Learning and Individual Differences 35 (2014) 56–61

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Preschoolers' oral language abilities: A multilevel examination of dimensionality

Jason L. Anthony *, Claire Davis, Jeffrey M. Williams, Teresa I. Anthony

Children's Learning Institute, University of Texas Health Science Center at Houston, United States

ARTICLE INFO

Article history:
Received 25 June 2013
Received in revised form 8 July 2014
Accepted 17 July 2014
Available online xxxx

Keywords:
Oral language
Preschool
Factor analysis
Vocabulary
Grammar
Phonology

ABSTRACT

This study explored the dimensionality of children's oral language using a fully crossed design, where modality (expressive versus receptive) was crossed with linguistic domain (grammar, semantics and phonology). The present study was also the first of its kind to employ multilevel factor analysis to control confounding classroom effects. Assessments of oral language were completed by 529 children who attended 85 preschool classrooms. The sample was heterogeneous with regard to language ability, ethnicity, sex and SES. Classroom membership was more highly associated with preschoolers' semantics and grammar than with preschoolers' phonological systems. A general language ability was found to drive preschoolers' performances on tests of semantics and grammar, regardless of modality. Furthermore, articulation and speech perception were found separate but correlated abilities. Implications for assessment and diagnosis of oral language impairments are discussed.

> "Oral language development is a critical foundation for reading, writing, and spelling, and it is the engine of learning and thinking".
> [(National Institute for Literacy, 2010)]

1. Introduction

1.1. Conceptual basis for distinctions among oral language abilities

A longstanding linguistic conceptualization of oral language asserts separate language systems underlying each domain. Pinker (1997, 1998) proposed a separate mental grammar for rules and a separate mental dictionary for words. However, an extensive review by Bates and Goodman (2001) articulates a more recent linguistic perspective that spoken language is acquired and processed by a unified processing system, rather than discrete and discontinuous mental grammar and mental dictionaries.

Clinical neuropsychology has also provided insights into the processing of oral language. Patients with agrammatic aphasia (i.e., a difficulty in composing sentences) can analyze regular past tense forms of words but not irregular ones, whereas patients with anomic aphasia (i.e., a difficulty in retrieving words) can analyze irregular past tense forms of words but not regular ones. Because both types of patients have injury to different brain regions, their linguistic weaknesses suggest a disassociation between the circuitry responsible for the grammatical analysis of words and for the memorization of words (Ullman et al., 1997). However in their review of the neuropsychology literature, Bates and Goodman (2001) argue that essentially all aphasic patients with deficits in grammar also demonstrate deficits in some aspects of lexical processing.

The American Psychiatric Association considered modality of deficit important for differential diagnosis of Expressive Language Disorder versus Mixed Receptive–Expressive Language Disorder in the prior version of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR; American Psychiatric Association, 2000). However, the recently published DSM-5 (American Psychiatric Association, 2013) foregoes the modality distinction and instead specifies that a Language Disorder is marked by persistent difficulties in acquisition and use of language *across* modalities. Furthermore, DSM-5's Language Disorder includes deficits in vocabulary, grammar, *and* discourse. Persistent difficulty with speech sound production was called Phonological Disorder in DSM-IV and subsequently called Speech Sound Disorder in DSM-5.

Speech–language pathologists generally work from a multidimensional framework, believing that distinctions along both linguistic domain and modality are important (Paul, 2001). Diagnostic evaluations are comprised of tests of semantics (e.g., vocabulary), grammar (i.e., morphology and syntax), and phonology (e.g., articulation, speech perception, phonological awareness, phonological memory) in both expressive and receptive domains. Multidimensional assessment permits treatment planning that follows differential diagnosis of those communication disorders outlined in DSM and other more fine grained

* Corresponding author at: Children's Learning Institute, University of Texas Health Science Center at Houston, 7000 Fannin Street, Suite 2377, Houston, TX 77030-1501, United States. Tel.: +1 713 500 3704.
E-mail address: Jason.L.Anthony@uth.tmc.edu (J.L. Anthony).

http://dx.doi.org/10.1016/j.lindif.2014.07.004
1041-6080/

language disorders, such as Specific Language Impairment, which is marked by a specific weakness in morphology in both modalities (Rice, 2003; Rice & Wexler, 1996; Rice, Wexler, & Hershberger, 1998).

In summary, oral language supports healthy academic, professional, family, community, and societal functioning. Many disciplines have contributed to the conceptualization of and understanding of the development and disorders of oral language. Oral language abilities are distinguished by linguistic domain (i.e., semantics, grammar, and phonology) and modality (i.e., expressive versus receptive). Many of these distinctions have been debated inconclusively, in part because most dissociation paradigms are prone to viable alternative interpretations.

1.2. Factor analytic research concerning dimensions of oral language

Exploratory factor analysis is a statistical method that can be used to examine distinguishability of oral language abilities. Most factor analytic research concerning the dimensionality of language has been carried out exclusively with children with atypical language abilities (e.g., Aram & Nation, 1975; Bishop & Edmundson, 1987; Rapin et al., 1996; Wilson & Risucci, 1986). Such methods may yield artificial distinctions among language abilities due to selection biases and attenuation of associations among language abilities due to restriction of range. To our knowledge, only three factor analytic studies have included typically developing children.

Colledge et al. (2002) administered a large assessment battery to 310 pairs of 4-year-old twins with normal language development. Exploratory factor analysis revealed a general Language factor and a Nonverbal ability factor. Most important for the topic at hand is that the Language factor was comprised of measures of semantics, grammar, and phonology. These results suggest that in the population of typically developing children individual differences in language domains overlap substantially, such that individuals who perform well in one domain tend to perform well in other domains too. Unfortunately, Colledge et al.'s study was not designed to specifically ask questions concerning the distinguishability of language skills along the dimensions of domain and modality, as many of their individual measures assessed multiple domains and some individual measures assessed multiple modalities. Also, Colleague et al.'s study did not detail how they excluded children with low language abilities. Such procedures could bias the results against finding factors that reflect separate abilities if the unique clustering of skills is present in children with low incident disorders, which is presumed by fields of speech and language pathology, special education, and psychology.

The most methodologically rigorous studies of the dimensionality of oral language were conducted by Tomblin and colleagues. Tomblin, Zhang, Weiss, Catts, and Ellis-Weismer (2004) studied 379 second grade children with normal language abilities and 225 second grade children with poor language abilities. The participants were administered both receptive and expressive measures of vocabulary and grammar. The participants were also administered expressive measures of phonological short-term memory, phonological awareness, and rapid naming. Exploratory factor analysis revealed that measures of expressive and receptive vocabulary and measures of expressive and receptive grammar formed a single Language factor and measures of expressive phonology formed a Phonological Processing factor. Tomblin and Zhang (2006) extended this work by performing analyses of these children's vocabulary and grammar scores obtained at kindergarten, second grade, fourth grade, and eighth grade. Exploratory and confirmatory factor analyses of data from each separate time point suggested that vocabulary and grammar formed a single language factor. Minor shortcomings of the studies by Tomblin and colleagues, for the purpose at hand, were that they did not include traditional expressive measures of articulation, and they did not include receptive measures of phonology that would be needed to complete a factorial design that fully crossed domain and modality.

1.3. Classroom effects on oral language

Complicating examination of relations among potentially distinguishable language abilities is the fact that children spend much of their time in classrooms. Moreover, much of the language instruction that young children receive is delivered by teachers. Teachers and classrooms differ systematically from one another, even within the same school. As such, the language abilities of children in the same classroom are more alike than those of children from different classrooms. For example, Branum-Martin et al. (2009) found that the clustering of children into classrooms, a.k.a., classroom nesting, accounted for up to 28% of the variance in children's vocabulary scores. The important implication is that statistical analyses that do not separate variability of children's performances into variability associated with classrooms and variability associated with individuals may yield erroneous results. Therefore, any substantive theory of language should be based on research that has accounted for the effects of classroom contexts, either methodologically by including only one child from each classroom or statistically by employing multilevel modeling.

1.4. Study purpose

Because this was the first study to examine dimensionality of oral language at the classroom-level and because it was the first study to examine dimensionality of preschool children's oral language while controlling for classroom nesting, which could have substantial impacts on the magnitude and directionality of child-level covariances, the present study was considered exploratory. Multilevel exploratory factor analysis was employed to simultaneously investigate dimensionality of oral language at both the child and classroom levels.

Extending prior research, the present study included assessments of three linguistic domains (i.e., semantics, grammar, and phonology) in both receptive and expressive modalities. The fully crossed design would permit factors reflecting separate domains or factors reflecting separate modalities to emerge from the factor analysis. A large and heterogeneous sample of preschool children was selected to reflect the full range of abilities in the population at large. However, like Tomblin et al. (2004), we purposefully overrepresented children at risk for having language impairments, so that the study had the power to detect patterns of associations among language skills that occur less frequently in at risk and disordered subgroups.

2. Method

2.1. Participants

Classroom inclusion criteria were full day programming and enrollment in the Texas Early Education Model (TEEM). This empirically validated model emphasizes frequent, intensive, and ongoing professional development for early childhood educators, onsite mentoring, regular monitoring of children's academic progress, and choice from among a list of research-based curricula (Landry, Anthony, Swank, & Monsegue-Bailey, 2009; Landry, Swank, Anthony, & Assel, 2011). Although utilizing a common educational model, TEEM classrooms are very diverse. For example, about an equal number of federally funded Head Start classrooms, state funded public school prekindergarten classrooms, and privately funded child care classrooms participated in the present study. Given that Head Start and state funded public pre-K programs both have eligibility requirements that include financial need, approximately two thirds of the sample was from a low SES background. Whereas all of the teachers from public pre-K programs had college degrees and were state certified in early childhood, teachers from Head Start and private programs ranged in educational attainment from high school through college.

Inclusion criteria for children were a nonverbal IQ score above 70 as estimated by the Block Design and Pattern Analysis subtests of the *Stanford Binet IV*, a vocabulary age equivalent above 2 years and 6 months as assessed by the *Expressive One-Word Picture Vocabulary Test*, and being a native, monolingual speaker of English. One older child with mental retardation was excluded from the analysis. Children learning English as a second language were excluded. Children with documented or observed speech and language delays were included. The sample of 529 children had average nonverbal cognitive abilities and average to low average language abilities (see Table 1). Most participants were 4 years old when tested (Mean = 4 years, 5 months; *SD* = 5 months). The sample was 50% female. The ethnic breakdown of the sample was 56% African-American, 23% Hispanic American, 15% Caucasian, and 6% multiracial.

2.2. Testing procedures

The children were tested individually at their preschools in locations that school administrators designated for testing. The assessment team was comprised of students enrolled in graduate, undergraduate, and post-bachelorette programs in speech–language pathology. The examiners attended a two-day training workshop led by a clinical child psychologist and speech language pathologist. Following training and ample practice, the examiners demonstrated competence on all tests during test-out sessions with the principle investigator. Testing was conducted in 20 to 45 min sessions depending on a given child's attention span and desire to continue. Testing of the entire sample took place over a six week period in late fall. The children were given verbal praise, physical praise (e.g., high fives), and tangible reinforcements (e.g., stickers) for participating.

2.3. Measures

2.3.1. Semantics

2.3.1.1. Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2000a). The examiners presented the children with colored line drawings that depicted an action, object, category, or concept. The children were asked to label each drawing. Prescribed cuing methods were used for elicitation if the children responded to the wrong part of an illustration or if they provided a response at the wrong level of abstraction. Standardized administration and scoring procedures were followed.

Table 1
Descriptive statistics for sample data.

Variable	Standard score *M*	Standard score *SD*	Raw score *M*	Raw score *SD*	Skew	Kurtosis	ICC	DE
RVCT	–[a]	–[a]	14.0	2.1	−1.55	3.19	–[b]	–[b]
EOWPVT	90.6	14.6	39.5	11.6	0.74	0.74	.23	2.2
ROWPVT	92.7	13.9	44.7	12.5	0.07	0.03	.17	1.9
CELFP-2 SS	89.9	14.6	11.2	4.7	−0.28	−0.60	.21	2.1
CELFP-2 WS	86.0	17.1	10.3	5.7	0.13	−0.97	.23	2.2
GFTA-2	107.3	10.5	10.3	8.7	1.72	3.82	.12	1.6
Artic Correction	–[a]	–[a]	9.3	6.9	1.56	3.45	.13	1.7
HSITA	–[a]	–[a]	10.3	7.0	1.45	3.39	.13	1.7
Artic Judgment Con.	–[a]	–[a]	6.2	3.8	−0.34	−1.30	.11	1.6
Artic Judgment Vow.	–[a]	–[a]	8.8	5.1	−0.33	−1.15	.16	1.8
NVIQ	86.2	8.6	–[b]	–[b]	–[b]	–[b]	–[b]	–[b]

Note. *N* = 529. ICC = Intraclass correlation; DE = Design effect. RVCT = Receptive Vocabulary Control Task (Anthony et al., 2010); EOWPVT = Expressive One-Word Picture Vocabulary Test (Brownell, 2000a); ROWPVT = Receptive One-Word Picture Vocabulary Test (Brownell, 2000b); CELFP-2 SS = Clinical Evaluation of Language Fundamentals Preschool—Second Edition Sentence Structure subtest (Wiig et al., 2004); CELFP-2 WS = Clinical Evaluation of Language Fundamentals Preschool—Second Edition Word Structure subtest (Wiig et al., 2004); GFTA-2 = Goldman–Fristoe Test of Articulation—Second Edition (Goldman & Fristoe, 2000); Artic Correction = Articulation Correction-Accuracy (Anthony et al., 2010); HSITA = Houston Sentence Imitation Test of Articulation (Anthony et al., 2010); Artic Judgment Con = Articulation Judgment scored for consonants; Artic Judgment Vow = Articulation Judgment scored for vowels (Anthony et al., 2010); NVIQ = Nonverbal IQ estimated from the Pattern Analysis and Block Design subtests of the Stanford Binet IV.
[a] Measures without national norms.
[b] Measures not used in factor analyses.

2.3.1.2. Receptive One-Word Picture Vocabulary Test (ROWPVT; Brownell, 2000b). The examiners stated a word and the children were required to point to one of four colored line drawings that corresponded with the stated word. Standardized administration and scoring procedures were followed.

2.3.1.3. Receptive Vocabulary Control Task (RVCT; Anthony, Williams, Aghara, Dunkelberger, & Novak, 2010). The children pointed to one of three pictures that was named by an examiner. The RVCT included 16 items that were used to verify receptive knowledge of the vocabulary words included in the Articulation Accuracy and Articulation Judgment tests (see below). Foils were from the same semantic field as the target (e.g., ambulance, fire truck, police car).

2.3.2. Grammar

2.3.2.1. Sentence Structure from the Clinical Evaluation of Language Fundamentals Preschool—Second Edition (CELFP-2; Wiig et al., 2004). This measure of receptive grammar required children to point to one of four colored line drawings that correctly illustrated a stimulus sentence. Testing followed standardized procedures.

2.3.2.2. Word Structure from the CELFP-2. This measure of expressive grammar employed a cloze procedure to assess children's competence with prepositions, possessive pronouns, regular and irregular verb tenses, subject/verb agreement, and inflections that denoted verb tense, possession, plurality, comparatives and superlatives. Testing followed standardized procedures.

2.3.3. Phonology

2.3.3.1. Goldman–Fristoe Test of Articulation—2 (GFTA-2; Goldman & Fristoe, 2000). The *Sounds-in-Words* subtest is commonly used by speech–language pathologists in clinical and school settings to evaluate accuracy of speech sound productions. It is a standardized test of expressive phonology, or more precisely articulation. A picture naming task elicits single word utterances from children. Articulation accuracy of 23 consonant sounds is evaluated from one to three times across initial, medial, and final positions. Target words are short, high frequency words that contain at least one targeted consonant or consonant cluster. Standardized administration procedures were followed.

2.3.3.2. Articulation Correction (Anthony et al., 2010). Because the GFTA-2 evaluates single word utterances of short, high frequency words, it provides a shallow assessment of articulation accuracy (Fowler & Swainson, 2004; Swan & Goswami, 1997). The Articulation Correction test was developed to assess articulation accuracy in the context of long, multisyllabic, phonologically complex words. The examiners displayed an illustration in view of both a puppet and the child. The puppet looked down at the picture, looked up at the child, and then named the picture in a much reduced form. Specifically, the puppet deleted a number of phonemes, usually from an unstressed syllable (e.g., /b^flaI/ for butterfly). Auditory stimuli were English nonwords. The child was instructed to tell the puppet how to say the word correctly. The number of consonant phonemes that were correctly articulated by the child was tallied. The 16 target words contained 24% early consonants, 59% middle consonants, and 18% late consonants (Shriberg, 1993; Shriberg & Kwiatkowski, 1982). Articulation correction demonstrated high internal consistency (Cronbach's alpha = .92).

2.3.3.3. Houston Sentence Repetition Test of Articulation (HSRTA; Anthony et al., 2010). The HSRTA was designed to assess articulation accuracy in the context of continuous speech. This measure was expected to provide a deeper evaluation of articulation accuracy than the GFTA-2 and Articulation Correction, given that children are able to accurately produce a given consonant phoneme in the context of single word utterances before they can in continuous speech (Andrews & Fey, 1986; Morrison & Shriberg, 1992). HSRTA required the children to repeat sentences spoken by an examiner. Consonant omissions, substitutions, and additions were scored as errors. Stimuli included 15 sentences that contained 34 early-developing consonants, 29 middle-developing consonants, and 20 late-developing consonants. Sentences conformed to the lower end of the typical range of mean length of utterance (MLU) for preschool children (MLU = 3.93, range = 2 to 7 morphemes), to minimize demands on memory. HSRTA demonstrates high internal consistency (Cronbach's alpha = .92), measures a single latent ability, demonstrates convergent validity with standardized measures of articulation, and demonstrates discriminant validity with standardized measures of oral language and phonological memory (Davis, Anthony, Dunkelberger, Aghara, & Williams, submitted for publication).

2.3.3.4. Articulation Judgment (Anthony et al., 2010). For this receptive test of phonological representation, a picture was displayed in view of a puppet and child. The picture illustrated a multisyllabic, phonologically complex word. The puppet looked down at the picture, looked up at the child, and then either correctly or incorrectly pronounced the name of the picture. The children were asked to indicate whether or not the puppet said the word correctly.

Articulation judgment yielded two scores. Of the 16 items that assessed phonological representation for vowels, about half involved the puppet misarticulating one vowel sound and the rest involved the puppet misarticulating two or three vowel sounds. Vowel manipulations included one deletion and 25 substitutions. Of the 11 items used to assess phonological representation for consonants, about half involved the puppet misarticulating one constituent consonant and the rest involved misarticulation of two or three consonants. Consonant manipulations included substitutions, cluster reductions, fronting, and spoonerisms. The brief articulation judgment measures demonstrated good internal consistencies (Cronbach's alphas = .89 and .88 for vowels and consonants, respectively).

3. Results

3.1. Preanalysis inspection of data

Descriptive statistics (see Table 1) indicated a ceiling effect on the 16-item Receptive Vocabulary Control Task (Mean = 14.0, SD = 2.0). Vocabulary items were correctly recognized by 72% to 98% of the participants (Mean = 90%, SD = 8%), significantly above the 33% chance level (zs = 14.94 to 25.34, $ps < .00001$). Thus, the words used as stimuli in the Articulation Correction and Articulation Judgment tests were within the participants' receptive vocabularies.

Scores on the three articulation tests were positively skewed and positively kurtotic, as to be expected from normative performances of 4-year-old children. Scores on the two receptive phonological representation tests evidenced some negative kurtosis. Because several variables evidenced mild departure from normality, subsequent analyses were performed with Mplus, version 6, using Robust Maximum Likelihood estimation method, Satorra–Bentler scaled chi-square, and adjustments to the standard errors to the extent of the nonnormality.

Intraclass correlations indicated classroom nesting accounted for 11% to 23% of the variance in children's performances (see Table 1). Considering ICCs in the context of the distribution of cluster sizes (Mean = 6.2, SD = 1.9), design effects ranged from to 1.6 to 2.3, indicating need to separate classroom-level variance from child-level variance. Measures of vocabulary and grammar had higher ICCs and design effects than measures of articulation and phonological representation. This finding highlights the need to separate variance in performances associated with children from that associated with classrooms, as not doing so may make children's semantics and grammar appear more highly correlated than warranted.

3.2. Multilevel exploratory factor analysis of data

To examine dimensionality of children's oral language, we first reviewed child-level eigenvalues (eigenvalues = 4.06, 1.71, 1.27, .51, .37, .34, .32, .27, and .13). Three factors produced eigenvalues greater than 1.0, and the screen plot suggested that models with 1, 2, or 3 factors were worthy of exploration. We tested one-, two-, and three-factor exploratory models at the child-level (L1), while allowing all covariances at the classroom-level (L2) to be freely estimated. The top panel of Table 2 shows that an adequate fitting L1 model was not achieved until three factors were modeled, given that good fitting models have SRMRs less than .05.

To examine dimensionality of classroom differences, we reviewed the classroom-level eigenvalues (eigenvalues = 5.81, 1.73, 1.05, .24, .12, .04, .01, .002, and .001). Three factors produced eigenvalues greater than 1.0, and the screen plot suggested that L2 models with 1, 2, and 3 factors were worth exploring, while allowing all covariances at L1 to be freely estimated. Values of SRMR reported in the middle panel of Table 2 indicate that an adequate fitting L2 model was not achieved until three factors were modeled. Overall model fits were good for all L2 models (i.e., CFIs and TLIs > .95), indicating children's performances were mostly attributable to individual differences, consistent with the intraclass correlations.

The final model included three factors at L1 and three factors at L2. This model yielded excellent overall model fits (CFI = .99, TLI = .98) and excellent level specific model fits (SRMRs = .01 and .03). Examination of obliquely rotated factor loadings (see Table 3) revealed that receptive and expressive semantics along with receptive and expressive grammar indexed a single latent Language ability at both levels. The two receptive phonological representation measures formed a Speech Perception factor at both levels, and the three expressive articulation measures comprised an Articulation factor at both levels. Factor intercorrelations were small but reliable at L1 (rs = .26 to .45 $ps < .001$) and L2 (rs = .47 to .54 $ps < .001$).

4. Discussion

4.1. Main findings

As the first multilevel investigation of oral language, this study revealed some interesting results concerning contextual effects of classrooms. One, the classroom context was found more highly associated with children's semantics and grammar than with children's phonology. Two, results from the classroom-level factor analysis imply that the contextual effects of classrooms on children's semantics were redundant with the contextual effects of classrooms on children's grammar. These novel findings are quite reasonable considering that general education is more likely to focus on vocabulary and grammar than on articulation and speech perception. Early childhood educators typically dedicate a large portion of their daily activities and curriculum to providing language instruction and they spend much less time providing feedback concerning accuracy of speech production or speech perception, which may be left to school-based speech–language pathologists.

There are two important practical implications of findings concerning classrooms' contextual effects. One, they highlight the benefit for researchers in separating variance in children's test scores due to children from variance in test scores due to classrooms because not doing so may skew results of statistical tests of effects at the child-level. Two, they provide reason for future studies to try to identify sources of classrooms' contextual effects that impact children's

Table 2
Fit statistics for alternative multilevel exploratory factor models.

Number of L1 factors	Number of L2 factors	SRMR for L1	SRMR for L2	Overall model CFI	Overall model TLI
Examination of factor structure at the child level					
1	Saturated	.16	.00	.59	.09
2	Saturated	.10	.00	.79	.20
3	Saturated	.01	.00	1.00	.97
Examination of factor structure at the classroom level					
Saturated	1	.00	.27	.98	.96
Saturated	2	.00	.16	.99	.98
Saturated	3	.00	.03	1.00	1.00
Final multilevel exploratory factor model					
3	3	.01	.03	.99	.98

Note. N = 529. L1 = Child level; L2 = Classroom level; SRMR = Standard Root Mean Residual; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index.

language development. For example, classroom differences that may directly impact children's language development include teacher knowledge and preparedness, frequency and quality of teacher–child interactions, curricula, full-day versus half-day programming, etc. Alternatively, classroom nesting effects could be explained by factors external to the classroom that determine which schools and classrooms children attend. For example, factors such as children's ages, the language or languages spoken in children's homes, family income, parents' religion, and other factors that lead parents to select which neighborhood to live could all help explain the contextual effects of classrooms on children's oral language development.

Having modeled both classroom-level covariances and child-level covariances, we have increased confidence in the present study's findings concerning the dimensionality of children's oral language. First, factor analysis at the child-level indicated no domain distinction or modality distinction with regard to semantics and grammar. Specifically, our preschool children's receptive and expressive semantics along with their receptive and expressive grammar reflected a single Language ability, extending findings of Tomblin et al. (2004), Tomblin and Zhang (2006) and Colledge et al. (2002) down to preschool-age children. Therefore contrary to Pinker's (1997, 1998) hypothesis of a mental dictionary separate from a metal grammar, these findings are consistent with a unified lexicalist view (Bates & Goodman, 2001), which holds that these language domains are operated by the same underlying system. Second, the present study found that general language ability was separable from phonology. This finding resembles that of Tomblin et al. (2004), whose factor analysis identified a latent general language ability separate from a latent phonological ability. Third, our finding that articulation and speech perception are distinguishable phonological skills parallels findings of Foy and Mann (2001). These later results are also consistent with Martin and Saffran's (2002) suggestion that input and output phonological abilities are separate abilities that are functionally related.

Table 3
Factor loadings from final multilevel exploratory factor model.

Variable	F1	F2	F3
Child level			
EOWPVT	.80	−.03	−.07
ROWPVT	.83	.03	−.02
CELFP-2 Sentence Structure	.70	.02	.04
CELFP-2 Word Structure	.70	−.17	.02
GFTA-2	.02	.86	.10
Articulation Correction	−.08	.77	−.02
Articulation Judgment Consonants	−.01	−.01	1.03
Articulation Judgment Vowels	.02	.01	.83
HSITA	−.05	.85	.00
Classroom level			
EOWPVT	.76	−.28	−.02
ROWPVT	1.06	.00	−.12
CELFP-2 Sentence Structure	1.10	.30	.01
CELFP-2 Word Structure	.81	−.18	.10
GFTA-2	.24	1.04	−.01
Articulation Correction	−.01	1.15	.32
Articulation Judgment Consonants	.19	.00	.87
Articulation Judgment Vowels	.01	−.11	.96
HSITA	.01	.87	−.13

Note. N = 529. EOWPVT = Expressive One-Word Picture Vocabulary Test (Brownell, 2000a); ROWPVT = Receptive One-Word Picture Vocabulary Test (Brownell, 2000b); CELFP-2 = Clinical Evaluation of Language Fundamentals Preschool—Second Edition (Wiig et al., 2004); GFTA-2 = Goldman–Fristoe Test of Articulation—Second Edition (Goldman & Fristoe, 2000); HSITA = Houston Sentence Imitation Test of Articulation (Anthony et al., 2010).

4.2. Implications for assessment and diagnoses of oral language impairments

The present findings call into question the practice of administering a large battery of multiple tests of semantics, morphology, and syntax, when perhaps only one or two such tests are needed to provide a reliable index of general language ability. Of course, precision of estimation of a child's language ability will improve with administration of additional tests that measure the same construct, but the law of diminishing returns applies. Besides being an inefficient use of assessment time, administration of multiple tests of semantics, morphology, and syntax will increase the likelihood of mistakenly interpreting small variations in less reliable subtest scores as strengths and weaknesses that are important for classification. Nonetheless, comprehensive assessment remains informative for treatment planning.

The present findings do support the current practice of separately assessing children's phonological abilities from their general language ability. Furthermore, perhaps speech–language pathologists should consider assessing both articulation and speech perception. However, this recommendation should be considered tentative until further research demonstrates the diagnostic utility, predictive validity, and practical importance of separate indices of articulation accuracy and speech perception, given that it may only be the variance which is shared among them that is important for development of oral and written language.

The present findings are consistent with DSM-IV and DSM-5 classifications of Phonological Disorder and Speech Sound Disorder, respectively, as separate from general language ability. The present findings and those of Tomblin and Zhang (2006) support DSM-5 foregoing modality and domain distinctions in its definition of Language Disorder as marked by persistent difficulties in vocabulary and morphosyntax across modalities. We echo Leonard's (2009) assertion that the former diagnosis of Expressive Language Disorder in the context of normal receptive language skills should have been viewed with caution because most expressive language tasks require comprehension.

4.3. Limitations

First, although we measured three language domains, we did not measure pragmatics. Pragmatics is challenging to measure because social norms are culturally specific. Consequently, there is little agreement on how to validly measure pragmatics (McNamara & Roever, 2006).

Second, although the present study included a large number of classrooms, the intraclass correlations were small, which limit the factorability of the classroom-level data. As such, conclusions concerning the dimensionality of classroom's oral language skills should be considered tentative until replicated.

4.4. Summary

Preschool children's oral language is characterized by three latent abilities, namely general language, articulation, and speech perception. These findings are consistent with much prior research and recent changes in nosology. Results support the common practice of evaluating speech and language abilities separately. However, they call into question the common practices of distinguishing expressive language from receptive language and distinguishing semantics from grammar, at least for the purpose of classification.

Acknowledgments

The research reported here was supported in part by the W.K. Kellogg Foundation, through Grant P30041790 to the University of Texas Health Science Center at Houston. The opinions expressed are those of the authors and do not represent views of W.K. Kellogg.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association (Text revision (DSM-IV-TR)).

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association (DSM-5).

Andrews, N., & Fey, M. E. (1986). Analysis of the speech of phonologically impaired children in two sampling conditions. *Language, Speech, and Hearing Services in the Schools*, *17*, 187–198.

Anthony, J. L., Williams, J. M., Aghara, R., Dunkelberger, M., & Novak, B. (2010). Assessment of individual differences in phonological representation. *Reading and Writing: An Interdisciplinary Journal*, *23*, 969–994.

Aram, D., & Nation, J. (1975). Patterns of language behavior in children with developmental language disorders. *Journal of Speech and Hearing Research*, *18*, 229–241.

Bates, E., & Goodman, J. (2001). On the inseparability of grammar and the lexicon: Evidence from acquisition. In M. Tomasello, & E. Bates (Eds.), *Language development: The essential readings* (pp. 134–162). Oxford, England: Blackwell.

Bishop, D.V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds—Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, *52*, 156–173.

Branum-Martin, L., Mehta, P. D., Francis, D. J., Foorman, B. R., Cirino, P. T., Miller, J. F., et al. (2009). Pictures and words: Spanish and English vocabulary in classrooms. *Journal of Educational Psychology*, *101*(4), 897–911.

Brownell (2000a). *Expressive one word picture vocabulary test*. Novato, California: Academic Therapy Publications.

Brownell (2000b). *Receptive one word picture vocabulary test*. Novato, California: Academic Therapy Publications.

Colledge, E., Bishop, D.V. M., Koeppen-Schomerus, G., Price, T. S., Happe, F. G. E., Eley, T. C., et al. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology*, *38*, 749–757.

Davis, C., Anthony, J. L., Dunkelberger, M., Aghara, R. G., & Williams, J. (2013n). Development and validation of a brief assessment of preschoolers' articulation: The Houston sentence imitation test of articulation Manuscript submitted for publication.

Fowler, A. E., & Swainson, B. (2004). Relationships of naming skills to reading, memory, and receptive vocabulary: Evidence for imprecise phonological representations of words by poor readers. *Annals of Dislexia*, *54*, 247–280.

Foy, J. G., & Mann, V. A. (2001). Does strength of phonological representations predict phonological awareness in preschool children? *Applied Psycholinguistics*, *22*, 301–325.

Goldman, R., & Fristoe, M. (2000). *Goldman Fristoe 2 Test of Articulation*. Pearson Assessments.

Landry, S., Anthony, J. L., Swank, P., & Monsegue-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology*, *101*, 448–465.

Landry, S., Swank, P., Anthony, J. L., & Assel, M. (2011). An experimental study evaluating professional development activities within a state funded pre-kindergarten program. *Reading and Writing: An Interdisciplinary Journal*, *24*, 971–1010.

Leonard, L. B. (2009). Is expressive language disorder an accurate diagnostic category? *American Journal of Speech–Language Pathology*, *18*, 115–123.

Martin, N., & Saffran, E. M. (2002). The relationship of input and output phonological processing: An evaluation of models and evidence to support them. *Aphasiology*, *16*, 107–150.

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research*, *35*, 259–273.

National Institute for Literacy (2010). *Learning to talk and listen: An oral language resource for early childhood caregiver*. Washington, DC: National Institute for Literacy https://lincs.ed.gov/publications/pdf/LearningtoTalkandListen.pdf.

Paul, R. (2001). *Language disorders from infancy through adolescence: Assessment and intervention*. St. Louis: Mosby.

Pinker, S. (1997). Words and rules in the human brain. *Nature*, *387*, 547–548.

Pinker, S. (1998). Words and rules. *Lingua*, *106*, 219–242.

Rapin, I., Allen, D. A., Aram, D., Dunn, D., Fein, D., Morris, R., et al. (1996). Classification issues. In I. Rapin (Ed.), *Preschool children with inadequate communication* (pp. 190–228). London: MacKeith.

Rice, M. L. (2003). A unified model of specific and general language delay: Grammatical tense as a clinical marker of unexpected variation. In Y. Levy, & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of SLI* (pp. 63–94). Mahwah, NJ: Erlbaum.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*, *39*, 1239–1257.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1412–1431.

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech, Language, and Hearing Research*, *36*, 105–140.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders*, *47*, 256–270.

Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, *66*, 18–41.

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.

Tomblin, J. B., Zhang, X., Weiss, A., Catts, H., & Ellis-Weismer, S. (2004). Dimensions of individual differences in communication skills among primary grade children. In M. Rice, & S. F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 53–76). Mahwah, NJ: Erlbaum.

Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., et al. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, *9*, 289–299.

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals Preschool - Second Edition. San Antonio, TX*.Inc: Harcourt Assessment.

Wilson, B. C., & Risucci, D. (1986). A model for clinical-quantitative classification. Generation I: Application to language-disordered preschool children. *Brain and Language*, *27*, 281–309.