Oral Language and Listening Comprehension: Same or Different Constructs?

Language and Reading Research Consortium (LARRC)

Accepted for publication in Journal of Speech, Language and Hearing Research.

Author Note

This paper was prepared by a Task Force of the Language and Reading Research Consortium (LARRC) consisting of Shelley Gray (Convener), Hugh Catts, Jessica Logan, and Jill Pentimonti. LARRC project sites and investigators are as follows:

Ohio State University (Columbus, OH): Laura M. Justice (Site PI), Richard Lomax, Ann O'Connell, Jill Pentimonti[1], Stephen A. Petrill[2], Shayne B. Piasta

Arizona State University (Tempe, AZ): Shelley Gray (Site PI), Maria Adelaida Restrepo.

Lancaster University (Lancaster, UK): Kate Cain (Site PI).

University of Kansas (Lawrence, KS): Hugh Catts[3] (Site PI), Mindy Bridges, Diane Nielsen.

University of Nebraska-Lincoln (Lincoln, NE): Tiffany Hogan[4] (Site PI), Jim Bovaird, J. Ron Nelson.[5]

[1] Jill Pentimonti is now at American Institutes for Research.

[2] Stephen A Petrill was a LARRC co-investigator from 2010-2013.

[3] Hugh Catts is now at Florida State University.

[4] Tiffany Hogan is now at MGH Institute of Health Professions.

[5] J. Ron Nelson was a LARRC co-investigator from 2010-2012.

This work was supported by grant # R305F100002 of the Institute of Education Sciences' Reading for Understanding Initiative. We are deeply grateful to the numerous staff, research associates, school administrators, teachers, children, and families who participated. Key personnel at study sites include: Lisa Baldwin-Skinner, Garey Berry, Beau Bevens, Jennifer Bostic, Shara Brinkley, Janet Capps, Beth Chandler, Lori Chleborad, Willa Cree, Dawn Davis, Michel Eltschinger, Kelly Farquharson, Tamarine Foreman, Rashaun Geter, Sara Gilliam, Miki Herman, Trudy Kuo, Gustavo Lujan, Junko Maekawa, Carol Mesa, Denise Meyer, Maria Moratto, Kimberly Murphy, Marcie Mutters, Amy Pratt, Trevor Rey, Amber Sherman, Shannon Tierney, and Stephanie Williams.

The views presented in this work do not represent those of the federal government, nor do they endorse any products or findings presented herein. Correspondence concerning this work should be sent to Shelley Gray, Department of Speech and Hearing Science, PO Box 870102, Tempe, AZ 85287-0102 Shelley.Gray@asu.edu.

Abstract

Purpose: The purpose of this study was to add to our understanding of the dimensionality of oral language in children and to determine whether oral language and listening comprehension are separate constructs in children enrolled in preschool (PK) through third grade.

Method: In the spring of the school year children from four states (n=1,869) completed multiple measures of oral language (i.e., expressive and receptive vocabulary and grammar) and listening comprehension as part of a larger study of the language bases of reading comprehension.

Results: Initial confirmatory factor analysis found evidence that measures of oral language and listening comprehension loaded on two separate factors in preschool through 3rd grade; however, these factors were highly correlated at all grades.

Conclusion: These results suggest that oral language and listening comprehension are best characterized as a single oral language construct in grades PK through 3. The implications for early identification and intervention are discussed.

Oral Language and Listening Comprehension: Same or Different Constructs?

The question of whether oral language and listening comprehension are different constructs follows from oral language and reading research where terms such as ‘oral language comprehension,’ ‘linguistic comprehension,’ ‘verbal comprehension,’ ‘story comprehension,’ ‘comprehension of spoken text,’ and ‘listening comprehension’ are often used interchangeably.

Nearly 30 years ago in their seminal text on language development and disorders, Bloom and Lahey (1978) described language as encompassing form (grammar and morphology), content (semantics), and use (pragmatics). There has been general agreement among oral language researchers and clinicians about these structural components of language, but less focus or consensus on the construct of listening comprehension. Oral language researchers and clinicians tend to think of listening comprehension as the construct presented in norm-referenced oral language tests by the same name; however, the content of listening comprehension measures varies substantially across tests. For example, for the ‘Understanding Spoken Paragraphs’ subtest of the *Clinical Evaluation of Language Fundamentals – 5th Edition* (CELF-5; Wiig, Semel, & Secord, 2015), the examiner reads a paragraph to the child then the child answers questions about the paragraph’s main idea, details, sequencing, and inferential information. Most would agree this is an assessment of listening comprehension. In contrast, the Listening Comprehension subtest of the *Oral and Written Language Scales – Second Edition* (OWLS-II; Carrow-Woolfolk, 2011) assesses children’s understanding of single words, phrases, and sentences using a picture pointing task.

Meanwhile reading researchers have focused on listening comprehension because of its central role in reading comprehension. According to the Simple View of Reading theoretical framework (Gough & Tunmer, 1986; Hoover & Gough, 1990), listening (or linguistic) comprehension refers to *comprehension of written text read out loud*. According to this definition, Understanding Spoken Paragraphs from the CELF-5 would be a measure of listening comprehension, but the Listening Comprehension subtest from the OWLS-II would not.

Thus it is clear, but not surprising, that researchers interested in listening comprehension do not agree on the basic construct. Some propose that oral language contributes to listening comprehension, some that listening comprehension is part of a larger construct of oral language, and some that oral language and listening comprehension are separate constructs. For example, several contemporary researchers describe oral language skills as essential building blocks for the construct of listening comprehension. In their study of the role of inference making and oral language skills in narrative listening comprehension, Lepola, Lynch, Laakkonen, Silven, and Niemi (2012) stated that, “We still have much to learn about the development of the individual skills necessary for narrative listening comprehension (e.g., inference making, vocabulary), how these skills may influence each other across time, and how they become integrated to produce skilled listening comprehension” (p. 260). Similarly, in their article on the importance of listening comprehension, Hogan, Adlof, and Alonzo (2014) described vocabulary, inferencing, and background knowledge as influencing listening comprehension.

Other researchers place listening comprehension within the construct of oral language or use oral language components, in lieu of listening comprehension, as predictors of reading comprehension. For example, in their study of skills predicting reading comprehension in elementary school, Kendeou, van den Broek, White, and Lynch (2009) included listening comprehension, television comprehension, and vocabulary under the category of oral language skills and in a different study investigating the structure of oral language and reading in relation to comprehension in grades K-2, Foorman, Herrera, Patscher, Mitchell, and Truckenmiller

(2015) included vocabulary, syntax, and listening comprehension in their oral language construct. Similarly, Catts, Herrera, Nielson, and Bridges (2015) included word-reading precursors (letter knowledge, phonological awareness, rapid naming), word reading, receptive and expressive vocabulary, and narrative comprehension and production measures in kindergarten to predict reading comprehension in third grade. They included the vocabulary and narrative tasks in their oral language factor, and found that along with other measures, the oral language factor predicted unique variance in subsequent reading comprehension.

Finally, some researchers define components of oral language and listening comprehension as separate constructs in their literacy models (e.g. Juel, Griffith, & Gough, 1986; Kim & Phillips, 2014). In these studies a distinction is drawn between children's performance on component oral language skill measures, such as vocabulary, and performance on listening comprehension measures that test understanding of aurally presented sentences or texts.

Empirical Studies of the Dimensionality of Oral Language

Although there are relatively few studies of the dimensionality of oral language, three studies suggest either that oral language is unidimensional into adolescence, or that it is unidimensional in young children (prior to first grade), with multidimensionality emerging as children progress through school. These studies have not included measures of listening comprehension.

In a longitudinal study of school-age English-speaking children who were tested on receptive and expressive vocabulary and grammar as they progressed through kindergarten, second, fourth, and eighth grades, Tomblin and Zhang (2006) found that a two-factor model best fit the data at all grade levels, especially in eighth grade; however, the authors argued that the fit for the one- and two-factor models was so close in kindergarten, second, and fourth grades (vocabulary and grammar factors were correlated at *r* = .94, .93, .90, and .78, respectively), that the most parsimonious interpretation of results was that a 'general language trait' (p. 1206) underpins language ability throughout elementary school.

In their longitudinal study evaluating the stability of oral language in English-speaking children at ages 20 months, 4, 10, and 14 years, Bornstein, Hahn, Putnick, and Suwalsky (2014) found that multiple measures of oral language, including language sample analyses, maternal reports from the Vineland Adaptive Behavior Scales Communication Domain (Sparrow, Balla, & Cicchetti, 1984), and verbal subtests from either the Wechsler Preschool and Primary Scale of Intelligence – Revised (Wechsler, 1989) (age 4) or the Wechsler Intelligence Scale for Children, 3^{rd} Edition (Wechsler, 1991) (ages 10 and 14), consistently loaded onto the same single core language factor. The authors interpreted their results as evidence for a single robust and stable core language skill from the end of infancy through adolescence.

A similar conclusion was drawn in a multilevel, cross-sectional study of 529 children with typical development who were enrolled in 85 different preschool classrooms (Anthony, Davis, Williams, & Anthony, 2014). Children completed receptive and expressive measures of vocabulary, grammar, and articulation. The authors found a single latent language ability factor at both the child and classroom levels that included receptive and expressive vocabulary and grammar. They also found separate speech perception and articulation factors that included the receptive and expressive articulation measures, respectively.

Taken together these three studies of the dimensionality of oral language indicate that oral language can best be characterized as a single construct. None of these studies included discourse level measures or measures of listening comprehension; therefore, we next consider studies that included these measures.

Empirical Studies of the Dimensionality of Oral Language including Discourse or Listening Comprehension

Two recent cross-sectional studies assessing the dimensionality of oral language were conducted by the Language and Reading Research Consortium (LARRC). The first involved 286 Spanish-English dual-language preschoolers who were assessed in Spanish the year prior to kindergarten entry (LARRC, 2015). Tests and experimental tasks assessed receptive and expressive vocabulary and grammar at the single word and sentence levels, as well as listening comprehension. In the listening comprehension tasks children were asked to answer comprehension questions after hearing paragraph level text read to them. A bifactor model with a single underlying language trait, plus two additional group traits of word knowledge and integrative language knowledge, best fit the data. The word knowledge factor included vocabulary and background knowledge measures. The integrative language knowledge factor included grammar, morphology, and listening comprehension measures. The finding of a single underlying language trait was consistent with Tomblin and Zhang's and Bornstein et al.'s proposal of a general language trait in children.

The second cross-sectional study included 915 English-speaking children enrolled in preschool (PK; *n*=420), kindergarten (K; *n*=124), first (G1; *n*=125), second (G2; *n*=123), and third grades (G3; *n*=123) (LARRC, in press). Children completed norm-referenced tests and experimental tasks designed to assess receptive and expressive vocabulary, grammar, and higher-level discourse skills including comprehension monitoring, understanding of narrative text structure, and inferencing. We tested whether oral language was a unitary construct or instead, was best represented by a two- (vocabulary/grammar, discourse) or three-factor model (vocabulary, grammar, discourse). Results of confirmatory factor analyses suggested that in PK and K, a one-dimensional model best fit the data. In G1 and G2, a two-factor model best fit the data, but there was substantial overlap between the vocabulary/grammar and discourse constructs in each grade (r^2=.72 and .64, respectively). In G3 there was evidence for further emergence of multidimensionality, with a three-factor model (vocabulary, grammar, discourse) best fitting the data.

The latter two studies illustrate the importance of using a wide array of language measures to evaluate the structure of oral language. By adding discourse level measures (which could be considered listening comprehension measures) to measures of vocabulary and grammar like those employed by Bornstein et al. (2014) and Tomblin and Zhang (2006), the LARRC studies (2015; in press) were able to show the unidimensionality of oral language prior to G1, consistent with previous studies, but also the emergence of separate vocabulary and grammar factors as children progressed from G1 to G2, then the emergence of an additional higher-level language discourse factor, akin to listening comprehension, in G3.

A recent cross-sectional study investigating the structure of oral language and reading in low-income English-speaking children enrolled in K, G1, and G2 evaluated five measurement models in K and four in G1 and G2 to determine whether various combinations of oral language components, including listening comprehension, syntax, vocabulary, and phonological awareness (along with decoding in G1 and G2), loaded on separate factors (Foorman, Herrera, Patscher, Mitchesll, & Truckenmiller, 2015). Results of confirmatory factor analyses indicated that in K, a single, second-order oral language factor consisting of listening comprehension, syntax, vocabulary, and phonological awareness provided the best model fit. In G1 and G2, listening comprehension, syntax, and vocabulary measures all loaded on a single oral language factor.

Similarly, in their study investigating components of the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990) with 488 Greek children enrolled in G3 through G5, Protopapas, Simos, Sideridis, and Mouzaki (2012) tested whether measures of receptive and expressive vocabulary, verbal instruction comprehension, and listening comprehension loaded on the same language factor. Their measure of listening comprehension included two narrative and one expository passage read to children, followed by four multiple-choice comprehension questions. They tested one- (vocabulary and listening comprehension) and two-factor models (vocabulary, listening comprehension) and found that the single factor model best fit the data.

Studies investigating the contribution of vocabulary to reading comprehension also report results that are largely consistent with a unitary view of oral language and listening comprehension in young elementary-age children. In their investigation of whether oral vocabulary explained the oral language contribution that listening comprehension makes to reading comprehension, Ouellette and Beers (2010) found that in G1, vocabulary depth and breadth did not contribute additional variance to reading comprehension over and above listening comprehension. In G6, however, vocabulary breadth accounted for variance above that explained by listening comprehension. A similar study by Tunmer and Chapman (2012) to test whether oral language comprehension and word recognition independently contributed to reading comprehension in G3 when vocabulary knowledge was included as a separate factor, reported that listening comprehension (.89) and receptive vocabulary (.89) each loaded highly on the same Linguistic Comprehension factor.

Together these studies have assessed components of oral language, including higher level discourse and listening comprehension, in different combinations, sometimes for the express purpose of investigating the dimensionality of oral language and sometimes to predict reading comprehension. No study has expressly tested whether oral language and listening comprehension are separate constructs when vocabulary, grammar, and listening comprehension are comprehensively assessed. Thus, the purpose of this study was to determine whether listening comprehension is part of a broader oral language construct that also includes vocabulary and grammar, or whether it is separable in children enrolled in PK through G3. Based on previous research assessing oral language discourse, we hypothesized that listening comprehension would not be separable from vocabulary and grammar in PK or K, but that we could see the emergence of a separate listening comprehension factor as children moved into G3.

Method

Participants

Participants were enrolled in a multi-site, five-year longitudinal research project conducted by LARRC. The purpose of the LARRC longitudinal study was to identify and model language processes important for reading comprehension in children enrolled in PK through G3. In the present study we utilized concurrent data from children across all grades (PK – G3) enrolled in the first three years of the study (2011-2014).

Children in each grade level were selected from four research sites in Arizona, Kansas, Nebraska, and Ohio. Across research sites school districts were selected based on size and diversity of their student populations, as well as willingness to participate in the project. Cooperating teachers in consenting districts received recruitment packets to send home for all students in their class. From among those children whose parents consented to participation, we randomly selected approximately equal numbers of children per grade level at each research site.

Demographic information for the sample is presented in Table 1. Overall seventy-eight percent of families reported speaking primarily English at home; other languages spoken at home included Spanish, Chinese, Amharic, and Vietnamese. Seventy percent of children resided in two-parent households. Nearly 10% of children had Individual Education Plans (IEPs) and 16% qualified for free/reduced lunch. Given the characteristics of our sample, caution should be taken when generalizing our results to minority populations or to children from families that speak languages other than English.

Procedures

Children completed a comprehensive assessment battery in the second half of the academic year during a 20-week window from January through May. Assessments were divided into 11 blocks approximately 30 minutes in duration. Typically one or two blocks were administered per day. All measures were administered individually by trained assessors at each of the assessment sites. Assessors underwent comprehensive training and in-lab observations to ensure consistent administration and scoring procedures across sites. This training included the completion of on-line training modules (including quizzes) and direct observation by supervising assessors.

Measures

Our assessment battery included multiple measures of vocabulary, grammar, and listening comprehension. Different measures were appropriate for different age ranges; thus, all children did not complete the same measures. Table 2 lists raw scores on measures administered at each grade level, indicating which measures were post-scored in the lab from audio recordings to insure high scoring reliability. Inter-rater reliability was acceptable for all post-scored measures, with Intra-class correlations (ICC's) ranging from .74 to .99 as calculated for the sample from Year 1 of the longitudinal study. Each of the measures is described below.

Grammar. Six measures were used to assess grammar. The Word Structure subtest of the Clinical Evaluation of Language Fundamentals - 4th Edition (CELF-4; Semel, Wiig, & Secord, 2003) assessed children's ability to apply word structure rules to indicate inflections, derivations, and to select appropriate pronouns to refer to people, objects, and possessive relationships. The Recalling Sentences subtest of the CELF-4 assessed children's ability to listen to spoken sentences of increasing length and complexity and repeat those sentences without changing word meanings, inflections, derivations, or sentence structure. The Past Tense Probe of the Rice/Wexler Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001) assessed children's production of regular and irregular past tense verbs and the Third Person Singular Probe of the TEGI assessed children's abilities to produce /-s/ or /-z/ in present tense verb forms with singular subjects, and was administered to children in PreK and K only (see Table 2). The Test for Reception of Grammar – Version 2 (TROG-2; Bishop, 2003) assessed children's grammatical comprehension and understanding of English grammatical contrasts marked by inflections, function words, and word order. Finally, the Morphological Derivation task described by Spencer and colleagues (2015) assessed children's knowledge of derivational morphology. For this measure of grammar, the assessor presented children with a base word (e.g., farm) and an incomplete sentence for which children provided a derived form of the base word (e.g., My uncle is a ______).

Vocabulary. Three measures were used to assess vocabulary. The Peabody Picture Vocabulary – Fourth Edition (PPVT-4; Dunn & Dunn, 2007) assessed children's recognition of the meaning of spoken words. The Expressive Vocabulary Test – Second Edition (EVT-2;

Williams, 2007) assessed children's expressive vocabulary. The Word Classes receptive and expressive subtests from the CELF-4 assessed children's ability to understand relationships between words related by semantic class features and to verbally express the similarities and differences concerning those relationships. The receptive and expressive subtest scores were combined into one vocabulary score.

Listening comprehension. The Listening Comprehension Measure (LCM) was administered to assess children's ability to listen to, comprehend, and answer inferential and non-inferential questions about spoken narrative and expository passages. This experimental measure consisted of seven narrative passages and questions, with some modifications, taken from the Qualitative Reading Inventory – Fifth Edition (QRI-5; Leslie & Caldwell, 2011), and seven new passages and questions, one narrative and six expository, created specifically for the project. All passages adhered to appropriate length and lexile level for each grade. Participants listened to passages (one expository and two narrative passages for PK - G1, and two of each type for G2 - G3) and then answered between 4 and 8 open-ended implicit and explicit questions. Children's responses were audio recorded and post-scored.

A researcher-developed measure, the Inference Task, based on work by Cain and Oakhill (1999) and Oakhill and Cain (2012), was used to assess a child's ability to construct a mental model of a passage read to them. It evaluated the child's ability to make two type of inferences: integration between sentences in a story and integration between story information and general knowledge. In this task children listened to two narrative passages read aloud and were asked a series of inference-based questions. Children's responses were audio recorded and post-scored.

Data Preparation. Data were obtained separately for each grade, and the tests used to measure constructs varied by grade. Therefore all latent models described in the next section were estimated separately for each grade. Prior to entry in the latent models, the data for each grade was analyzed and determined to be missing completely at random for all grades with the exception of grade 1 (Little's MCAR test chi-squared values all >158, *DF*> 125, *p*s > .11). Missing data was accounted for in latent models using Restricted Maximum Likelihood Estimation in Mplus v7.11 (Muthén & Muthén, 1998-2012).

Results

The primary aim of the present study was to determine whether oral language (receptive and expressive vocabulary and syntax) and listening comprehension are unique constructs in children enrolled in grades PK through G3. We addressed this by examining two models to determine the best conceptualization of constructs across grades PK to G3. Specifically, we fitted a taxonomy of latent-variable models allowing for competing configurations of constructs. These confirmatory factor analyses (CFA) were conducted in MPlus v7.11 (Muthén & Muthén, 1998-2012) using maximum likelihood estimation with robust standard errors (MLR) to adjust for non-independence within classrooms and slight non-normality of the data (note that analyses were also conducted in a two-level CFA with minimal to no changes in factor loadings, factor correlations, or subsequent conclusions). MLR also allowed for the estimation of factor solutions for all cases, even those with missing data. Two models were compared at each grade for quality of fit: a one factor model included oral language and listening comprehension measures, whereas the two-factor included an oral language factor composed of vocabulary and grammar measures and a listening comprehension factor.

For all model comparisons made within each grade, less complex models (i.e., one factor) were always estimated as constrained versions of the more complex models, such that the models can be considered to be nested. This was accomplished by constraining factor correlations to 1.0

between factors that are collapsed in less complex models. This method allowed for comparative model fit assessed using a chi-square difference test. We present a graphical representation of the preschool model in Figure 1. The models for kindergarten, G1, G2, and G3 were very similar (see Appendices A-E). A few observed measures changed across the grades (see Table 1).

The fit of each model was measured with four different static fit indices. First, the Root Mean Square Error of Approximation (RMSEA), which ranges from 0 to 1; values less than .08 suggest that model fits the data well (Browne & Cudeck, 1993; MacCallum & Austin, 2000). Second was the Standardized Root Mean Square Residual (SRMR), which follows a similar cutoff to the RMSEA, such that good model fit is indicated by a SRMR value smaller than .05 (Byrne, 2012). Also included were the Comparative Fit and Tucker-Lewis Indices (CFI; TLI), for which values greater than .90 indicate good model fit (Hu & Bentler, 1999; Lomax, 2013).

In addition to the four static indices of model fit, three comparative fit indices were examined. The Akaike's Information Criterion (AIC) and Bayesian Information Criteria (BIC) where in both cases smaller values indicate better model fit (Kline, 2005). Next, for models with multiple factors, the correlations between factors were examined. Brown and Moore (2014) note that when factor correlations are close to 1.0, it is questionable as to whether the factors are representative of separate constructs. Therefore both Brown (2006) and Brown and Moore (2014) recommend that highly-correlated dimensions be collapsed into a single factor. For our purposes, we determined a-priori that factors correlated above .90 could be considered to be highly correlated and would be used as evidence that the two correlated factors were not distinct. Lastly, we tested the significance of changes in model fit between nested models using a chi-square difference test. If the test is statistically significant, it indicates that the increase in model fit is sufficiently large to warrant the decrease in degrees of freedom sacrificed to fit the model, and that the more complex model should be retained. Finally, when model fit was unclear an additional fit index was examined: Moran, Marsh, and Nagengast (2013) suggest that when comparing two nested models, the more parsimonious model should be retained if the incremental change in the CFI is .01 or smaller. A similar index has been suggested for the RMSEA (Chen, 2007), where a change less that .015 is considered unimportant. No one fit index was weighted more heavily when determining which model to retain. Instead all indices are reported and were considered simultaneously (e.g., Lomax, 2013; Mueller & Hancock, 2010).

Model Fit Results

Model fit and nested model comparisons for one- and two- factor models are presented in Table 3. The one-factor model included oral language and listening comprehension measures, whereas the two-factor included an oral language factor composed of vocabulary and grammar measures and a listening comprehension factor. The results across all grades were very similar. The absolute fit statistics were slightly better for the two-factor model: RMSEA and SRMR were slightly smaller, and CFI and TLI were slightly larger. For the comparative fit indices, AIC and BIC were slightly smaller for the two-factor model, and the chi-square difference test suggested that the two factor model fit significantly better than the one-factor model (delta chi square significance test < .003 for all grades). In PreK and K, the observed difference between the one- and two-factor models in their RMSEA (.023 and .027) and CFI (.022 and .024) were slightly larger the identified cutpoints for meaningful fit differences (Δ RMSEA = 0.015; Chen, 2007 and Δ CFI = 0.01; Moran et al., 2013), whereas in grades 1-3 the model differences were minimal. Finally, the correlation between the two factors was very strong in all grades (ranging from 0.87 to 0.91), suggesting that the two factors are not necessarily unique (Brown, 2006). Therefore, the fit indices are split in terms of which model should be considered the best fit to the data across grades, but with slightly less evidence in G1 – G3 for the uniqueness of language and listening comprehension. Because the chi-square difference test can be biased towards significance when sample sizes are large (Tomarken & Waller, 2003), we therefore conclude that listening comprehension is not unique from oral language at any grade.

Discussion

The purpose of this study was to determine if listening comprehension, vocabulary, and grammar are part of a single language construct, or whether they are separable constructs in children enrolled in PK through G3. In recent work, particularly that related to reading comprehension, some researchers have made a distinction between these aspects of oral language (Braze, Tabor, Shankweiler, & Mencl, 2007), while others have treated them as if they were the same construct (Catts et al., in press; Kendeou et al., 2009). Our results indicate that measures of oral language and listening comprehension appear to assess the same underlying construct. Analyses did show that two-factor models had slightly better fit than one-factor models. However, the factors in the two-factor models (i.e., oral language and listening comprehension) were highly correlated and thus may operate as a single construct (Brown, 2014).

It is important to note that our conceptualization of listening comprehension is the understanding of written discourse that has been read aloud. We framed listening comprehension in this manner rather than in terms of asking students to understand complex spoken sentences or multistep spoken directions. The latter approach seems more central to oral language and is often the approach taken to measure to language abilities in standardized assessments (e.g., Bishop, 2003). Our approach, on the other hand, has been used more often by reading researchers to assess the linguistic component of reading comprehension beyond word recognition (Gough & Tunmer, 1986; Hoover & Gough, 1990). Nevertheless, both approaches appear to be measuring the same underlying language abilities.

We anticipated that we would find evidence of a single language construct in PK/K, but that listening comprehension and an oral language construct including vocabulary and grammar might separate by G3. This hypothesis was largely based on the findings of our previous examination of the dimensionality of language (LARRC, in press). In that study we found distinct constructs for vocabulary, grammar, and discourse in G3 children. The discourse construct in that earlier study included an inferencing task similar to the listening comprehension measures in this study in that it asked children to answer open-ended questions about passages read to them. But the discourse measures in that study also included comprehension monitoring and text structure knowledge measures. It may be that the inclusion of comprehension monitoring and text structure measures was sufficient to identify a discourse construct separate from vocabulary and grammar. The former measures may be especially impacted by children's experience with literacy and thus show some separation from oral language measures involving vocabulary and syntax. However, the present results indicate that when discourse measures are restricted to those that specifically assess listening comprehension, these measures are better construed as being part of a larger oral language construct at all grade levels.

Our results are consistent with studies that have examined these measures in relationship to the simple view of reading. According to the simple view, reading comprehension is the product of word recognition and language comprehension (Gough & Tunmer, 1986). Measures of listening comprehension are most often used as a proxy for language comprehension. In some studies, however, measures of vocabulary and grammar are also considered. In one of the latter investigations, Braze et al. (2007) reported that measures of vocabulary accounted for variance in reading comprehension over and above that explained by listening comprehension. Such a result implies that vocabulary and listening comprehension may be partially distinct. However, follow-up studies have more often found that vocabulary and listening comprehension load on the same construct and jointly predict reading comprehension (Protopapas et al., 2012; Tunmer & Chapman, 2014). Specifically, Protopapas et al., (2012) reported that the systematic variance associated with their two measures of vocabulary completely overlapped with the variance shared by their two measures of listening comprehension. They suggested that any unique variance accounted for by vocabulary in other studies may be due to the fact that vocabulary measures are often more reliable than listening comprehension measures, and thus account for more systematic variance.

Our findings have implications for the early identification and intervention of language impairments, especially as they relate to problems in reading comprehension. As noted above, the simple view of reading posits that word recognition and listening comprehension are predictive of reading comprehension. A direct extension of this view is that measures of listening comprehension could be used in the early school grades to forecast subsequent problems in reading comprehension. However, listening comprehension measures are often lengthy and highly dependent on background knowledge. Given that other oral language measures may fall within the same construct as measures of listening comprehension, we may be able to use the former measures to more efficiently predict problems in reading comprehension. Many of these measures take less time to administer and are less dependent on background knowledge. For example, measures of vocabulary and sentence repetition have been shown to be unique predictors of subsequent reading comprehension (Catts, Fey, Zhang, & Tomblin, 2001; Catts, Nielsen, Bridges, Liu, in press). In fact, Catts et al. (in press) found that a measure of expressive vocabulary at the beginning of kindergarten performed as well as a lengthy test of narrative production and comprehension in the unique prediction of reading comprehension in G3. While it is not common to use measures of oral language in screening for risk for reading disabilities, these results suggest that measures of vocabulary may be an efficient way to identify language problems that may be related to later reading difficulties. A promising approach may be the use of computer adaptive tests of vocabulary. Recently, Foorman, Petscher, and Schatschneider (2015) used computer adaptive tests to quickly estimate language skills in an early screening battery.

Our results also raise interesting questions about early intervention for children at risk for language and reading disabilities and also about preschool and elementary language arts curricula. Given that oral language appears to operate as a single construct in the early elementary school years, would children at risk for language and reading difficulties benefit more from interventions that provide broad and rich language experiences versus instruction in a single component of language (e.g. syntax or morphology)? Similarly, would language arts curricula focused on school-based language proficiency, or what some have termed academic language (Schleppegrell, 2012; Uccelli et al., 2014), promote better language development and subsequent reading comprehension than more diverse approaches? These are testable hypotheses, generated by theoretical research, that inform next steps in reading comprehension research.

In conclusion, our study comprehensively assessed vocabulary, grammar, and listening comprehension. Our results suggest that components of oral language and listening comprehension are part of the same oral language construct in PK through G3. In the future if these results are replicated longitudinally, researchers should investigate whether it may be possible to more efficiently index children's oral language in the clinic and in research studies using a single measure of oral language.

References

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

Babayigit, S. (2014). The role of oral language skills in reading and listening comprehension of text: A comparison of monolingual (L1) and bilingual (L2) speakers of English language. *Journal of Research in Reading, 37,* S22-S47.

Baker, H. J., & Leland, B. (1967). *Detroit Tests of Learning Aptitude.* Indianapolis, IN: Bobbs-Merrill.

Berninger, V. W., & Abbott, R. D., (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 2, 5, and 7. *Journal of Educational Psychology, 102,* 635-651.

Bishop, D. V. (2003). *Test for Reception of Grammar (2nd ed*.). Minneapolis, MN: Pearson.

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

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

Braze, D., Tabor, W., Shankweiler, D., & Mencl, W.E. (Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, *40*, 226-243.

Brown, T.A. & Moore, M.T. (2014). Confirmatory Factor Analysis. In Rick H. Hoyle (Ed.), Handbook of Structural Equation Modeling (pp. 361-379). New York, NY: Guilford Press.

Carrow-Woolfolk, E. (2011). Oral and Written Language Scales – Second Edition. Torrence, CA: Western Psychological Services.

Catts, H., Fey, M.E., Zhang, X., & Tomblin, J.B. (2001). Estimating risk for future reading difficulties in kindergarten children: A research-based model and its clinical implications. *Language, Speech, and Hearing Services in Schools, 32,* 38-50.

Catts, H. W., Herrara, S., Nielsen, D., & Bridges, M. (2015). Early prediction of reading comprehension within a simple view framework. *Reading and Writing: An Interdisciplinary Journal*, 28, 1407-1425

Catts, H.W., Nielsen, D., Bridges, M., & Liu, Y. (2015). Early identification of reading comprehension difficulties. *Journal of Learning Disabilities, 48*, 281-297.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, *14*(3), 464-504.

Colvin, R. J., & Root, J. H. (1976). Reading Evaluation – Adult Diagnosis (READ). Syracuse, NY: Literacy Volunteers of America.

Curtis, M. E., (1980). Development of components of reading skill. *Journal of Educational Psychology*, *72*, 656-669. http://dx.doi.org/10.1037/0022-0663.72.5.656

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

Florida Department of Education. (2009). Florida Assessments for Instruction in Reading K-2 Technical Manual. Tallahassee, FL: Author. Retrieved October 22, 2014, from http://www.fcrr.org/FAIR/Final_K-2_Technical%20Manual_2010.pdf

Florida Department of Education. (2009–2013). Florida Assessments for Instruction in Reading (FAIR). Tallahassee, FL: Author.

Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in Kindergarten through Grade 2. *Reading and Writing*, *28,* 655-681.

Gold, P. C. & Johnson, J. A. (1982). Entry level achievement characteristics of youth and adults reading below fifth grade equivalent: A preliminary profile and analysis. *Psychological Reports, 50,* 1011-1019. doi: 10.2466/pr0.1982.50.3.1011

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10.

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance on listening comprehension. *International Journal of Speech-Language Pathology, 16,* 199-207. DOI: 10.3109/17549507.2014.904441

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160.

Horowitz, R., & Jay, S. S. (1985). Reading and listening to expository text. *Journal of Reading Behavior, 17,* 185-198.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78,* 243-255.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*, 765-778.

Kim, Y. & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly, 49,* 269-281.

King, W. H. (1959). An experimental investigation of the relative merits of listening and reading comprehension for boys and girls of primary school age. *British Journal of Educational Psychology, 29,* 42-49.

Language and Reading Research Consortium (2015). The dimensionality of language ability in young children. *Child Development*, 86, 1948-1965.

Language and Reading Research Consortium (2015). The dimensionality of language in young Spanish-English dual language learners. *Journal of Speech, Language, and Hearing Research*, 1-13, DOI: 10.1044/2015_JSLHR-L-13-0266.

Lepola, J., Lynch, J., Laakkonen, E., Silven, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4-6-year-old children. *Reading Research Quarterly*, 47, *259-282. DOI:10.1002/RRQ.020*

Lloyd, J., Cullinan, D., Heins, E. D., & Epstein, M. H. (1980). Direct instruction: Effects on oral and written language comprehension. *Learning Disability Quarterly, 3,* 77-76.

Marlowe, W., Egner, K., & Foreman, D. (1979). Story comprehension as a function of modality and reading ability. *Journal of Learning Disabilities*, *12,* 194-197. http://dx.doi.org/10.1177/002221947901200316

Nelson, N. W., Plante, E., Helm-Estabrooks, N., & Hotz, G. (2015). Test of Integrated Language & Literacy Skills. Baltimore, MD: Paul H. Brookes Publishing Company.

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, *23,* 189-208.

Protopapas, A., Simos, P. G., Sideridis, G. D., & Mouzaki, A. (2012). The components of the simple view of reading: A confirmatory factor analysis. *Reading Psychology, 33*, 217-240. DOI: 10.1080/02702711.2010.507626

Rice, M. L., & Wexler, K. (2001). *Rice/Wexler Test of Early Grammatical Impairment*. San Antonia, TX: The Psychological Corporation.

Schleppegrell, M.J. (2012). Academic language in teaching and learning: Introduction to special issue. *Elementary School Journal, 112*, 409-418.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals (4th ed.*). Bloomington, MN: Pearson.

Spache, G. D. *Diagnostic Reading Scales*. Monterey, Calif.: California Test Bureau, 1963.

Sparrow, SS.; Balla, DA.; Cicchetti, DV. (1984). *Vineland Adaptive Behavior Scales Survey Form Manual* (Interview Edition). American Guidance Service; Circle Pines, MN.

Sternberg, R. J., & Powell, J. S., (1883). Comprehending verbal comprehension. *American Psychologist, 38*, 878-893. http://dx.doi.org/10.1037/0003-066X.38.8.878

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

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities, 45*, 453-466. DOI: 10.1177/0022219411432685

Uccelli, P., Barr, C., Dobbs, C., Galloway, E.P., Meneses, A., & Sanchez, E. (2014). Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics, 35*, 1-33.

Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., Tighe, E. L., & Bishop, M. D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing, 28,* 959-988.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence--Revised*: Manual. The Psychological Corporation; San Antonio, TX.

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children—Third Edition*. The Psychological Corporation; San Antonio, TX.

Wechsler, D. (2001). *Wechsler Individual Achievement Test* (2nd ed.). San Antonio, TX: Psychological Corporation.

Wilkinson, A. C., (1980). Children's understanding in reading and listening. *Journal of Educational Psychology, 72*, 561-574. http://dx.doi.org/10.1037/0022-0663.72.4.561

Wiig, E., Semel, E., & Secord, W. (2015). *Clinical Evaluation of Language Fundamentals – Fifth Edition*. Bloomington, MH: Pearson.

Williams, K. T. (2007). *Expressive vocabulary test (2nd ed.*). Bloomington, MN: Pearson.

Table 1

Participant Demographics

			Race						Ethnicity		
Grade	N	Female	American Indian	Asian	Black	Other Pacific Islander	White	Not Reported	Hispanic or Latino	Not Hispanic or Latino	Not Reported
PK	420	41.4	2.1	3.8	5.0	0.2	92.4	1.0	9.3	87.9	3.1
K	509	44.4	2.4	4.3	6.9	0.2	90.4	0.8	9.8	87.8	2.4
1	593	48.6	2.5	4.9	7.3	0.2	87.4	3.7	9.4	86.0	1.9
2	366	50.8	1.6	4.9	9.0	0.3	83.1	6.3	10.7	82.5	6.8
3	371	52.3	1.3	6.7	8.9	0.3	81.9	6.7	8.1	84.6	7.3

Note. Percentages for Race may sum to more than 100% because parent could select more than one.

Table 2

Means (SDs) for Measures by Grade and Cronbach's Alpha Range by Measure

	PK	K	G1	G2	G3	Cronbach's Alpha (Range across Grades)
Grammar						
CELF-4 Word Structure Raw Score*	15.59 (5.56)	20.54 (4.59)	24.05 (4.09)	26.59 (3.39)	28.04 (2.78)	.63 - .83
CELF-4 Recalling Sentences Raw Score		43.14 (14.21)	52.08 (14.66)	60.86 (13.15)	66.74 (13.82)	.91 - .92
TEGI – Past Tense Raw Score	8.55 (4.32)	10.25 (4.45)				.86 - .89
TEGI – Third Person Singular Raw Score	7.00 (2.93)	7.88 (2.62)				.84 - .85
TROG Standard Score	6.25 (3.76)	10.75 (3.91)	13.74 (3.41)	15.29 (3.23)	16.41 (2.65)	.77 - .86
Morphological Derivation Raw Score			9.12 (4.34)	13.21 (5.05)	16.87 (4.97)	.78 - .84
Vocabulary						
PPVT-4	93.80 (19.29)	113.30 (17.57)	127.90 (16.20)	140.05 (16.28)	152.71 (16.79)	.95 - .96
EVT-2	70.06 (13.78)	84.39 (12.98)	95.71 (13.24)	106.40 (13.78)	116.31 (14.60)	.93 - .95
CELF-4 Word Classes Receptive and Expressive Total*		30.23 (6.92)	34.18 (4.10)	36.31 (3.16)	17.95^ (5.54)	.84 - .95
Listening Comprehension						
Test of Narrative Language Receptive*	15.58 (6.90)	23.04 (5.74)	26.92 (4.54)	29.76 (4.02)	31.50 (3.43)	.58 - .87
Listening Comprehension Measure*	7.02 (3.23)	9.84 (2.70)	11.71 (2.70)	19.66 (4.62)	21.51 (4.79)	.66 - .83

Inference Task*	0.83 (0.42)	1.12 (0.37)	1.18 (0.35)	1.88 (1.20)	1.53 (0.32)	.64 - .78

Note. All means reported are for raw scores. * = Post-scored measure. CELF-4 = Clinical Evaluation of Language Fundamentals – Fourth Edition (Semel, Wiig, & Secord, 2003); TEGI = Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001); TROG = Test for Reception of Grammar – Second Edition (TROG-2; Bishop, 2003); PPVT-4 = Peabody Picture Vocabulary Test – Fourth Edition (Dunn & Dunn, 2007); EVT-2 = Expressive Vocabulary Test – Second Edition (Williams, 2007). ^ = Only receptive measure was administered.

Table 3

Model Fit for PK, K, G1, G2, and G3

	PK *n* = 420		K *n* = 500		G1 *n* =587		G2 *n*=362		G3 *n*=365	
	Two-factor	One-Factor	Two-factor	One-Factor	Two-factor	One-Factor	Two-factor	One-Factor	Two-factor	One-Factor
Static Indicators										
RMSEA	0.058	0.081	0.048	0.075	0.041	0.053	0.053	0.064	0.063	0.078
SRMR	0.027	0.034	0.027	0.034	0.020	0.024	0.027	0.030	0.034	0.038
CFI	0.979	0.957	0.983	0.959	0.989	0.982	0.982	0.972	0.975	0.961
TLI	0.971	0.943	0.978	0.947	0.985	0.976	0.975	0.963	0.966	0.948
Comparative Indicators										
AIC	22045	22090	29251	29321	30606	30628	18930	18947	19041	19067
BIC	22174	22216	29398	29464	30746	30764	19055	19068	19166	19188
-2LL	-10990	-11014	-14590	-14626	-15271	-15283	-9433	-9443	-9489	-9503
Free Parameters	32	31	35	34	32	31	32	31	32	31
Language with LC ρ	0.91*	1[a]	0.89*	1[a]	0.91*	1[a]	0.91*	1[a]	0.87*	1[a]
Chi Square Δ test		*p* <.001		*p* <.001		*p* =.001		*p*=.002		*p* <.001
ΔRMSEA value		0.023		0.027		0.012		0.011		0.015
ΔCFI value		0.022		0.024		0.007		0.010		0.014

Note: * = Correlation is significantly different from zero, $p < .05$. [a] = path fixed at 1.0. Rules of thumb for static indicators: CFI and TLI > .90, RMSEA < .08, SRMR < .05. Smaller AIC and BIC indicate better fit. -2LL values closer to zero indicate better fit. Language with LC ρ is the size of the standardized correlation between the factors (Figure 1, path "a"). Chi Square Δ test, a significant result favors the more complex model. A Δ RMSEA value .015 or smaller and Δ CFI value .01 or smaller suggests the more parsimonious model is a better fit.

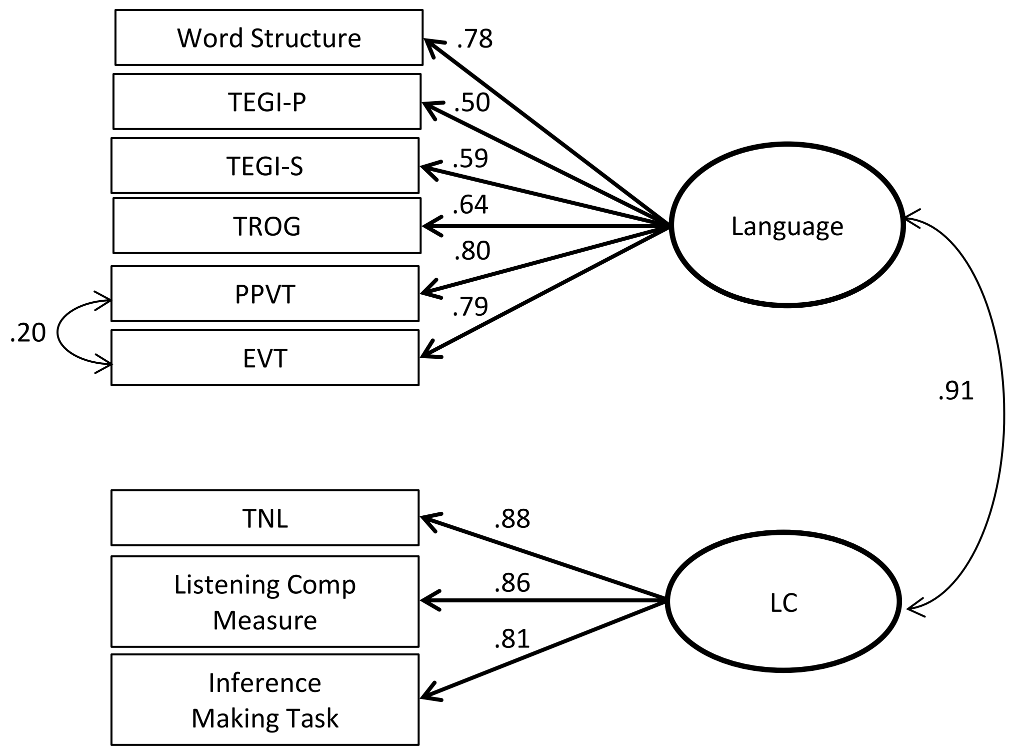
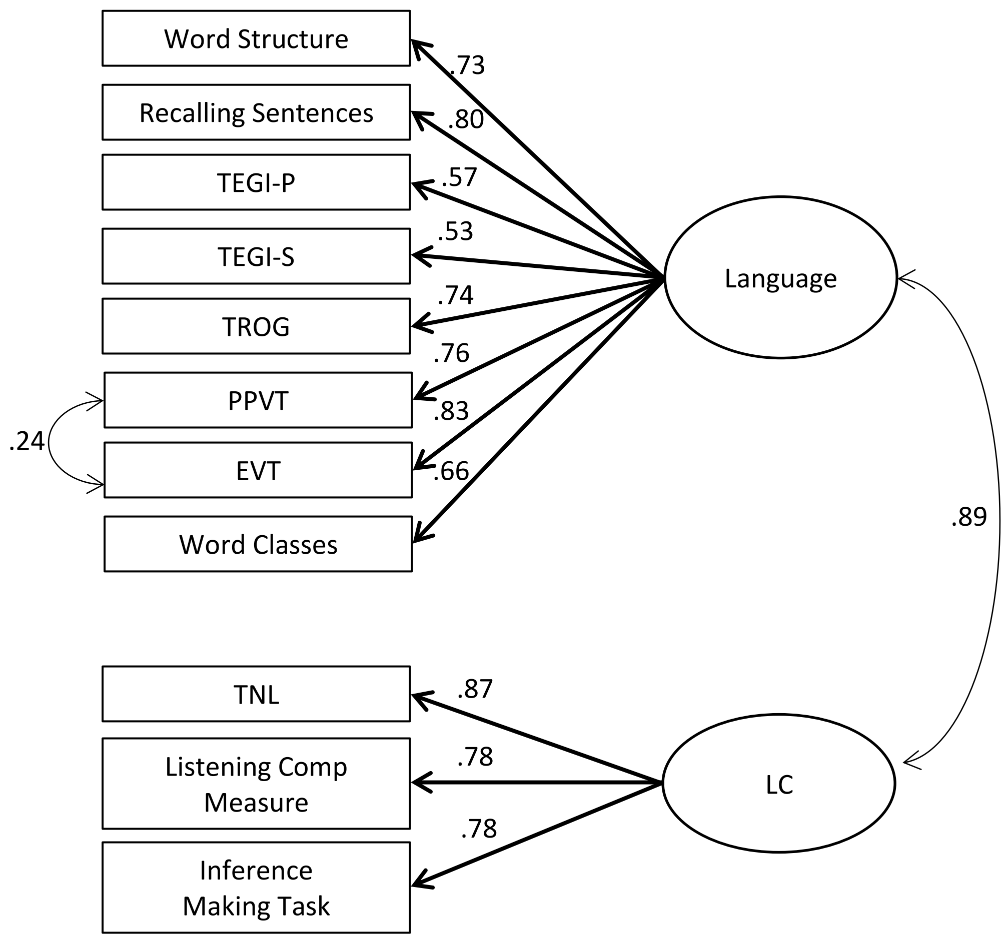


Figure 1. Structural model for PK, with similar models fitted for K, G1, G2, and G3 shown in Appendices. In the two-factor model the correlation between Language and Listening Comprehension (LC) is freely estimated. In the one-factor model, it is constrained to 1.0. Observed indicators are described in the measures section. Not all assessments were given at each grade, see Table 2.

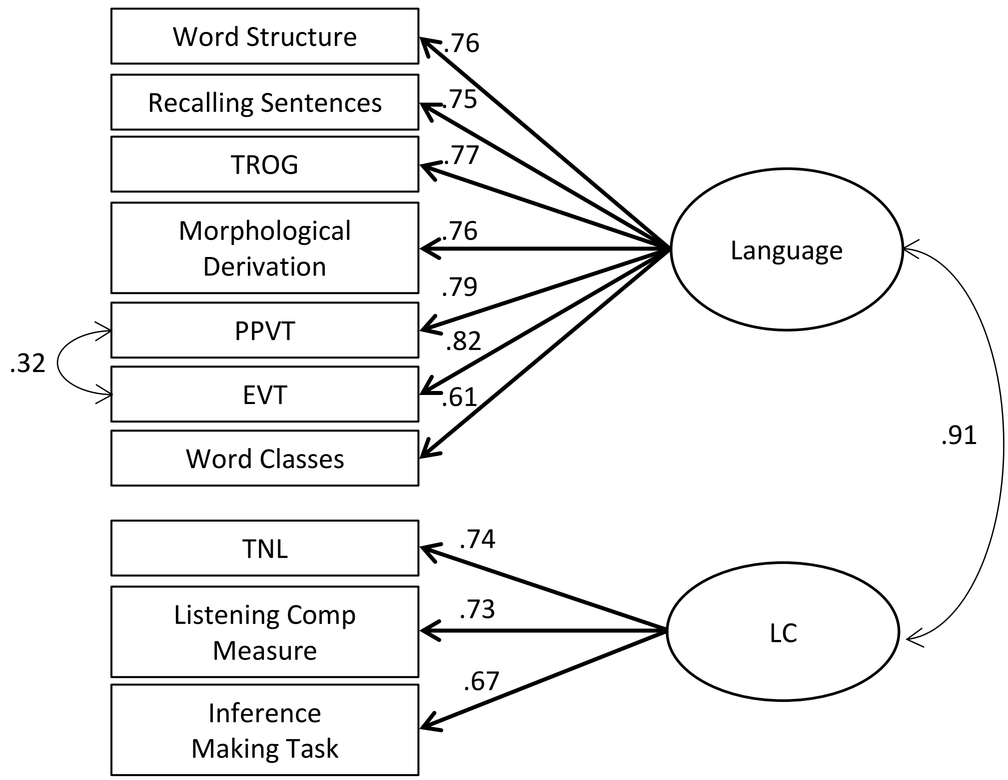
Appendix A

Structural Model for Kindergarten



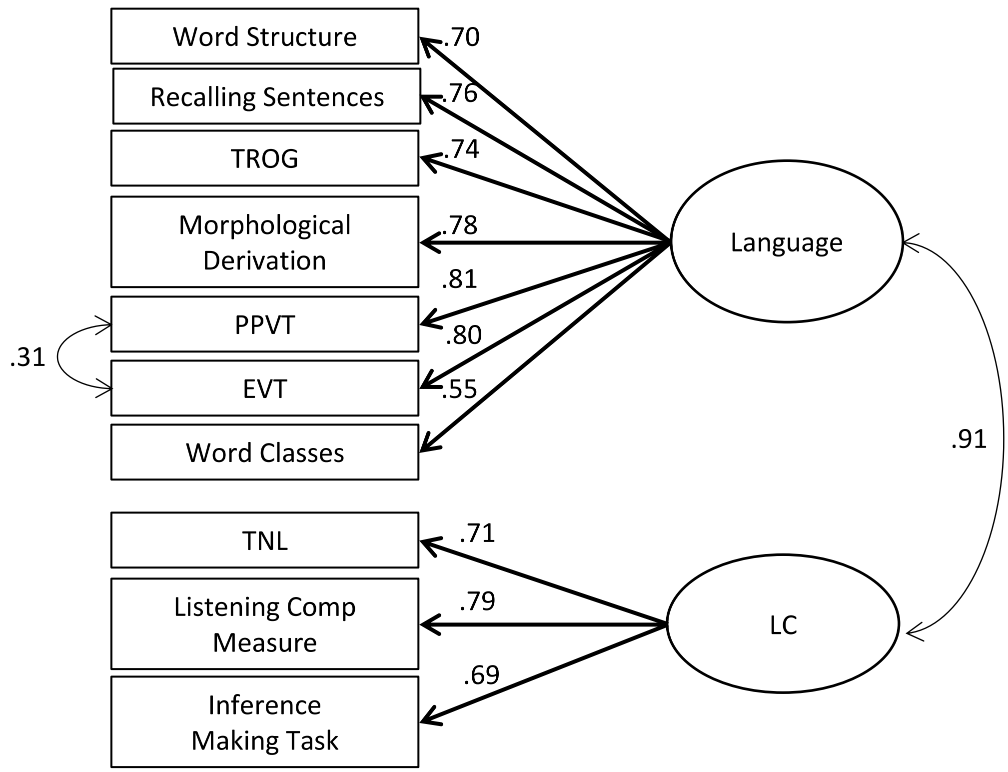
Appendix C

Structural Model for Grade1



Appendix D

Structural Model for Grade 2



Appendix E

Structural Model for Grade 3