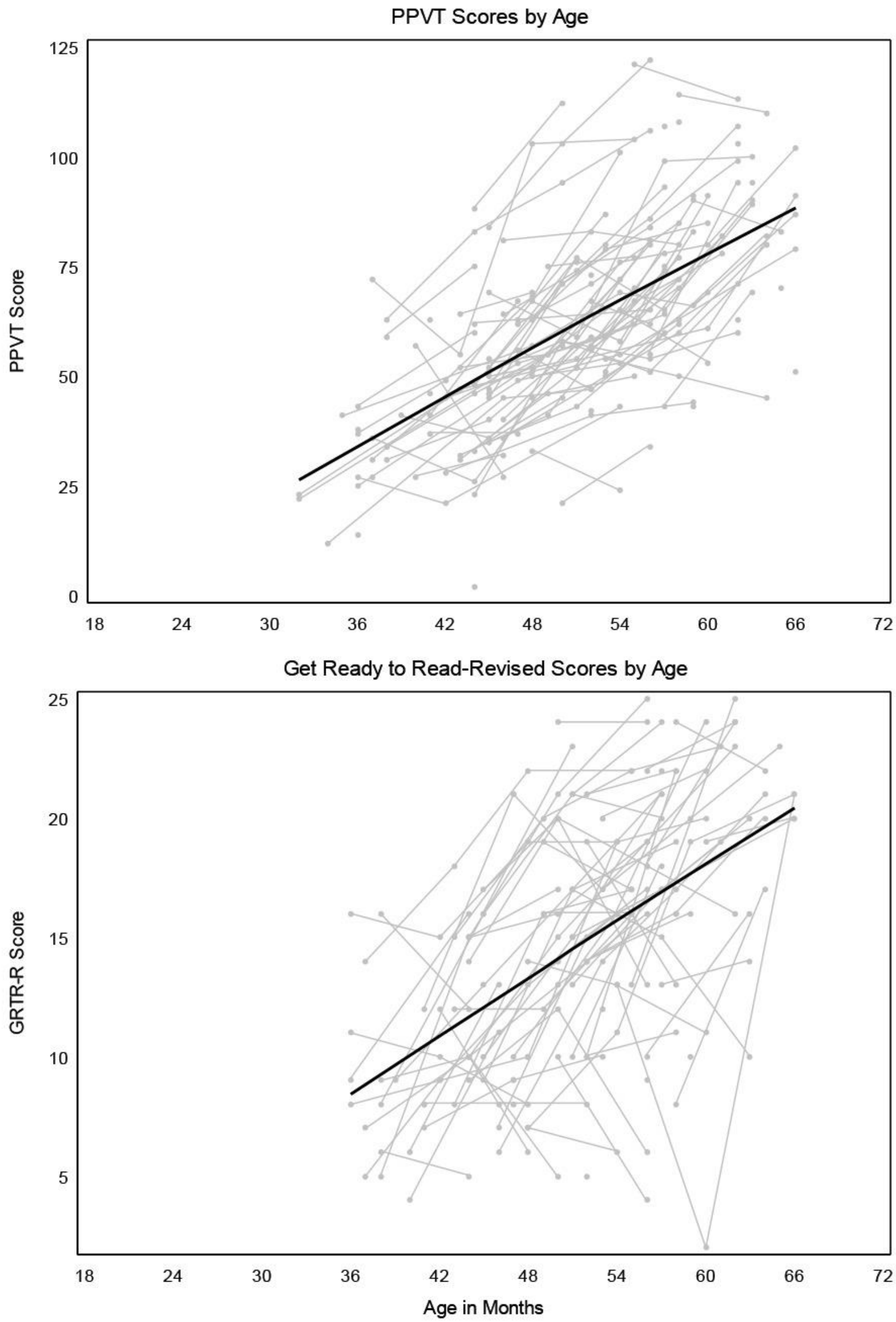

Figure 3. Longitudinal plots of the PPVT-IV and GRTR-R
Note. Observed growth of the Peabody Picture Vocabulary Test-IV shown at the top of the figure and growth in the Get Ready to Read-Revised shown at the bottom of the figure. The dark line is a loess line of fit.

3.2 Models

The general growth curve model can be written as regression equations (from Hox & Stoel, 2005):

$$y_{ti} = \lambda_{0t}\eta_{0i} + \lambda_{1t}\eta_{1i} + \varepsilon_{ti}$$

$$\eta_{0i} = \upsilon_0 + \zeta_{0i}$$

$$\eta_{1i} = \upsilon_1 + \zeta_{1i}$$

where y_{ti} is the predicted outcome of individual *i* at time *t*. The first equation includes three latent scores representing the person-specific intercept (η_{0i}), person-specific slope or rate of change over time (η_{1i}), and individual measurement error (ε_{ti}). The time of measurement is denoted by λ_{1t}, and λ_{0t} represents a constant equal to 1, these are the loading parameters. The person-specific intercept is estimated by the mean intercept (υ_0) and a random error component (ζ_{0i}) or the person's deviation from the mean. The person-specific growth trajectory is estimated by the mean slope (υ_1) and a random error component (ζ_{1i}). Branum-Martin (2013) discusses the correspondence between conventional notation in hierarchical linear modeling and structural equation modeling. The first model, the Trivariate Growth Model, can be thought of as three growth models joined together, so that there would be three sets of the equations above, one for each measure.

While the equations cannot show the relations among multiple measures included in a model (the covariances), these relations can be explicitly represented in SEM figures. All model figures shown are multilevel SEM representations of individual growth models (Branum-Martin, 2013; Mehta & Neale, 2005; Mehta, 2013; Mehta & West, 2000). The bottom of each figure has three observed variables (rectangles), which represent the expressive communication portion of the PLS-5, the auditory comprehension portion of the PLS-5 and the PPVT. The upper part of each diagram has latent factors for each person, one for the Intercept (starting level) and one for the Slope (individual rate of change over time). The triangles represent the mean structure—a typical feature of SEM individual growth models (Little, 2013; Mehta & West, 2000)—with an overall average for each intercept and slope factor. Since age is centered at 48 months and the time scale is in months, the intercept is interpreted as the estimated score at the age of four and the slope represents the estimated amount of growth per month. Latent processes are related within person and are represented by the curved arrows between factors. For model selection, using the logic outlined by Raftery (1995), a better fitting model is both more parsimonious (fewer estimated parameters) and has a smaller Bayesian Information Criterion (BIC).

3.2.1 Trivariate Growth Model

The first model tested was the Trivariate Growth Model, shown in Figure 3. This model allows a separate intercept and growth factor for each of the three measures, and models the relations among the intercepts and slope for all three measures.

Since all of the measures are related to language, at least moderate correlations are expected among the intercepts and slopes—correlations that are too strong indicate that separate constructs may not be necessary. The correlations between the six factors (three intercepts and three slopes) are shown in Table 3. The correlations are strong and positive across the measures—children who have a high score in one measure are likely to have a high score in another measure and are likely to grow faster in all of the measures. The high correlations indicate that a more parsimonious model with fewer latent factors might be more appropriate than a model with a slope and growth factor for each measure. The parameter estimates for the Trivariate Growth Model can be found in Appendix A.

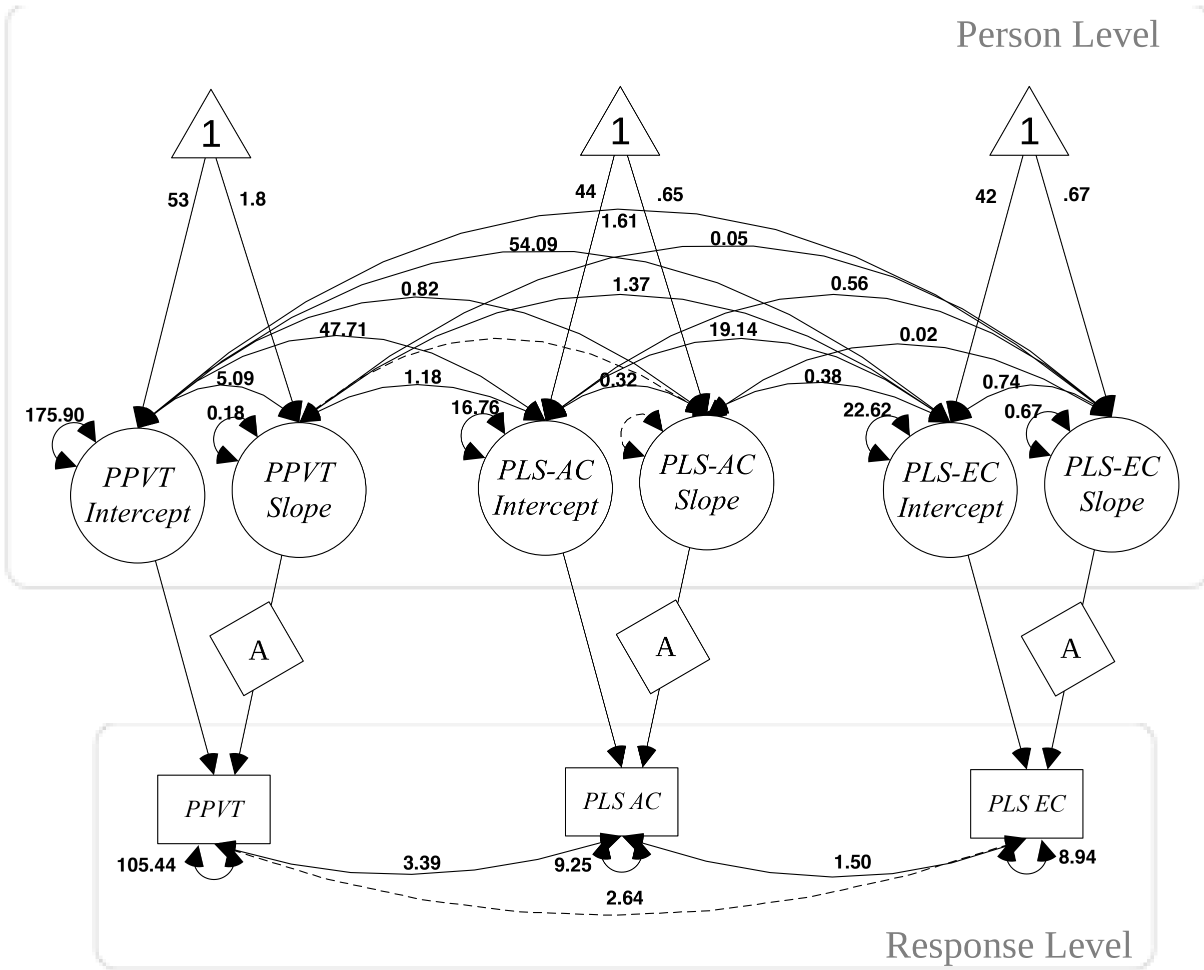

Figure 4. Trivariate Growth Model
Note. Diamonds represent definition variables of age, allowing age to be treated as individual-specific (see Mehta & West, 2000). Dashed lines represent parameters that were not statistically significant. The large boxes drawn are to clarify the nested nature of the data.

Table 3. *Correlations Among Latent Variables in the Trivariate Growth Model*

	1.	2.	3.	4.	5.	6.
1. Intercept PLS EC	1					
2. Intercept PLS AC	0.98	1				
3. Intercept PPVT	0.87	0.88	1			
4. Slope PLS EC	0.78	0.69	0.62	1		
5. Slope PLS AC	0.85	0.82	0.65	0.80	1	
6. Slope PPVT	0.67	0.67	0.90	0.62	0.59	1

3.2.2 Method Factor Model

It is possible that the strong correlations seen in Table 3 are partly a result of a method effect—two out of the three measures are part of the same test, with similar items and administration. This model includes separate starting levels and growth trajectories for the PPVT and the PLS-5, where the PLS-5 factor is estimated by both subtests of the PLS-5. Since the two subtests of the PLS use similar task types and are scored similarly with an almost equal score range, it is reasonable to equate the factor loadings, allowing the model to be identified. In this model, there are only four latent factors: Intercept and Slope, each for the PLS-5 and the PPVT (The parameter estimates for the Method Factor Model can be found in Appendix B).

As can be seen in Table 4, the more parsimonious Method Factor Model did not fit better than the Trivariate Growth Model: there is an increase in BIC (though small) despite reducing the number of estimated parameters. The Likelihood Ratio Test (Table 4) also shows that the Method Factor model fits significantly worse than the Trivariate Growth Model. Since there was no evidence that the Method Factor Model was an improvement in fit, the model was rejected.

3.2.3 Second-Order Language Factor Model

To represent a modality-general language structure, with a general language intercept and general language slope, second order latent variables of intercept and slope were added to the model. This second-order approach was chosen over using all three measures to estimate a single first order latent intercept and slope because of the scaling concerns mentioned previously. As can be seen in Figure 5, the general Language intercept is estimated by the first-order latent intercepts of the PLS-EC, PLS-AC, and PPVT. The general Language slope is estimated by the first-order latent slopes of the PLS-EC, PLS-AC, and PPVT.

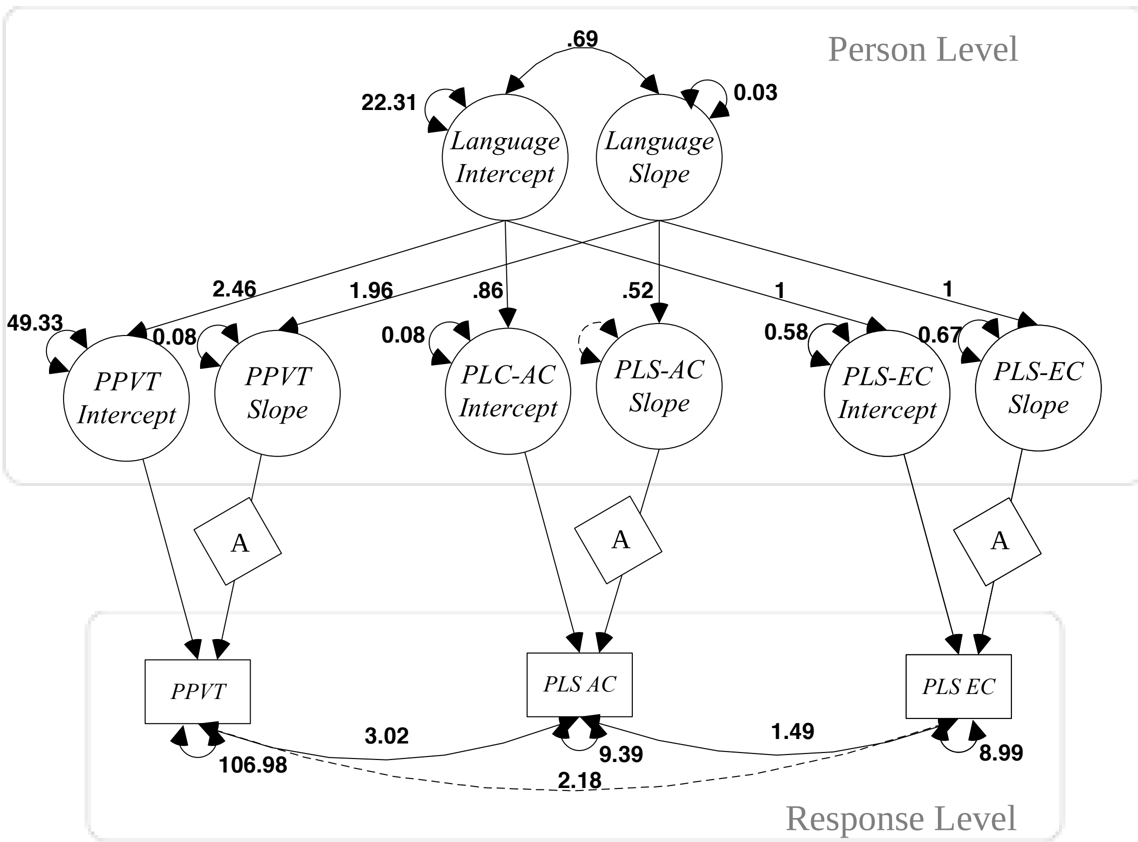

Figure 5. Unified Language Model with general language intercept and slope
Note. Diamonds represent definition variables of age, allowing age to be treated as individual-specific (see Mehta & West, 2000). Dashed lines represent non-significant parameters. The large boxes drawn are to clarify the nested nature of the data.

The fit of the models can be seen in Table 4 below. The Likelihood Ratio Test results shown in Table 4 reject the Unified Language Model in favor of the Trivariate Model, however, the Trivariate Model is not viewed as the most appropriate model. The pattern of high correlations among the latent constructs (see Table 3) needs to be accounted for, which is achieved by the Unified Language Model. Using the logic of Raftery (1995), which favors parsimony and smaller BIC values to support model selection, the Second Order model was selected as the best-fitting model. Fit indices commonly used to evaluate structural equation models, such as root mean square error of approximation (RMSEA) and confirmatory fit index (CFI) are not available for the multilevel structural equation modeling approach used here. The model parameters for the Unified Language Model are displayed in Table 5.

Table 4. Model fit

Model	Estimated Parameters	BIC	ΔBIC	Likelihood Ratio Test
1. Trivariate Model	33	19594		
2. Modality Model	-	-		
3. Method Factor Model	20	19663	+ 69*	χ^2 (13) = 159.11, $p < .05$*
4. Unified Language Model	25	19569	- 24	χ^2 (8) = 31.02, $p < .05$*
5. Full Model with GRTR and covariates	39	23972		

Note. Change in BIC and Likelihood Ratio Test use the Trivariate Model as the comparison model. For the likelihood Ratio Test, * indicates the nested model fit is worse than the Trivariate model.

The correlation between the second order language intercept and growth is .89, indicating that children who start off with higher levels of language grow at a faster rate. Looking at the model loadings show that for both the second order slope and intercept constructs, the PPVT has the strongest loadings—since the model is scaled to the PLS-EC, the loadings can be interpreted relative to the loading of 1 for the PLS-EC. The stronger loadings for the PPVT are a reflection of the scaling differences of the tests—the PPVT has a larger range and may be more sensitive to detecting differences too small to be captured by the scale of the PLS-5.

Means of the second order latent intercept and slope are zero, where zero is the average level and the average growth of language. For models with a second order latent factor type of model, the estimates for the first order factor become intercepts, or the expected values at average levels of the second order factor. At average levels of language, the expected values of the PLS-AC and the PLS-EC are at the 30th percentile, and the PPVT is at the 32nd percentile according to the relevant manuals. The expected growth estimate is the amount of growth expected on the assessment per month. Multiplying by 12 gives the expected yearly growth, which is 8 points on the PLS-EC and 7.75 points on the PLS-AC. At average language levels at 48 months and average expected growth, the projected scores at age five would remain at the 30th percentile for both of the PLS-5 subscales. For the PPVT, the projected average score at age five of 75 items correct is at the 34th percentile. Considering that the sample was drawn from a population frequently demonstrated to have lower than average scores on language assessments, the lower mean score is in line with expectations. The relations between language—as represented by the PLS-5 and the PPVT—and reading readiness were of further interest to this project, leading to the full model presented next.

Table 5. Parameter estimates for the Unified Language Model

	Parameter	Estimate	*SE*	*p*
	2nd Order Person Level			
Covariance	Language with Language slope	0.692	0.098	< .001
Variances				
	Language	22.309	2.498	< .001
	Language slope	0.027	0.009	.004
Factor loadings				
	PLS-EC	1	0	—
	PLS-AC	0.864	0.037	< .001
	PPVT	2.464	0.134	< .001
	SLOPEEC	1	0	—
	SLOPEAC	0.521	0.084	< .001
	SLOPEPPVT	1.962	0.41	< .001
	Person Level			
Intercepts				
	PLS-EC	42.252	0.256	< .001
	PLS-AC	43.574	0.220	< .001
	PPVT	52.688	0.743	< .001
	SLOPEEC	0.67	0.016	< .001
	SLOPEAC	0.646	0.013	< .001
	SLOPEPPVT	1.835	0.051	< .001
	Response Level			
Residual variances				
	PLS-EC	0.577	0.680	.396
	PLS-AC	0.083	0.507	.870
	PPVT	49.328	8.911	< .001
	SLOPEEC	0.011	0.006	.073
	SLOPEAC	0.001	0.004	.888
	SLOPEPPVT	0.079	0.082	.336
Residual covariances				
	PLS-AC with PLS-EC	1.493	0.467	.001
	PPVT with PLS-EC	2.177	1.438	.130
	PPVT with PLS-AC	3.017	1.458	.038

Note. Dashes indicate a parameter fixed for model identification.

3.3 Full Model

Since the Unified Language Model was selected as the best-fitting model, the GRTR-R and covariates of gender and special services were added to the Unified Language Model to create the Full Model. The Full Model is visually displayed in Figure 5 and the parameter estimates are shown in Table 6.

Looking first at the two person-level covariates included into the model, Figure 5 shows with a dashed line that the Special Services variable did not significantly predict language status or growth in this sample. This should be interpreted with caution since the sample mean is quite a bit lower than the norming mean—there may not be the sensitivity necessary to detect differences between typically developing children and those with special needs when the sample mean is significantly lower than the norming mean. Since no information was available about specific diagnoses (such as a diagnosed language delay) the Special Services variable was used to test if model fit differed by group (those identified as needing special services compared to the other students). Additionally, there is not information on what services are included in the IFSP/IEP – these can cover a wide range of non-language services such as behavioral interventions, and so this variable may not be sufficient to identify a subgroup of children needing language-specific support services. The gender covariate (male) shows that boys have lower starting language levels and grow more slowly than girls. This pattern is consistent with previous literature indicating higher language performance for girls than boys in the early years (Bornstein, Hahn, & Haynes, 2004).

Turning to the growth model for the GRTR-R in this model, it should be noted that no paths were included to predict variation in the GRTR-R slope factor. Since the GRTR-R was administered twice to each child (though there were some exceptions two administrations were in the design), there was not enough information to estimate random slopes for the GRTR-R: only an average growth parameter was estimated. Since individual slopes were not estimated, no predictors of variation in slope were incorporated into the model. The lack of variation in the

GRTR-R latent slope is also why the covariance between the GRTR-R intercept and slope is non-significant.

The GRTR-R intercept—the model-predicted mean value at 48 months—is 13. This mean score is in the basic understanding category range of the screener. A score of 13 is also in the range of 12-13 which was suggested by Whitehurst (2001) as the lower bounds of scores likely to predict successful reading in 2nd grade. The dashed line from the second order language intercept to the GRTR-R intercept shows that language starting level is not a significant predictor of reading readiness. The dashed line from language growth to GRTR-R shows that language growth is also not a significant predictor of reading readiness.

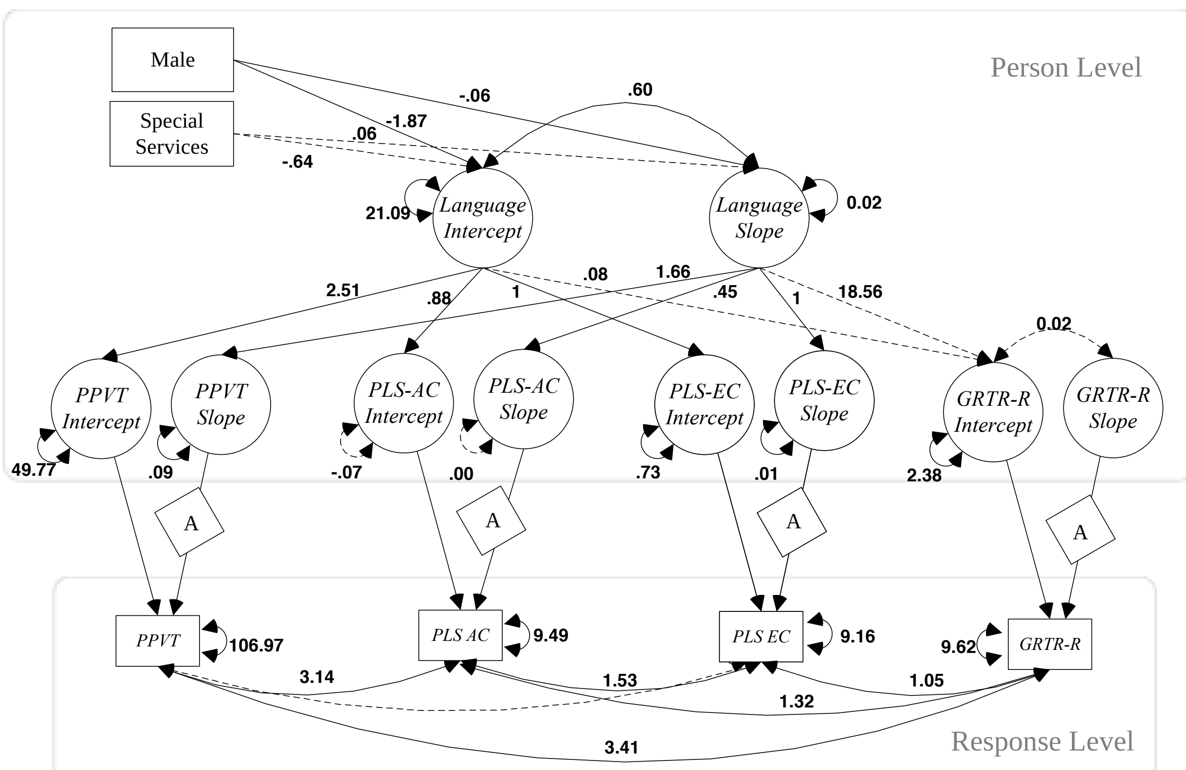

Figure 6. Full model with covariates.
Note. Diamonds represent definition variables of age, allowing age to be treated as individual-specific (see Mehta & West, 2000). Dashed lines represent parameters which were not statistically significant. The large boxes drawn are to clarify the nested nature of the data.

Table 6. Parameter estimates for Unified Language Model with GRTR-R and covariates

Level	Type	Relation	Estimate	*SE*	*p*
Person	Regression	Language on MALE	-1.867	0.478	< .001
		Language on SPECSER	-0.638	0.771	.408
		Language slope on MALE	-0.060	0.020	.004
		Language slope on SPECSER	0.056	0.043	.197
		GRTR-R on Language	0.084	0.306	.784
		GRTR-R on Language slope	18.56	10.188	.068
	Covariance	Language with Language slope	0.604	0.083	< .001
		GRTR with SLOPEGRTR	0.015	0.021	.724
	Loadings	PLS-EC	1	0	—
		PLS-AC	0.877	0.036	< .001
		PPVT	2.516	0.133	< .001
		SLOPEEC	1	0	—-
		SLOPEAC	0.451	0.093	. < .001
		SLOPEPPVT	1.663	0.440	< .001
	Intercepts	PLS-EC	43.283	0.332	< .001
		PLS-AC	44.497	0.281	< .001
		PPVT	55.386	0.909	< .001
		GRTR-R	13.001	0.298	< .001
		SLOPEEC	0.670	0.019	< .001
		SLOPEAC	0.634	0.014	< .001
		SLOPEPPVT	1.815	0.056	< .001
		SLOPEGRTR	0.368	0.019	< .001
	Residual variances	Language	21.089	2.345	< .001
		Language slope	0.019	0.006	. < .001
		GRTR-R	2.375	1.196	0.047
		PLS-EC	0.727	0.597	.224
		PLS-AC	-0.070	0.460	.879
		PPVT	49.771	8.886	< .001
		SLOPEEC	0.012	0.007	.046
		SLOPEAC	0.000	0.004	.998
		SLOPEPPVT	0.090	0.078	.248
Response	Residual covariances	PLS-AC with PLS-EC	1.528	0.439	.001
		PPVT with PLS-EC	2.512	1.37	.067
		PPVT with PLS-AC	3.138	1.402	.025
		GRTR-R with PLS-EC	1.047	0.456	.022
		GRTR-R with PLS-AC	1.319	0.389	.001
		GRTR-R with PPVT	3.414	1.377	.013
	Residual variances	PLS-EC	9.155	0.657	< .001
		PLS-AC	9.487	0.771	< .001
		PPVT	106.970	11.698	< .001
		GRTR-R	9.617	0.771	< .001

Note. Dashes indicate a parameter fixed for model identification.

4 DISCUSSION

The purpose of this study was to examine language structure and the nature of growth in language as it relates to language modality—is there support for the division of language into expressive and receptive language for children in pre-kindergarten and younger? The sample used was primarily typically-developing, low SES African American children. With a longitudinal dataset of 475 children from urban childcare centers assessed at approximately six-month intervals with the PLS-5, PPVT-4, multi-level structural equation modeling was used to test a series of growth models for language structure and the relation of language to reading readiness as measured by the GRTR-R. This study adds to the literature in providing insight into the structure of language, particularly because it utilizes an understudied population of young, low SES African American children. Also, this study uses employs random-effects SEM to account for the differences in starting ages and assessment intervals in data collection (Mehta & Neale, 2005; Mehta & West, 2000; Muthen, 2017). Though the technique is not new in the field of psychometrics and statistics, it is still underutilized in longitudinal developmental designs where clear-cut cohorts are frequently not achieved due to the practicalities of data collection.

4.1 Validation of the PLS-5

This study is the first published factor analysis of the PLS-5. The models imply that the PLS-5 may be better represented as a single test, rather than two subtests of expressive and receptive language. The Trivariate growth model showed the correlation between the PLS-AC and PLS-EC latent intercepts was .98, indicating the PLS-AC and PLS-AC may be indistinguishable at the population level. The correlation between the PLS-AC and PLS-EC latent slopes was also high at .80, suggesting that growth rates are measured essentially the same way between the tests. The strong correlations of the intercept factors and the slope latent factors suggest there is so much overlap between the two subscales that they reasonably represented as a single test rather than two. There was also no evidence of discriminant validity: the PLS-5 subtests did not show differential relations to the PPVT, an external measure of language ability. To support the expressive/receptive divide of the PLS-5 subscales, differential relations would be expected. The PLS-AC, a receptive language measure, would be expected to have a stronger correlation with the PPVT, another receptive language measure. The correlation between the PLS-EC, an expressive language measure, and the PPVT would be expected to have a moderate correlation as both are language measures but the correlation would not be as strong as that between the PLS-AC and the PPVT. Instead, the correlations were indistinguishably high and homogeneous. This provides another layer of evidence that the PLS-5 is adequately represented as a single test of language rather than two subscales.

4.2 Language Modality and the Structure of Language

The final language structure model selected in this study is that of a general construct of language. Even though not all of the model selection evidence supports the selection of the Unified Language Model, most of it does. The final model includes second order constructs of general language level and growth, estimated by the first order growth factors for the PLS-AC, PLS-EC, and PPVT: the high correlations among the three language measures in the Trivariate growth model indicate the necessity of simplifying the structure. While the likelihood ratio test rejects both alternate models (the Method Factor Model and Unified Language Model) in favor of the Trivariate Model, the smaller BIC value supports the selection of the more parsimonious Unified Language Model.

The Unified Language Model is also consistent with the majority of prior measurement studies of language (e.g., Tomblin & Zhang, 2006; Sommers, Erdige, & Peterson, 1978) that concluded against separate constructs of expressive and receptive language, though those studies included children older than the children in the current study. One of these studies, however, Lonigan and Milburn (2017), included prekindergarten children in their cross-sectional study, and found for that grade only the two-factor structure of expressive and receptive language fit better than a single factor structure. There are several possible reasons for why the current study had opposite conclusions. First, the Lonigan and Milburn (2017) study was cross-sectional rather than longitudinal. The motivation for conducting longitudinal studies in spite of the increased cost and time is precisely because they can reveal patterns that are masked in a cross-sectional design (McArdle 1988; 2009). Secondly, different language assessments were administered. Thirdly, there are differences in the participant characteristics; while the Lonigan and Milburn (2017) sample also included a majority of low SES children most of the children were White while more than 99% of the current sample were African American. The difference in findings between the current and this prior study highlight the need for further research, with careful consideration of the design, language assessments, and participant characteristics.

This study extends the previous body of language measurement research by using a younger sample of children, pre-kindergarten and younger, who are primarily African American and from low SES families. This study is also longitudinal in design, unlike the majority of previously reviewed language measurement studies, which were primarily cross-sectional in design.

4.3 Language Growth

Since the measures were administered multiple times, this study examined language growth in addition to language structure. The linear growth trajectory pattern indicates steady and consistent growth in language over the age range covered in the sample. A child with average language at age 4 whose language developed at the average rate would be at the 30th percentile for all measures at age 4 and remain at approximately the 30th percentile for all of the measures at age 5. A mean at the 30th percentile is quite a bit lower than the 50th percentile mean of the norming sample, but the results are consistent with previous literature on children from low SES families indicating lower language levels (e.g., García, 2015).

The high correlation of .89 between the latent general language intercept and slope means that children who start with higher levels of language tend to increase their language skills at a faster rate. On a more promising note, the average rate of growth estimates language scores that stay in the same percentile range from age four to age five. This means that while the average rate of language growth is not great enough to close the gap between this lower SES sample and the norming sample, it is also not so slow as to widen the gap between the groups.

The person-level covariates show that boys have significantly lower language levels than girls, and boys' language growth rate is slower than girls' language growth rate as well, which is consistent with prior research on gender differences in early language (Bornstein, Hahn, & Haynes, 2004) particularly for children from low SES backgrounds (Barbu et al., 2015). Also, children with an IFSP or IEP were no different than other children in either language levels or growth. This lack of difference may be due to the overall lower language of the entire sample. Alternatively, the simple, global administrative designation of having an individual plan may be insufficient to show a reliable relation: many of these children may have conditions not related to language disorders and so would not be expected to have a measureable language difference. The current approach may be useful for a closer, examination of detailed, clinical diagnoses of disabilities.

4.4 Relation to reading readiness

The regression paths of the GRTR-R intercept on the second order language intercept and second order language slope were not significant. Though growth in language was not significant at $p = .06$, it is a better predictor of reading readiness than language level. The regression weight of the GRTR-R on language slope is interpreted as any other regression. With all other parts of the model held equal, one standard deviation increase in language level ($SD = .14$) multiplied by the regression weight of 18.56 gives an expected 2.62 points higher on the GRTR-R.

The weak relation between language and reading readiness is likely due to the fact that the GRTR-R assesses a different set of skills that are less related to language comprehension. For example, the assessment includes items asking the child to identify the cover of a book. This is not related to language comprehension. Though based on the model here the GRTR-R is only weakly related to the language aspect of reading; it is likely that there is a stronger relation to the decoding aspect of reading skills (Hoover & Gough, 1990) because it contains items that require recognition of letters. The GRTR-R is intended to measure readiness to read rather than actual reading comprehension. As such, the lack of relation to language and language growth is not surprising.

4.5 Limitations

There are several limitations to this study. The results are limited to a typically developing population considered at-risk of low academic achievement due to socio-economic status. The structure of language may be different for a disordered population, such as children with a diagnosed language delay or disability. The primary purpose of the PLS-5 is to diagnose language disorders or delays and identify children who may need additional language support services. For this reason, conclusions about changing the structure of a language assessment should not depend solely on a study with primarily typically developing children. The same assessments would need to be administered in the same longitudinal design to children with a language diagnosis to compare language structure and growth of a disordered population. It may be that there is clinical utility in the division of a language assessment into separate measures of expressive and receptive language. However, there is likely substantial overlap between this sample of children and those with diagnosed language disorders considering the low language scores, so it would not be surprising to find the same results with a sample of children with an ‘official’ diagnosis.

It should be noted that the study sample comes from a population identified in the literature as being at high risk for language delay/disorder and later reading problems related to environmental factors such as maternal education, household income, and exposure to words and language experiences such as turn-taking conversations and dialogic reading (Rinderman & Baumeister, 2015). However, with only one measure for expressive language and two for receptive language, we cannot provide a definitive evaluation of the nature of language or change in language even for this population. Also, though the sample included a set of children identified as needing special services based on having an IFSP or IEP at any point in the study, those services were not specified as language services. The relation of special services status and language status and growth provides limited insight on relational differences for children with a diagnosis that is specific to language.

A limitation of the modeling approach is that classroom variability is ignored, and previous research shows significant effects of classroom on language and literacy (e.g., Mehta, Foorman, Branum-Martin, & Taylor, 2005). Unfortunately, there was too much missing information about classroom assignment or teacher assignment to account for the nesting of children in classrooms. Incorporating this structure into the model could have altered the final model chosen. For example, Mehta et al. (2005) showed different language and literacy structures at the individual and classroom levels.

4.6 Future Directions

This study demonstrates the use of an individual-specific growth curve model of language and reading readiness measures. Rather than grouping children into a more coarsely grained time scale, the fine-grained time scale of age in months was retained. This approach allows the estimation of unbiased parameters of the intercept and covariance between the intercept and slope, identical to multilevel modeling (Mehta & West, 2000). This is the appropriate approach for data collection designs where participants start at different ages and are not assessed at exact intervals. Future studies could incorporate additional language measures to allow a greater selection of alternate language structure models. Additionally, an actual reading measure at an appropriate age (e.g., seven or eight), rather than a reading readiness screener, would allow for more concrete conclusions about the relation of early language and language growth to early reading.

Though the final model shows latent correlations greater than .8 for the PLS-5 subscale intercepts and growth, administration of only one subscale should be carefully considered and should not be based solely on the results of this study. The sample is not representative of the wider population of US schoolchildren: it is a sample of low SES African American children. Therefore, the results may not generalize beyond this group. The homogeneity of the current sample could also have inflated the relationship of the subscales or even between the second order language intercept and slope. The sample used here is also not a clinical sample: the structure of language may be different for clinical populations of children with language delay or disability.

4.7 Conclusions

This study shows that for a typically developing low SES African-American population with the available assessments in the data set, there is not support for expressive and receptive language as separate constructs. Rather, the structure of the PLS-5 is adequately represented by a single factor. In other words, language performance among preschoolers seems to be relatively neutral to modality. Thus, for the purpose of assessing language in an African American low SES (specifically non-clinical) population, there seems to be little utility in dividing the PLS-5 scores into expressive and receptive language scores. Either score individually or the total test score may be an adequate indicator of language status and growth.

REFERENCES

Anthony, J.L, Davis, C. Williams, J.M., & Anthony, T.L. (2014). Preschoolers' oral language abilities: A multilevel examination of dimensionality. *Learning and Individual Differences,* 35, 56-61. doi: 10.1016/j.lindif.2014.07.004

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Publishing, Arlington, VA.

Arnold, D. H., & Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*, *54*(1), 517-545. doi: 10.1146/annurev.psych.54.111301.145442

Barbu, S., Nardy, A., Chevrot, J. P., Guellaï, B., Glas, L., Juhel, J., & Lemasson, A. (2015). Sex differences in language across early childhood: Family socioeconomic status does not impact boys and girls equally. *Frontiers in Psychology*, *6*, 1-10. doi: 10.3389/fpsyg.2015.01874

Bauer, D.J., Goldfield, B.A., & Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics, 23*(3), 313-335.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6*, 183-200.

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. New York, NY: Wiley.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley-Interscience.

Bornstein, M. H., Hahn, C. S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, *24*(3), 267-304. doi: 10.1177/0142723704045681

Bornstein, M. H., Hahn, C.S., & Putnick, D. L. (2016). Stability of core language skill across the first decade of life in children at biological and social risk. *Journal of Child Psychology and Psychiatry, 57*(12),1434-1443. doi: 10.1111/jcpp.12632

Bornstein, M. H., & Hendricks, C. (2012). Basic language comprehension and production in > 100,000 young children from sixteen developing nations. *Journal of Child Language, 39*(4), 899-918.

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Suwalsky, J. T. (2014). Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Development*, *85*(4), 1346-1356.

Bornstein, M. H., & Putnick, D. L. (2012). Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Developmental Psychology*, *48*(2), 477.

Branum-Martin, L. (2013). Multilevel modeling: Practical examples to illustrate a special case of SEM. In Y. Petscher, C. Schatschneider & D.L. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 95-124). New York: Routledge.

Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. Cambridge University Press.

Carrow, E. (1973). *Test for Auditory Comprehension of Language*, 5th ed. Boston: Teaching Resources Corp.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language,* *and Hearing Research*, *45*(6), 1142-1157. doi: 10.1044/1092-4388(2002/093)

Denman, D. Speyer, R. Munro, N., Pearce, W.M., Chen, Y., and Cordier, R. (2017). Psychometric properties of language assessments for children aged 4-12 years: A systematic review. *Frontiers in Psychology*, 8, 2-28, doi: 10.3389/fpsyg2017.01515

Dietrich, S.E., Assel, M.A., Swank, P., Smith, K.E., & Landry, S.H. (2006). The impact of early maternal verbal scaffolding and child language abilities on later decoding and reading comprehension skills. *Journal of School Psychology, 43,* 481-494. doi: 10.1016/j.jsp.2005.10.003

Dockrell, J.E. & Marshall, C.R. (2015). Measurement issues: Assessing language skills in young children. *Child and Adolescent Mental Health, 20*(2), 116-125. doi:10.1111/camh.12072

Dunn, L. M., & Dunn, L. M. (1981). *Manual for the Peabody Picture Vocabulary Test-Revised.* Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test manual* (4th ed.). Minneapolis, MN: Pearson Assessment.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1-185.

Foorman, B.R., Koon, S., Petscher, Y. Mitchell, A., Truckenmiller, A. (2015). Examining general and specific factors in the dimensionality of oral language and reading in 4th-10th grades. *Journal of Educational Psychology, 107*(3), 884-899. doi: /10.1037/edu0000026

García, E. (2015). Inequalities at the Starting Gate: Cognitive and Noncognitive Skills Gaps between 2010-2011 Kindergarten Classmates. Report. *Economic Policy Institute*.

Gathercole, S., & Baddeley, A. (1996). *Children's Test of Non Word Repetition (CNrep)* Sydney, NSW: Psychological Corporation.

Goldfield, B.A., & Reznick, S.J. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language, 17*(1), 171-183.

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua, 120*(8), 1887-1897.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*. doi: 10.1007/BF00401799

Hox, J. , & Stoel, R.D. (2005). Multilevel and SEM approaches to growth curve modeling. In B. S. Everitt and D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science,* Vol. 3 (pp. 1296–1305). Chichester, UK: John Wiley & Sons.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological bulletin*, *104*(1), 53. doi: 10.1037//0033-2909.104.1.53

Kaplan, D., & Elliott, P. R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(1), 1-24. doi: 10.1080/10705519709540056

Kieffer, M. J., Petscher, Y., Proctor, C. P., & Silverman, R. D. (2016). Is the whole greater than the sum of its parts? Modeling the contributions of language comprehension skills to reading comprehension in the upper elementary grades. *Scientific Studies of Reading, 20*(6), 436-454. doi: 10.1080/10888438.2016.1214591

Klem, M., Gustafsson, J., & Hagtvet, B. (2015). The dimensionality of language ability in four-year-olds: Construct validation of a language screening tool. *Scandinavian Journal of Educational Research, 59*(2), 195-213. doi:10.1080/00313831.2014.904416

Language and Reading Research Consortium. (2015a). The dimensionality of language in young

Spanish–English dual-language learners. *Journal of Speech, Language, and Hearing Research*, 58, 754–766. doi: 10.1044/2015_jslhr-l-13-0266

Language and Reading Research Consortium. (2015b). The dimensionality of language ability in young children. *Child Development*, 86, 1948–1965. doi: 10.1111/cdev.12450

Language and Reading Research Consortium. (2016). Oral language and listening comprehension: Same or different constructs. *Journal of Speech, Language, and Hearing Research, 60,* 1273-1284. doi: 10.1044/2017_JSLHR-L-16-0039

Lee, L. (1969). *The Northwestern Syntax Screening Test.* Evanston, Ill.: Northwestern University Press.

Lee, L. (1974). *Developmental sentence analysis*. Evanston, Ill.: Northwestern University Press.

Leonard, L.B. (2009). Is expressive language disorder an accurate diagnostic category. *American Journal of Speech-Language Pathology, 18,* 115-123. doi: 1058-0360/09/1802-0115

Levelt, W. (1993). The architecture of normal spoken language use. In G. Blanken, J. Dittmann, H. Grimm, J. Marshall, & C-W. Wallesch (Eds), Linguistic disorders and pathologies: An international handbook (pp. 1–15). Berlin, Germany: Walter de Gruyter.

Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. New York, NY: Guilford Press.

Little, T.D., Bovaird, J.A., & Slegers, D.W. (2006). Methods for the analysis of change. In D. Mroczek & T.D. Little (Eds.), *Handbook of Personality Development* (pp. 181-211). Mahwah, NJ: Erlbaum.

Little, T.D., Card, N.A., Preacher, K.J., McConnell, E. (2009). Modeling longitudinal data from research on adolescence. In R.M. Lerner & L. Steinberg (Eds.), *Handbook of Adolescent Psychology- 3rd edition* (pp. 15-54). New York: Wiley.

Lonigan, C. J., & Milburn, T. F. (2017). Identifying the dimensionality of oral language skills of children with typical development in preschool through fifth grade. *Journal of Speech, Language, and Hearing Research, 60*(8), 2185-2198. doi: 10.1044/2017_jslhr-l-15-0402

Lonigan, C. J., & Wilson, S.B. (2008). Report on the Revised Get Ready to Read! Screening Tool: Psychometrics and normative information.

Matov, J., Mensah, F., Cook, F., & Reilly, S. (2018). Investigation of the language tasks to include in a short-language measure for children in the early school years. *International Journal of Language & Communication Disorders*. doi: 10.1111/1460-6984.12378

Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology, 4,* 1-13. doi:10.3389/fpsyg.2013.00169

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In *Handbook of multivariate experimental psychology* (pp. 561-614). Boston, MA: Springer.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60,* 577-605. doi: 10.1146/annurev.psych.60.110707.163612

McCarthy, D. (1930). The Language Development of the Pre-School Child. *Institute of Child Welfare Monographic Series, 4*. University of Minnesota Press, Minneapolis, MN.

Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading*, *9*(2), 85-116. doi: 10.1207/s1532799xssr0902_1

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259. doi: 10.1037/1082-989x.10.3.259

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*(1), 23. doi: 10.1037//1082-989x.5.1.23

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher*, *18*(2), 5-11. doi: 10.3102/0013189x018002005

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741. doi: 10.1037//0003-066x.50.9.741

Miles, S., Fulbrook, P., & Mainwaring-Mägi, D. (2016). Evaluation of standardized instruments for use in universal screening of very early school-age children: Suitability, technical adequacy, and usability. *Journal of Psychoeducational Assessment, 1,* 1-21. doi: 10.1177/0734282916669246

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*(4), 338-354. doi: 10.1111/j.1745-3984.1991.tb00363.x

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*(3), 376-398. doi: 10.1177/0049124194022003006

Muthén, L. K. (2017). Mplus User's Guide, Ed 7, Los Angeles, CA, Muthén & Muthén, 1998–2012.

Newcomer, P., & Hammill, D. (1988). Test of Oral Language Development—2: Primary. Austin, TX: Pro-Ed.

Putnick, D. L., Bornstein, M. H., Eryigit-Madzwamuse, S., & Wolke, D. (2017). Long-term stability of language performance in very preterm, moderate-late preterm, and term children. *The Journal of Pediatrics*, 181, 74–79.e3. doi: 10.1016/j.jpeds.2016.09.006

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental psychology*, *28*(3), 406. doi: 10.1037//0012-1649.28.3.406

Rindermann, H., & Baumeister, A. E. (2015). Parents' SES vs. parental educational behavior and children's development: A reanalysis of the Hart and Risley study. *Learning and Individual Differences, 37*, 133-138. doi: 10.1016/j.lindif.2014.12.005

Semel, E., Wiig, E. H., & Secord, W. A. (1995). Clinical evaluation and language fundamentals – 3rd edition (CELF-III). San Antonio, TX: Psychological Corporation.

Shanahan, T., & Lonigan, C. J. (2010). The National Early Literacy Panel: A summary of the process and the report. *Educational Researcher*, *39*(4), 279-285. doi: 10.3102/0013189X10369172

Sommers, R. K., Erdige, S., & Peterson, M. K. (1978). How valid are children's language tests?. *The Journal of Special Education*, *12*(4), 393-407. doi: 10.1177/002246697801200405

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, *27*(4), 720-731.

Straight, S. H. (1976). Comprehension versus production in linguistic theory. *Foundations Of Language*, *14*(4), 525-540.

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1193-1208.

Van Hout, A., Harrigan, K., & de Villiers, J. (2010). Asymmetries in the acquisition of definite and indefinite NPs. *Lingua*, *120*(8), 1973-1990. doi: 10.1016/j.lingua.2010.02.006

Wallace, G., & Hammill, D. D. (1994). *Comprehensive receptive and expressive vocabulary test*. Austin, TX: Pro-Ed.

Whitehurst, G.J. (2001). The NCLD Get Ready to Read! screening tool technical report. Retrieved from http://www.getreadytoread.org/images/content/downloads/GRTR_screening_tool/technicalreport.pdf

Whitely, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological bulletin*, *93*(1), 179. doi: 10.1037//0033-2909.93.1.179

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). PLS-5: Preschool Language Scale-5 [measurement instrument]. *San Antonio, TX: Psychological Corporation*.

APPENDICES

Appendix A

Table 7. Parameter Estimates for the Trivariate Growth Model

Level	Parameter Type	Relation	Estimate	*SE*	*p*
Person	Means				
		PLS-EC	42.27	0.26	< .001
		PLS-AC	43.58	0.22	< .001
		PPVT	52.67	0.74	< .001
		SLOPEEC	0.67	0.02	< .001
		SLOPEAC	0.65	0.01	< .001
		SLOPEPPVT	1.83	0.05	< .001
	Variances				
		PLS-EC	22.62	2.58	< .001
		PLS-AC	16.76	1.97	< .001
		PPVT	175.90	19.31	< .001
		SLOPEEC	0.04	0.01	< .001
		SLOPEAC	0.01	0.01	0.16
		SLOPEPPVT	0.18	0.08	0.02
	Covariances				
		PLS-EC with SLOPEEC	0.74	0.10	< .001
		PLS-EC with SLOPEAC	0.38	0.08	< .001
		PLS-EC with SLOPEPPVT	1.37	0.33	< .001
		PLS-AC with SLOPEAC	0.32	0.07	< .001
		PLS-AC with SLOPEEC	0.56	0.09	< .001
		PLS-AC with SLOPEPPVT	1.18	0.29	< .001
		PPVT with SLOPEPPVT	5.09	0.90	< .001
		PPVT with SLOPEAC	0.82	0.23	< .001
		PPVT with SLOPEEC	1.61	0.31	< .001
		SLOPEAC with SLOPEEC	0.02	0.01	0.007
		SLOPEAC with SLOPEPPVT	0.02	0.02	0.116
		SLOPEAC with SLOPEPPVT	0.05	0.02	0.004
		PLS-EC with PLS-AC	19.14	2.07	< .001
		PLS-EC with PPVT	54.09	6.24	< .001
		PLS-AC with PPVT	47.71	5.35	< .001
Response	Residual variances				
		PLS-EC	8.94	0.64	< .001
		PLS-AC	9.25	0.71	< .001
		PPVT	105.44	11.60	< .001
	Residual covariances				
		PLS-AC with PLS-EC	1.50	0.47	0.00
		PPVT with PLS-EC	2.64	1.46	0.07
		PPVT with PLS-AC	3.39	1.54	0.03

Appendix B

Table 8. Parameter estimates for the Method Factor Model

	Parameter	Estimate	*SE*	*p*
	Person Level			
Means				
	PPVT	52.84	0.77	< .001
	PLS	43.00	0.26	< .001
	SLOPEPLS	0.66	0.01	< .001
	SLOPEPPVT	1.80	0.05	< .001
Variances				
	PPVT	180.15	116.89	.123
	PLS	19.71	20.38	.334
	SLOPEPPVT	0.16	0.04	< .001
	SLOPEPLS	0.01	0.01	.005
Covariances				
	PPVT with SLOPEPPVT	4.99	2.10	.02
	PPVT with PLS	52.20	48.73	.28
	PPVT with SLOPEPLS	1.10	0.83	.19
	PLS with SLOPEPLS	0.45	0.33	.18
	PLS with SLOPEPPVT	1.22	0.85	.15
	SLOEP PLS with SLOPEPPVT	0.03	0.01	.02
Loadings				
	PLS by PLS_EC	1	0	—
	PLS by PLS_AC	1	0	—
	Response Level			
Residual variances				
	PPVT	105.93	10.22	< .001
	PLS_AC	9.48	0.71	< .001
	PLS_EC	11.37	1.16	< .001
Residual covariances				
	PLS-AC with PPVT	3.12	1.39	.02
	PLS-EC with PPVT	3.10	1.67	.06
	PLS-EC with PLS_AC	0.57	0.62	.36

Note. Dashes indicate a parameter fixed for model identification.