The dimensionality of oral language ability: evidence from young Greek children

Angeliki Mouzaki
Primary Education Department
University of Crete
amouzaki@uoc.gr

Elli Spyropoulou
Psychology Department
University of Crete
elli_spiropoulou@yahoo.gr

Asimina Ralli
Psychology Department
National and Kapodistrian University of Athens
asralli@psych.uoa.gr

Faye Antoniou
Department of Philosophy, Psychology and Education
National and Kapodistrian University of Athens
fayeantoniou@ppp.uoa.gr

Vassiliki Diamanti
Department of Special Needs Education
University of Oslo
vassiliki.diamanti@isp.uio.no

Sophia Papaioannou
Special Education Department
University of Thessaly
papaioannousophia@gmail.com

Address correspondence to:
Angeliki Mouzaki
Primary Education Department, University of Crete
Gallos Campus, Rethymno 74100
Crete, Greece
E-mail: amouzaki@uoc.gr

Conflict of Interest

The language measures reported here form part of a commercially available assessment battery (Logometro, produced by Inte*Learn Multimedia Educational Applications) designed be the authors (AM, AR, FA, VD & SP), who receive part of the proceeds from its use.

Funding

This research was supported in part by a postodoctoral research fellowship to VD in the context of research program “The foundation of Reading and Writing in a Transparent Orthography: Oral Language Development and Early Literacy Skills” funded by the University of Crete Special Account of Research (PI:AM)

Abstract

Purpose: This study investigated component skills in oral language development utilizing and validating a new assessment battery in a large (N=800) and representative sample of Greek students 4-7 years of age.

Method: All participants enrolled in public schools from four geographical regions (Attica, Thessaly, Macedonia and Crete) that varied demographically (urban, semi-urban and rural). For the individualized language assessments we utilized mobile devices (tablet PC) to ensure children's interest and joyful participation as well as reliable administration procedures across sites. Results by Confirmatory Factor Analyses specified and validated five different models in each grade to identify the best conceptualization of language dimensionality in the respective age groups.

Results: Four-dimensional model provided a slightly better discriminant validity in language data of the preschool group. However, in kindergarten and first grades, the five-dimensional model had the best fit to the data to the four-dimensional.

Conclusion: These findings support the multidimensionality of oral language ability at this phase of development and increase of factor distinctiveness as children grow.

Keywords: language assessment, language components, early literacy

The Dimensionality of Oral Language Ability: Evidence from Young Greek Children

Language development in early childhood has been studied extensively in an effort to explain observed variations in both oral language and emerging literacy skills. The search for contributing factors is motivated in part by the need to refine psychoeducational prevention and early interventions addressing early reading difficulties (Fletcher, Shaywitz, Shankweiler, Katz, Liberman, Stuebing, Francis et al., 1994; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Morris, Stuebing, Fletcher, Shaywitz, Lyon, Shankweiler et al., 1998; Ramus, Rosen, Dakin, Day, Castellote, White et al., 2003; Stanovich & Siegel, 1994). Hypotheses related to the role of certain language skills in literacy development are regularly leading research efforts and informing educational practice without doubting the componential view of language domains. Accordingly, most language assessments currently in use include largely common groups of tasks. This notion of oral language dimensionality is related to the concept of modularity (i.e., vocabulary and grammar) assuming the existence of distinct dimensions and associated abilities. For example, child's understanding of the phonological units of spoken language is an important source of variance in language ability that associates strongly to literacy development and difficulties.

In this context, phonological awareness and oral vocabulary are among the earliest and most-well studied linguistic abilities. Several large-scale longitudinal studies strongly support the role of phonological awareness and oral vocabulary in preschool years for later reading achievement, including both word recognition and text comprehension skills (e.g., Fletcher, Foorman, Boudousquie, Barnes, Schatschneider, & Francis, 2002; O'Connor & Jenkins, 1999; Scarsborough, 1998; Torgesen, 2002; Vellutino, Scanlon, & Lyon, 2000; Wood, Hill, & Meyer, 2001). Much less research has been devoted to other language skills and their relation to reading achievement (Foorman, Koon, Petscher, Mitchell & Truckenmiller, 2015). Recently the contribution of morphological and syntactic skills to literacy development has come into focus yielding very promising findings toward understanding how oral language development facilitates access to written language (Dickinson, McCabe, Anastopoulos, Peisner-Feinberg & Poe, 2003; Foorman, Herrera, Petscher, Mitchell & Truckenmiller, 2015). Recent conceptualizations of language skills also emphasize developmental interactions, as well as distinctions, between language domains. Such interactions may introduce variations in the magnitude of associations between language domains and associated skills with age. Two language domains that may be related through bidirectional links are phonology and semantics. This notion is based on correlational findings in typical preschoolers (Bowey, 1994; Lonigan, Burgess, Anthony & Barker, 1998; Lonigan et al., 2000; Wagner, Torgesen, Laughton, Simmons & Rashotte, 1993; Wagner et al., 1997). Moreover, limited vocabulary size seems to be associated to phonology disorders, and poor phonemic awareness (Best, 2005) while phonological possessing skills seem to facilitate semantic and syntactic development (Chiat, 2001). Further, the presence of significant links between semantics and morphosyntax is supported by evidence from syntactic errors produced by children with language disorders as they are associated with impaired semantic possessing (McGregor, Berns, Owen, Michels, Duff, Bahnsen & Lloyd, 2012).

The pattern of associations between oral language skills in early childhood has implications for modeling key predictors of emerging literacy skills. Thus, poor phonological processing capacity may impact reading acquisition and achievement both directly (i.e., manifested as poor phonological awareness; Torgesen, Wagner & Rashotte, 1994, Torgesen, 1996), as well as indirectly by posing limitations on other domains such as verbal short-term memory (Rapala & Brady, 1990; Stone & Brady, 1995) and lexical retrieval (Catts, Fey, Zhang, & Tomblin, 1999, Wolf, Bowers & Biddle, 2000). Moreover, oral vocabulary may affect reading comprehension both directly by ensuring adequate text understanding, as well as indirectly by facilitating efficient word recognition (Ouellette & Beers, 2010). This dual effect of vocabulary appears to vary with age and reading experience: the effect of word knowledge on decoding diminishes with age as automaticity is achieved in upper grades and accounts for progressively larger amounts of variance in text comprehension (Kim, Al Otaiba, Puranik, Folsom & Greulich, 2014; Protopapas, Sideridis, Mouzaki & Simos, 2007; Suggate, Reese, Lenhard & Schneider, 2014).

In spite of the wide popularity of the componential view of language ability, the notion of multidimensionality of language function is not undisputable (LARRC, 2015a). Thus, there have been arguments in favor of a unitary view for the language construct, stirring a debate regarding the distinction among language modules and dimensions and questioning the validity of widespread measurement practices (Bates & Goodman, 2001; Tomblin & Zhang, 2006). Available empirical evidence has been neither conclusive nor sufficient to support conceptualization of distinct language dimensions compounding this “theoretical uncertainty” (Klem, Gustafsson & Hagtvet, 2015; LARRC, 2015a; LARRC, Yeoman-Maldonado, Bengochea & Mesa, 2018; Lonigan & Milburn, 2017; Tomblin & Zhang, 2006).

The identity of psychometrically distinct domains of linguistic capacity and the pattern of associations between them is important on both clinical and theoretical grounds. The early distinction between the expressive and receptive language modalities had been very popular in clinical applications however, without support by recent research, at least for children older than 4 years of age (Lonigan & Milburn, 2017; Tomblin & Zhang, 2006). Most of the existing evidence regarding language dimensionality comes from studies that focus on semantic and structural language data without examining metalinguistic awareness, as measured by phonological processing skills. Current models emphasize distinctions among semantics, grammar and pragmatics. Notably, the latter domain is also very often neglected in studies specifically addressing dimensionality, partly because of the required time and uncertainty in data scoring and coding procedures. Pragmatic ability refers to language use in specific communicative contexts and many assessment efforts are often insufficient since they are affected not only by normative assessment limitations but also by child's personal style of communication (Adams, 2002). The addition of narrative discourse has also been proposed as another dimension that needs to be taken into account along with vocabulary and grammar (LARRC, 2015a, 2017; Dickinson, McCabe, Anastopoulos, Peisner-Feinberg & Poe, 2003). Another large study (N=915) that also assessed comprehension of *oral discourse* utilizing measures of comprehension monitoring, narrative structure and inference skills indicated that discourse emerges as a distinctive language dimension compared to lower level skills (such as vocabulary and grammar) especially for older children (7-8 years) and constitutes an important predictive variable of literacy outcomes (LARRC, 2015a).

Oral Language Assessment

Dimensionality of language functions at various stages of language development has important implications for methods of assessments of related language skills. Assessing preschoolers' language development accurately is a very challenging task that requires not only psychometrically sound assessment tests but also amble experience in their administration. Children at this age demonstrate great variation in speech production and language skills and a more limited attention span compared to school age children (Dockrell, 2001; Dockrell & Marshall, 2015). Examiners ought to be engaging and skillful in keeping young children on task and eliciting their best verbal responses while adhering to test guidelines. The child's temper and overall attitude can be an additional challenge as well as their motivation to participate in any given task. Furthermore, language exchange at this age depends greatly upon the communicative context since many children can be quite reluctant speaking with unknown adults, especially in unfamiliar settings and their reservation could influence assessment results (Conti-Ramsden & Durkin, 2012).

On the other hand, accurate language assessment entails not only examination of children's expressive language but also evaluating their understanding of speech. Many children could appear reluctant to talk during assessment because they have difficulties understanding the new information presented to them or following directions given by the examiner. Listening comprehension problems have been considered as very important in detecting language disorders and predicting future outcomes, while children with difficulties only in expressive language are more often expected to develop adequate skills later (Conti-Ramsden & Durkin, 2012).

In all, associations among various oral language skills, as mentioned before, and testing challenges call for comprehensive assessments that cover thoroughly all language domains. Furthermore, as indicated in previous studies (Lonigan & Milburn, 2017), is of particular importance to utilize multiple assessment tasks that vary in format and modality of required responses for each of the constructs studied. This way, variance in underlying common skill found by utilizing several and dissimilar tasks could support a more accurate distinction among the emerging factors (Dockrell, 2001; Dockrell & Marshall, 2015).

The Present Study

The distinction among language dimensions and evidence of their interrelations are quite important for both theoretical and practical reasons especially as changes during development have been detected. Weaknesses at certain language skills could affect other skills or mask their development, having a distinct impact upon overall language ability and serving as predictors of oral and written language problems (i.e. specific language impairment and dyslexia). Untangling different language skills and studying their distinct contributions at different phases of literacy development is a very promising line of research and crucial for clinical and educational interventions. Therefore, comprehending complex interrelations among language dimensions in preschool years and their complex interactions with emergent literacy skills could have several implications for the design and interpretation of the results of psychometric assessments employed in the context of prevention/early intervention efforts targeting literacy difficulties (LARRC, 2015a; Lonigan & Milburn, 2017).

As part of the present cross-sectional study we developed a comprehensive battery assessing a wide range of linguistic skills in a developmentally appropriate and motivating manner to evaluate structural models of language dimensions. Despite increasing awareness of the importance of narrative discourse and pragmatics as key elements for everyday communication ability, they have received very little attention in studies addressing the dimensionality of language domains in early childhood (LARRC, 2015b). In addition, metalinguistic skills, such as phonological awareness, known for their importance for literacy skills, are not typically part of the assessment batteries employed in studies on the dimensionality of language (LARRC, 2015a; Lonigan & Milburn, 2017; Tomblin & Zhang, 2006). Study differences in the range of evaluated skills, sample age, and assessment methods may account for previous inconclusive results.

Another issue that deserves attention relates to the language under study. The bulk of existing evidence regarding language structure has been provided by investigations in English-speaking populations, therefore reflecting the features and structure of the English language. As it has been previously reported, the English morphology is sparse while has been found to emerge later than Spanish (LARRC, Yeoman-Maldonado, Bengochea & Mesa, 2018). In contrast to languages with rich morphology (i.e., Spanish, Greek, etc.) acquisition of language form in English is slower and may depend more heavily upon lexical development (Devescovi, Caselli, Marchione, Pasqualetti, Reilly & Bates, 2005; Marchman & Bates, 1994.) Dimensionality research in other languages is very limited with only two studies involving young Spanish speakers (LARRC, 2015b; LARRC, Yeoman-Maldonado, Bengochea & Mesa, 2018). Therefore, cross-linguistic evidence from other languages would greatly enhance our understanding as we anticipate findings to help us advance current models on structure of component language constructs.

Standard Modern Greek has its origin in Koine, the spoken language of the Hellenistic era and is written with the Greek script, which originates from the Phoenician script. With regard to phonology, the size of vowel and consonant inventories is of average size, the syllabic structure is complex and has lexical stress (for more information and references, see Protopapas, 2017). Greek retains most of the inflectional characteristics of the ancient language categorizing nouns into masculine, feminine, and neuter while adjectives agree in gender, number, and case with their nouns, as do the articles. Verbs are inflected for mood (indicative, subjunctive, imperative), aspect (perfective, imperfective), voice (active, passive), tense (present, past), and person (first, second, and third, singular and plural). Inflected Greek words indicate relevant grammatical information through their endings that express values for several categories simultaneously (Brian, 1990). Thus, Greek morphology is rich, both inflectional and derivational. The most common form of derivation is by suffixation (most derivation and inflection involve suffixes and/or vowel change). Extensive formation of new words is achieved by compounding of stems and word including noun-noun compounds, verb-noun compounds, etc. Greek syntax is characterized by flexible word order (Subject-Verb-Object/Verb-Subject-Object) (Holton, Mackridge, Philippaki-Warburton & Spyropoulos, 2012). In developmental context, morphological acquisition is to a great extent, if not completely, achieved, upon entrance in primary education (Diamanti et al., 2017). Grammatical gender in Greek is usually acquired early on (Tsimpli, 2014; Tsimpli & Hulk, 2013), as are case and number for adjectives and nouns, as well as person and number for verbs (Mastropavlou, 2006). Verb aspect is somewhat more challenging and some difficulties with gender may arise in certain categories of nouns with uncommon features (Varlokosta & Nerantzini, 2013, 2015). Morphological awareness of inflectional processes is usually achieved by the first school years, while derivational processes appear to develop later (Diakogiorgi, Baris & Valmas, 2005; Diamanti et al., 2017).

The present study aspires to contribute to current conceptualizations of language dimensionality and to address the aforementioned issues by assessing young speakers of a highly inflected language, characterized by rich morphology. Accordingly, the present study examines a) interrelations among language skills in a large population of young Greek children, b) different models of language dimensions using confirmatory factor analysis in order to determine the structure of language abilities as measured by a comprehensive battery of language tests (Logometro) varying in cognitive and response demands and in levels of linguistic complexity, and c) potential changes in language dimensionality from preschool to Kindergarten and first grade.

Method

Participants

The participant pool for the study were 995 4–7 year old Greek students (521 boys) recruited from several schools in rural (19.4%), semi-urban (11.7%) and urban (68.9%) areas of four geographically dispersed provinces of Greece (Attica, Thessaly, Macedonia and Crete) spanning a variety of socioeconomic backgrounds. Children eligible for participating in the study were native speakers of Greek and students without sensory deficits or diagnosis of developmental delays that could prevent them from enrolling in a typical school classroom. Children with diagnosis of language impairment were not included in the study. Students were randomly selected from each school but they were included in the final sample only if their parent/guardian had provided written informed consent (the overall return rate of signed consent forms was approximately 60%). Parental educational level was reported by 69.2% of male parents (23.8% had graduated from high school and 45.4% were College graduates) and 50.8% of female parents (24.8% had high school diploma and 26% had a College degree). The protocol of the study was approved by the Institute of Education, Greek Ministry of Education, and the Research Ethics Committee of the University of Crete.

Complete data were available on 800 students (384 boys) attending prekindergarten (PK; n=180, M = 4.68 years, SD = .45 years), kindergarten (K; n=269, M = 5.6 years, SD = .36 years), and Grade 1 (G1; n=351, M = 6.62 years, SD = .4 years).

Language Measures

All children were evaluated with the Logometro language assessment battery developed by our group specifically for the Greek language (Mouzaki, Ralli, Antoniou, Diamanti, Papaioannou, 2017). Logometro is administered through an Android application (app) for mobile devices (tablets) featuring automatic application of ceiling rules. It was designed to cover a wide range of language skills (semantic knowledge, phonological awareness and processing, morphological awareness, oral discourse and pragmatics) in children aged 4-7 years. Test items within each task assessing the various language domains were arranged in ascending order of difficulty adhering to floor and ceiling rules that were determined during extensive pilot testing (Kanellou, Korvesi, Ralli, Mouzaki, Antoniou, Diamanti & Papaioannou, 2016; Ralli, Kazali, Kanellou, Mouzaki, Antoniou, Diamanti & Papaioannou, submitted). Task presentation and item order within each task were the same for all participants who were initially trained through two practice items that were followed by proper feedback and repeated up to three times if needed.

Task sequence was designed to ensure variety in demand of language modality (receptive/expressive) and to maintain child's interest and task engagement (receptive vocabulary, phonological tasks, listening comprehension, retelling, story comprehension, word definitions, naming, narration, morphological tasks, and pragmatics). Scoring (number of correct responses) was recorded automatically. Psychometric properties of the Greek version of the test are reported elsewhere (Mouzaki, Ralli, Antoniou, Diamanti, Papaioannou, 2017).

Semantic tasks. Vocabulary knowledge was assessed with three different tasks (Listening comprehension task [Cronbach's α=0.78], Receptive vocabulary task [Cronbach's α=0.88], Naming task [Cronbach's α=0.72]), addressing both receptive and expressive/naming skills. Student's responses in all semantic tasks were scored with 1 or 0. Test items/words were arranged in ascending order of difficulty (based on pilot assessment data) and the administration of the task was stopped when children made a number of consecutive errors according to each task's ceiling criterion. The tasks are described in detail below.

In the listening comprehension task (16 items) each child was invited to look carefully at four different images and then to choose the appropriate image that corresponded to the sentence heard. For example, the child was asked to «point to the boy with short hair» with distractors a boy with long hair, a girl with short hair and a girl with long hair.

In the receptive vocabulary task (30 items) each child was presented with four different images and was asked to choose the image that best represented the word that was heard. For example, the child was given the word “dolphin” with distractors pictures of a shark, a whale and a scooter.

In the expressive vocabulary/naming task (20 items) each child was asked to look at a picture of an object and then to name it. Test items included soap, cherry, paddle, compass, etc.

Phonological awareness and processing tasks. For assessing phonological awareness as a unitary construct that varies on a continuum of complexity, we administered three tasks appropriate for younger children (i.e. 4-5 years): initial phoneme matching (Cronbach’s α=0.84), phonemic synthesis (Cronbach’s α=0.93), and phonemic deletion (Cronbach’s α=0.92).

In the *identification of similarities task* at the level of initial phoneme (7 items) each child was listening to the label of a target image as well as to the labels of three other images displayed simultaneously and then had to choose which of the three images began with the same phoneme as the target image. For example, the target image “flower” /luluδi/ “λουλούδι” was followed by images of a tree /δedro/ “δέντρο”, a rabbit /laγos/ “λαγός” and a pencil /molivi/ “μολύβι”.

In the *phonemic synthesis* task (7 items) each child had to compose words from a sequence of spoken phonemes. The first item constituted a two-phoneme word, followed by two items constituting three-phoneme-words, two items making up four-phoneme-words and two items making up five-phoneme-words.

Finally, in the *phonemic deletion* task (7 items) the child was asked to listen carefully to a word and then to repeat it by deleting the initial phoneme of the word heard. For example, the directions were: “say the word gifts /δora/”, “say the word again without the /δ/”. Administration of all phonemic awareness tasks was discontinued after four consecutive errors.

Morphological awareness tasks. Morphological awareness skills were assessed through three different tasks assessing judgment (8 items; α=0.80) and production (11 items; α =0.73) of inflectional suffixes and production of derivational suffixes (16 items; α=0.94). In order to account for pseudoword use in the tasks, children were presented with a series of pictures that introduced animal heroes (penguins) that were supposed to use their own language.

In the inflectional morpheme judgment task children were presented with a picture displaying either one or two turtles performing an action while listening to two sentences spoken by two penguin figures, and had to choose the right sentence. Each pair of sentences contained one pseudoword which was either singular or plural (differing in inflectional suffix), and thereby matched the picture context. For example, given a picture of two turtles taking photographs, the two sentences were “the turtles skeni (3rd singular) photos” and “the turtles skenoun (3rd plural) photos” (the critical pseudoword is denoted by italics).

In the *inflectional morphemes production task* children saw a pair of pictures, displaying actions performed by turtles while at the same time they listened to an oral description of the first picture with a sentence that included a pseudoword. Children were then given the beginning of a second sentence, matching the second picture, up to the subject of the verb, and were asked to complete the sentence. Changes in the pseudoword number (from singular to plural or from plural to singular) had to be made accordingly (e.g., First sentence/picture: “Turtle plays with zagon (pseudoword for wagon). Second sentence/picture: Turtle plays with…. requiring “zagons”).

In the *derivational morphemes production task* children saw a picture and at the same time listened to a sentence with a target word (a different one for each sentence) and the beginning of a second sentence that was syntactically altered; it required children to manipulate the derivational morpheme within each target word in order for the sentence to be completed correctly (e.g., “The sea deepens. The sea is…” requiring “deep”). The task covered a variety of derivational morphemes, denoting property, profession, establishment/institution, material, collection, comparatives, action, device, nationality/origin, etc.

Pragmatics tasks. Pragmatic skills were assessed by a task that invites children to produce suitable oral responses by answering twenty-one questions that accompany 11 illustrated scenarios. All scenarios were developed using attractive, child-friendly pictures illustrating familiar experiences for the respective age groups (i.e. eating lunch with a parent, going to bed, etc.). Participants were asked to answer verbally 1-3 questions based on communication goals and pragmatic aspects of each scenario that were adapted from “The pragmatics profile of everyday communication skills in children” (Dewart & Summers, 1995) by taking into account cultural and linguistic differences. Questions were validated during extensive pilot testing and were intended to assess the following aspects of pragmatics: (a) comprehension and interpretation of the communicative situation presented in the picture (Context interpretation: Participating in Interaction, Initiating Interaction, Maintaining an Interaction or Conversation, Joining a Conversation), (b) intention/ ability to communicate (Communicative intent: Attention Directing, Requesting, Rejecting, Greeting, Self-Expression and Self-Assertion, Naming, Commenting, Giving Information), and (c) interactional skills related to the contextual variation (proper response-context: Gaining Someone’s Attention, Interest in Interaction, Understanding of Gesture, Understanding of Speaker’s Intentions Anticipation, and Negotiation) (Cronbach’s α=0.81).

Students were familiarized with the task at the beginning of the pragmatic skills’ assessment by replying to 3 questions and receiving immediate feedback. Answers provided by the children were scored with 1 to 7 points for each of the main aspects depending on quality and richness. Scoring for each of the main aspects was determined by specific criteria set for each item/question that followed the scenario. For example, one scenario shows a happy child sitting on the only swing available and two sad children looking at him. One of the related questions was “What should the children say to the boy in order to swing too?” Criteria used to evaluate verbal responses to the previous scenario include: understanding of speakers’ intention (i.e. children are requesting something), request for a specific reaction (e.g. get down of the swing), seeking and maintaining personal attention through use of the appropriate referent (e.g. use the word «you»), politeness/kindness (e.g. use of the word «Please»), expression of feelings (e.g. we are sad that we cannot swing too), etc.

Oral discourse tasks. Oral discourse skills were assessed through 4 different tasks (word definition, retelling, narration and story comprehension).

In the *word definitio*n *task* (Cronbach's α=0.93) each child was asked to give a brief definition of a series of words. Scoring matched other similar tasks (i.e. WISC vocabulary) where a proper word definition received 2 points while examples of word use or descriptions were scored with 1 or 0 depending on word understanding and richness of expression. *Retelling* was assessed with the use of six simple pictures which accompanied a story (one short story for all age groups). Each child was invited to listen carefully to the story that was presented orally to them and then were asked to retell it. The story had a simple story structure: three leading roles, an introduction, an event, a problem and a solution. The theme was interesting and compatible with the existing knowledge of most children (Buck, 2001) referring to a simple incident faced by the two children playing in a park. Children's ability to produce a *narrative* (without listening previously to a story sample) was evaluated by utilizing again six pictures and following a task format similar to the retelling task. Children were presented with simple pictures of scenes of a child's birthday party and were asked to observe them carefully and then make up a story about the specific sequence of pictures. Finally, in the *Story Comprehension questions* task, children listened to another short story accompanied by six pictures and were asked to answer simple questions about the heroes, their actions as well as their interpretation of story's events. Children's answers in *Story Comprehension questions* were subsequently coded for accuracy and inferencing with scores of 0 (false), 1 (insufficient or underdeveloped) or 2 (complete or developed) and a total scoring range of 0-12.

Coding of children's retellings and narrations. Children's retellings were analyzed according both to microstructure and macrostructure criteria. All of the criteria were scored with "0" as "minimum/ immature", with "1" as "developing", and with "2" as "sufficient / mature". The specific criteria which were evaluated for microstructure were: (a) use of conjunctions (Halliday & Hasan, 1976; McCabe & Rollins, 1994; Peterson & McCabe, 1991) and lexical cohesion (relevance and quality of words) (Halliday & Hasan, 1976). Story retellings with no use of conjunctions (i.e. labelling) scored with 0. Retellings using only few coordinating conjunctions (i.e. "and") received 1 point while retellings using coordinating, subordinating and/or correlative conjunctions (and contained more grammatically complex sentences) were scored with 2 points.

Macrostructure was evaluated with Story Grammar (Stein & Glenn, 1979) and temporal sequencing of actions (Hipfner-Boucher et al., 2014; Mandler & Johnson, 1977; Stein & Albro, 1997). The story grammar included (a) narrative introduction (convention and/ or place/ time/ heroes), (b) development of characters (state of mind and feelings of heroes), (c) problem reporting (how and/ or why and solution), and (d) result/ conclusion. Temporal sequencing refers to the episodic structure and succession of story elements. No score was assigned to retellings with errors in ordering of events and/or omissions while simple repetitions of events or minor gaps received 1 point. Retellings containing accurate story succession were scored with 2 points. Individual scores were created by summing the scores for all six elements (conjunctions/cohesion, temporal sequencing, introduction, character development, problem reporting, result/cohesion) and the total scoring range was 0-12. Coding and scoring of the retellings according to the aforementioned criteria was held by four raters in a sample of retellings (N=30) randomly selected. Agreement percentage was calculated to examine interrater reliability for each of the six elements as well as on the ratings of retelling accuracy and sophistication of vocabulary used. Interrater agreement was on average higher than 90%.

Procedures

Examiners were mostly postgraduate students of psychology or education who were extensively trained and evaluated to ensure uniform administration. Research coordinators established required contacts with the schools, organized testing schedules, contacted field observations of assessment procedures and supervised data collection. After obtaining written parental approval and child oral assent, examiners administered the tasks in two to three 40-minute sessions within a two-week period (in the context of a variety of other tasks not reported here). Assessments were conducted individually in a quiet room at the school. Breaks were provided as needed. Task administration was completed through an Android application for mobile devices (tablets) that was developed especially for this project in order to provide the participating children with a familiar and attractive way for completing the assessment. Development of the application aimed to produce a comprehensive language assessment battery that later could be commercially available. For this reason, it was carefully designed and constructed according to authors' detailed specifications in incorporating all task directions, practice examples and proper feedback for the examinees. Moreover, application development enabled the use of proper voicing and child-friendly graphic designs that were also developed especially for each language task. The touch-screens enabled direct recording of children's oral (narratives) and manual responses (response choices) ensuring high assessment fidelity. Data (language samples) were uploaded to a parallel web-based application and were scored by study's authors across language domains.

Analysis strategy

Initially we evaluated cross-sectional differences in performance on different language domains for both genders in all age groups as well as interrelations among domains and language tasks. Descriptive statistics using SPSS 25 for all tasks in each grade appear on Table 1. As expected, mean scores for each language skill tended to increase gradually from PK to G1.

Dimensionality of language measures was assessed within the context of Confirmatory Factor Analysis (CFA), a standard, widely used analytic approach to assess construct validity in theory-driven instrument construction (Li, 2016). Specifically, we compared five alternative models, separately in each grade, using multiple recommended criteria and approaches (Table 2). The models tested were nested because they had exactly the same number of indicators and differed only in the number of latent factors (one to five). The first model (A) was unidimensional. The second model (B) evaluated the validity of a distinction between morphological awareness, phonological awareness and processing from semantic and discourse skills (vocabulary, narrative and pragmatic skills). Potential separation of the first factor was examined on the basis that involved metalinguistic word-level skills. Model C examined separation of the phonological awareness/processing factor from a word level (lexical/semantic skill factor) and a broad factor of oral discourse skills (morphological knowledge, narrative and pragmatic skills). Nesting for this model was based upon evidence from the LARRC study that supported distinction between a "lower level" language factor (grammar and vocabulary) and the oral discourse factor in grade 1 (LARRC, 2015) (Figures 1-3 and supplemental material). Model D evaluated distinctiveness for the four factors (lexical/semantic and discourse skills, pragmatics, morphology and phonological awareness and processing) as an alternative to model C by nesting lexical/semantic skills with discourse skills based on previous evidence supporting their distinguishability at older children. Finally, the last model (E) evaluated distinctiveness of all five factors.

All analyses were conducted in Mplus version 8.0 (Muthén & Muthén, 2017) using maximum likelihood estimation with robust standard errors (MLR; Yuan & Bentler, 2000) to account for missing data, and deviations from normality. Given that all correlation coefficients between measures were below <.90 (Hair, Blanck, Babin, & Anderson, 2010), and variance inflation factor <10 (Kline, 2011) multicollinearity problems were not addressed further (see Supplementary Table S2). The amount of missing data on individual items was <1% across all age groups. Detailed information regarding procedures employed to handle missing data and outliers can be found in the Supplementary Material. Briefly, The MLR method was used for handling missing data because it uses the available information of all missing data patterns and provides estimates that are robust to nonnormality and/or independence of observations (Savalei, 2010). We did not allow any measurement errors to correlate in our measurement models as this practice might bias parameter estimates (Tomarken & Waller, 2003) and lead to misconceptions about the underlying structure of modeled relationships (Gerbing & Anderson, 1984).

Several complementary indices were used to assess model fit, including the Satorra-Bentler scaled chi square fit index, the Comparative fit index (CFI), the Non-normed fit index (NNFI, also known as the Tucker-Lewis index, TLI), the Root Mean Square Error of Approximation (RMSEA), the Standardized Root Mean Square Residual (SRMR) and the Akaike Information Criterion (AIC). Conservative cut-off criterion of CFI $\geq$.95 (Brown, 2006; Hu and Bentler, 1999), TLI $\geq$ .95, RMSEA $\leq$.05, and SRMR $\leq$.05 were adopted. AIC is a log likelihood measure of fit adjusting for the number of estimated parameters (Burnham & Anderson, 2004) with smaller value indicating better fit (Kline, 2013). CFI's incremental change was also used in comparison of nested models, with a value <.01 supporting the more parsimonious one (Moran, Marsh & Nagengast, 2013). In addition to the above model fit indices, the scaled χ^2 difference test ($\Delta\chi^2$) was estimated based on the procedure described by Satorra and Bentler (2001) to compare pairs of nested models.

Given the nature of our dataset, involving measures that are inherently strongly intercorrelated, assessment of convergent and divergent validity of the various structural models examined presented crucial challenges. To address these challenges in establishing convergent validity we computed several indices within in each model across all grades. Convergent validity was evaluated through an assessment of standardized indicator factor loadings (value $\geq$ .32; Tabachnick & Fidell, 2001) and their statistical significance (p <.05), followed by the evaluation of factor average variance extracted (AVE $\geq$ .05, Diamantopoulos & Siguaw, 2000) and Composite Reliability (CR $\geq$ .6, Diamantopoulos & Siguaw, 2000) coefficients. Discriminant validity was examined by reviewing pairwise factor correlations (Venkatraman, 1989) and by comparing the square root value of AVE (√AVE) with the correlation of latent constructs, a technique explained by Fornell and Larcker (1981). Discrimination of measures is ensured when the factor correlation coefficients are lower than .85 or more strictly .80 (Brown, 2006) and AVE is larger than both the maximum shared variance (MSV) and average share squared variance (ASV). Furthermore the √AVE for each latent variable should be higher than any of the bivariate correlations involving the latent variables in question (Fornell & Larcker, 1981) to evidence discriminant validity. In sum, model comparison was multifaceted, taking into account a constellation of model fit indices, tests of nested model comparisons and evaluation of convergent and discriminant validity of latent factors.

Results

Prekindergarden: Model fit indices. The results of one-dimensional (χ^2 = 438.963, *df* = 104, *p* <.05, RMSEA = .134, CFI = .772, TLI = .737, SRMR = .080), two-dimensional (χ^2 = 337.622, *df* = 103, *p* <.05, RMSEA = .112, CFI = .841, TLI = .814, SRMR = .072) and three-dimensional (χ^2 = 333.299, *df* = 101, *p* <.05, RMSEA = .113, CFI = .842, TLI = .812, SRMR = .066) models were below satisfactory values for analyzed indices, indicating an ill-fitting (see Supplemental material). Four-dimensional (χ^2 = 163.404, *df* = 98, *p* <.05, RMSEA = .061, CFI = .956, TLI = .946, SRMR = .052) and five-dimensional (χ^2 = 151.465, *df* = 94, *p* <.05, RMSEA = .058, CFI = .961, TLI = .950, SRMR = .050) models did not have an exact fit but they provided a close fit to the data with the differences between them being too subtle to justify the unequivocal support for one over the other (Table 3). Given that a model fits better the closer the χ^2 value is to the degrees of freedom (Thacker, Fields & Tetrick, 1989) five-structure solution could be considered more adequate as the discrepancy between the χ^2 value and the corresponding degrees of freedom (χ^2 = 151.465, *df* = 94) was smaller comparatively to the four-structure solution (χ^2 = 163.404, *df* = 98). Furthermore, a closer examination of RMSEA in four- ($RMSEA_{modelD}$ = .061, *p* =.136, CI [.044, .077]) and five-dimensional models ($RMSEA_{ModelE}$ = .058, *p* =.206, CI [.040, .075]) did not provide clear and strong evidence in favor of either model: Confidence Intervals (CI) were very similar across the two competing factor solutions (see above) and both *p* exceeded >.05 in order for RMSEA to be considered sufficient small (Brown, 2006). The only difference that could give a slight precedence to the five-structure solution was its slightly smaller RMSEA's critical value (ModelE = .058 vs ModelD =.061). Additionally, the small increase in CFI between the two nested models (CFI_{ModelE}-CFI_{ModelD} = .005 < .01) indicated better fit of the more parsimonious, four-dimensional model. Akaike's information criterion (AIC) was smaller for the five-structure model, indicating a better fit. TLI and SRMR values for the two models were quite similar, further complicating their comparison. Overall the collection of the aforementioned fit indices did not clearly reveal the best-fitting model, necessitating the use of additional criteria in order to find the better conceptualization of language dimensionality in PK.

Nested Model Comparisons. Given the lack of clear differences between the two nested models, we examined the adjusted χ^2 difference test (see right-hand section of Table 3) which was significant, $\chi^2(4)$ = 12.35, *p* <.05, indicating that the "larger", five-factor structure model fitted the data better than the "smaller" four-factor structure model. Nevertheless, given the sensitivity of the chi square test to sample size and other design features (Tomarken & Waller, 2003) we supplemented our analysis by examining the construct validity of the latent factors for the two aforementioned competing models.

Construct Validity of Factors. The majority of standardized indicator factor loadings across both models (Figure 1) ranged between .60 to .90 (p <.001), indicating good indicator reliability (Hair, Blanck, Babin, Anderson, & Tatham, 2006), with the exception of the MORPH_DER and the RETEL variables (with loadings ranging from .39 to .49). Nevertheless, they were retained in further model testing because they exceeded the recommended cut-off value of .32 (Tabachnick & Fidell, 2001). All latent constructs had CR values higher than the reliability suggested threshold of 0.6 (Diamantopoulos & Siguaw, 2000) indicating good internal consistency. Convergent validity problems (Table 4) were not detected across models as AVE's values exceeded .50 (Afthanorhan et al., 2014; Hair et al., 2010).

Insert Figure 1 about here

All but one factor correlations in the two Models (see Figure 1) were equal to or lower than the recommended threshold of .80 for establishing discriminant validity (Brown, 2006). In the five-structure solution the high intercorrelation found between DISCR and SEMANT latent dimensions ($r_{\mathbf{RDISCR\text{-}SEMANT}}$ = .91) called into question the degree of their distinctiveness. To examine this issue further we applied the technique proposed by Fornell and Lacker (1981) combined with the requirements that AVE>MSV and AVE>ASV. For the four-dimensional model (Table 4), the √AVE for the

SEM_DISCR factor was .71, a value substantially smaller than those indicated by two out of three bivariate correlations involving this construct ($r_{SEM\_DISCR-PRAGM}$ =.80, $r_{SEM-DISCR-MORPH}$ = .77). Furthermore AVE_{DISCR}= .50 was smaller than the corresponding MSV = .64 and ASV = .52 values. Based on these results, the discrimination of the DISCR-parent factor was not fully supported.

Similarly, in the five-dimensional model (Table 4), the DISCR and SEMANT latent constructs explained an average of 82% and 79% of the variance in the set of their respective indicators. Their √AVE values were smaller than their respective intercorrelation coefficient ($r_{DISCR-SEMANT}$ =.91) casting doubt on their distinctiveness. Additionally, comparisons of AVE, MSV and ASV values for the DISCR and the SEMANT respectively confirmed their partial discriminant validity: AVE_{DISCR} (.67) was smaller than MSV_{DISCR} (.83) but higher than the ASV_{DISCR} (.58). Although both models appear equivalent in terms of the aforementioned criteria, discriminant validity (Figure 1) was established only for the four-dimensional model because none of the factor correlations exceeded the proposed threshold of .80 (Brown, 2006).

Summary: PK Results. One- to three- dimensional language models were rejected because they did not achieve a valid model fit. On the other hand, four- and five-factor models provided adequate fit to the current data. Nevertheless, the differences between them were too small to justify the clear support of one over the other. An argument could be made in favor of the more complex five-structure model based strictly on significance of the χ^2_{diff} test. In addition, a closer evaluation of the fit criteria did not clearly point to the best fitted model: although RMSEA, TLI, SRMR and AIC indicated the five-dimensional model as representing a closer fit to the data, the small incremental change of CFI (<.01) pointed to the more parsimonious four-dimensional model (Moran et al., 2013). Furthermore, internal consistency (CR) and convergent validity (AVE) were equally supported between the two models, although discriminant validity based on AVE, MSV and ASV estimators and on √AVE and inter-construct correlations, raised questions regarding the distinctiveness of several factors. It appears, however, that the four-factor model may afford greater discriminant validity than the five-factor model given acceptable values for all factor intercorrelations in the former model whereas in the latter one factor correlation exceeded the proposed upper value of .85 (Brown, 2006). Taking into consideration all these approaches used in the assessment of models, the four-dimensional solution at PK seemed preferable to the five-dimensional one.

Kindergarten: Model Fit Criteria. The middle section of Table 3 contains model fit statistics for the five confirmatory factor models conducted in the kindergarten sample. Similar to the prekindergarten results, the first three dimensional models (Models A to C) could undoubtedly be rejected as ill-fitting. Comparison of the four- and five-dimensional models based on goodness of fit criteria, gave strong initial support to the five-structure model. Further model evaluation steps are described below.

Nested Model Comparisons. A statistical comparison of the four- versus five-factor models yielded preference to the five-factor solution, $\chi^2(4) = 43.127$, $p <.001$.

Insert Figure 2 about here

Construct Validity of Factors. Standardized factor loadings in both models were statistically significant with the majority of them exceeding .60. AVE results indicated similar problems with convergent validity for SEM_ DISCR (AVE_{SEM_DISCR} =.39) and DISCR (AVE_{DISCR} = .39) factors. As can be seen SEM_DISCR indicators in the four-dimensional model were identical to those defining the DISCR and SEMANT factors in the five-dimensional model. Notably, although the SEMANT factor displayed marginal but acceptable convergent validity (AVE_{SEMANT} = .53), which was further supported by CR values > 0.6 (Diamantopoulos & Siguaw, 2000).

Factor discriminant validity, as indicated by factor intercorrelation coefficients, was supported in both models with values ranging from .21 to .78. On the other hand, evaluation of discriminant validity based on the criterion proposed by Fornell and Larcker (1981) indicated the same problems with SEM_DISCR and DISCR parent-factor distinctiveness found in the PK dataset. More precisely, the SEM_DISCR was .62, a value slightly lower than its intercorrelation with PRAGM ($r_{SEM\_DISCR\text{-}PRAGM}$ = .65). The subtle problem with distinctiveness detected in this factor was confirmed through comparison of its AVE value with the respective MSV and ASV values: AVE for SEM_DISCR (.39) was bigger than ASV (.35) and smaller than MSV (.42) suggesting partial distinctiveness. Similarly, the DISCR (.62) and SEMANT (.73) in the five-structure model were lower their respective intercorrelations ($r_{DISCR\text{-}SEMANT}$ = .78. and $r_{DISCR\text{-}PRAGM}$ = .70) raising doubts regarding factors distinctiveness. Comparison of AVE, MSV and ASV estimators confirmed the partial discriminant validity of SEM_DISCR, DISCR and SEMANT factor in the four- and five-dimensional models. Overall the assessment of convergent and discriminant validity revealed quite similar deficiencies in both models. However, given the fact that the highest factor correlation was found in the five-factor solution ($r_{DISCR\text{-}SEMANT}$ = .78), one could give a slight precedence to the four-dimensional model.

Summary: K results. Examination of convergent and discriminant validity gave a slight but not unequivocal support to the four-dimensional model. On the other hand, the evaluation of models based on the goodness of fit criteria and the chi-square difference test pointed to the five-factor model. Thus, from a statistical perspective, the five-dimensional model seemed preferable to the four-dimensional one among kindergarten students.

First grade: Model fit criteria. Similar to the prekindergarten and kindergarten results, one– to three-dimensional models did not seem to fit the data due to unacceptable fit indexes. The four-dimensional model exhibited marginally allowable solutions while the five-dimensional model provided the most appropriate fit (e.g. significant RMSEA *p* value, higher CFI and TLI, lower SRMR and AIC). Thus, based on model fit criteria, five-structure model was considered as having the best fit to the G1 data.

Nested Model Comparison. There was a statistical difference between the four- and the five-dimensional models, $\chi^2(4) = 41.340$, $p < .001$. Thus, the representation of language in the five-dimensional model that separated the discourse and semantic skills fitted the data significantly better than the four-dimensional model which combined the indicators of the two aforementioned factors, into one (SEM_DISCR).

Insert Figure 3 about here

Construct Validity of Factors. Although there was a statistical support for the five-factor solution, inspection of AVE and CR estimators showed a clear preference for the four-dimensional model. As can be seen in Table 3, in the five-dimensional model, values were under the recommended thresholds for DISCR (CR = .59 and AVE = .28) was marginally allowable for SEMANT (AVE = .42), indicating poor convergent validity for these two factors. Moreover, the measurement capacity of the seven indicators of the SEM_DISCR was rather poor (AVE = .29), although standardized factor loadings (see Figure 3) were statistically significant (as in the four-factor model) with the majority exceeding >.60. Discriminant validity problems were not addressed in either model in view of the relatively small factor correlations (between .19 and .74).

Furthermore, for all the latent factors in each model were higher than the intercorrelations between these factors. Only SEM_DISCR factor in the four-dimensional model had slightly smaller (.54) than its correlation with the factor PRAGM ($r_{SEM\_DISCR-PRAGM}$ = .55). The questionable distinctiveness of the SEM_DISCR factor was partially confirmed by its AVE value (.29) being inconsiderably smaller than MSV value (.30). Additionally, the smaller value of AVE (.42) compared to MSV (.55) in factor SEMANT indicated possible deficiencies in regard of factor's discriminant validity. Nevertheless, the overall estimation of these criteria gave a slight support to the five-dimensional model.

Summary: G1 results. Based on model fit criteria, the chi-square difference test, and the construct validity results, the five-dimensional model was preferable, with a cautionary note regarding the moderate convergent validity of DISCR and SEMANT factors.

Discussion

The present cross-sectional study examined the dimensionality in oral language skills in Greek students attending Prekindergarten through First Grade. A key feature of the study entails use of a comprehensive battery including tasks assessing narrative discourse, pragmatics, and metalinguistic capacities (phonological awareness and processing) in addition to the commonly examined skills of lexical/semantic knowledge (vocabulary, listening comprehension), and word-level morphology. Confirmatory Factor Analyses were employed to compare five nested structural models ranging between one and five dimensions in each age group. By considering several complementary approaches to evaluate comparative model fit, results indicated that the four-dimensional model had slightly better discriminant validity in the data from preschoolers, although both the four- and five-dimensional models provided adequate fit to the data. Conversely, the five-dimensional model (distinguishing lexical/semantic from discourse skills) demonstrated better fit as compared to the four-dimensional model in the data from kindergarten and first grades indicating a trend toward increased distinctiveness of language domains in this developmental period.

Developmental Changes in Language Dimensionality

The results of previous studies assessing developmental changes in the dimensional structure of language skills are characterized by notable discrepancies, to a large extent due to significant differences in the range of skills assessed. In the earliest study addressing this topic, data from a battery of standardized tests were best described by a single underlying dimension in K through grade 4 (Tomblin & Zhang, 2006) with a viable two-dimensional model first evident in grade 8 (grammar and semantics). The battery of tests employed in the subsequent LARRC study included tasks assessing discourse skills (LARRC, 2015a). Results supported a unidimensional structure of language skills in PK and kindergarten, although the distinction between a “lower level” language factor (grammar and vocabulary) and the oral discourse factor was first evident as early as in grade 1. A three-dimensional model was supported in grade 3 (grammar, vocabulary, discourse). In a more recent, large-scale study utilizing a comprehensive array of tasks assessing semantic knowledge (including depth and breadth of vocabulary and oral comprehension) and morphosyntactic knowledge and use, the two-dimensional model featuring grammar and semantic factors was optimal as early as preschool, as compared to the single-domain model (Lonigan & Milburn, 2017). Thus, with a notable exception (Tomblin & Zhang, 2006), semantics, grammar/morphology, and discourse have emerged as psychometrically distinct language domains, although the precise age at which this takes place varied across studies. This distinction was clearly evident in the data from preschool through Grade 1 children in the present study. Such evidence is consistent with a conceptualization of the language system as comprised of two partially independent subsystems responsible for processing lexical representations and rules, respectively.

Our data further suggest that a third domain, *oral discourse* may not be properly distinguishable from lexical/semantic skills until kindergarten. This finding is consistent with the notion that the development of oral discourse skills follows the consolidation of foundational skills, and comprises a higher and more complex level of language processing that develops later as a result of literacy experiences and schooling (LARRC, 2015a). In the only other study that examined oral discourse, differentiation emerged a year later (1st grade) (LAARC, 2015a). The observed time difference between the two studies, could be explained by dissimilarities in the tasks selected to assess oral discourse. In the LAARC (2015a) study, three of the used tasks focused upon comprehension monitoring and inference skills and only one task in narrative structure. In contrast, we assessed oral discourse skills through their ability to comprehend short stories presented orally (and to answer explicit and implicit questions), by evaluating specific characteristics of story retellings and narrations (i.e., coherence, temporal succession, cohesion, etc.), as well as by rating their expressive competence in defining words. Greek Kindergarten students were understandably showing facility with literal comprehension and narration tasks as in the Greek educational system, devote the largest amount of their instructional time interacting with stories and recitations. On the other hand, self-monitoring and inferencing abilities that were tested in the LARRC (2015a) study, are generally considered as more complex and higher-level abilities cultivated by formal literacy learning when lower level skills (vocabulary and grammar) are integrated and well specified (Hogan, Bridges, Justice & Cain, 2011). In general, evidence provided by both studies support the aforementioned notion of a hierarchical development of language skills from words to sentences to discourse (Tomblin & Zhang, 2006). We hypothesize that the basic oral discourse factor and relates to the literal understanding of narratives is distinguishable from vocabulary and grammar abilities as early as in Kindergarten. Conceptualization of a higher-level discourse factor that involves more complex skills (as inferencing and comprehension monitoring) is expected to start becoming distinct from the other language factors at the time of school entry.

Lexical/semantic knowledge is considered a core language domain in the context of investigations of language development and subsequent literacy learning. As noted by many, it is difficult to assess semantic development without implicating other language or cognitive abilities (Pinker, 1997). In this study, we conceptualized semantic knowledge as it is measured directly through three tasks which assessed children's vocabulary through especially constructed pictures and photos using both receptive and expressive modalities. Another vocabulary-related task of defining words was included in the discourse construct because of the cognitive demands in forming an appropriate definition for any given word, and the scoring criteria used for rewarding richness of expression. Such task appears to be much broader than an estimate of extend of semantic representations and categorizations among them. The selected option however, contributed to the high correlation found between the semantic and the discourse factors which at the preschool period do not emerge as distinct. This finding, however, could also be an indication of overreliance of oral discourse upon word knowledge that is expected in younger children (Rowe, 2012). The diminishing correlation between the two factors between preschool and first grade may indicate increasing independent variance within the discourse skills as children achieve basic linguistic competence. Support for substantial variability in the growth trajectories between vocabulary and the other language dimensions has also been provided by previous work that has examined in addition effects of developmental and socio-educational factors (Campisi, Serbin, Stack, Schwartzman & Ledingham, 2009; Rowe, 2012).

A further note is in order regarding word morphology which was assessed using a different format than in former investigations. Specifically, we focused on implicit knowledge of derivative formation and inflections in verbs and nouns by assessing ability to detect and manipulate morphemes in pseudowords embedded in sentences. As mentioned previously, Greek is a language with rich morphology, both inflectional and derivational (Ralli, 2003). Nouns and adjectives require obligatory inflection for gender, number and case, realized through suffixation. Verbs are inflected for tense, number, person, voice and aspect (Ralli, 2003). Different rules apply for nouns, adjectives and verbs, as regards their suffixes and alternations of stem (Holton et al., 2012). The formation of nouns, adjectives and verbs also relies on derivation. Nouns can be based on the stems of verbs and vice versa (Ralli, 2005). Awareness of inflectional processes has been found to be achieved by the first school years, which is earlier than that of derivational processes, which develop at a slower pace (Diakogiorgi, Baris & Valmas, 2005; Diamanti, Mouzaki, Ralli, Antoniou, Papaioannou & Protopapas, 2017). In addition, derivation in Greek usually relies on semantic cues, which are not readily available were pseudowords to be used. Findings across age groups in the present study support the psychometric independence of this factor that consists a marker of a meta-linguistic skill based on spoken language experience and it could be considered as a precursor of conscious morphological skills. Such evidence could be explained partly by participants' specific language characteristics but also could be an indicator for the developmental progression of metalinguistic awareness skills. As it has been previously shown, the rich morphological system of Greek language facilitates the development of morphological awareness at the implicit/epilinguistic level well before the onset of formal literacy instruction (Diamanti et al., 2017). Moreover, it seems that implicit word morphology consists a factor distinguishable from other observed indicators of linguistic competence even from an early age.

An important feature of the present study is the inclusion of tasks assessing pragmatic and phonological skills. It should be noted that phonological and pragmatics do not contradict previous findings as these studies haven't included relevant measures (Tomblin & Zhang, 2006, LARRC, 2015a, Lonigan & Milburn, 2017). Awareness and processing of phonemes in spoken words (which has been studied extensively for its contribution to reading outcomes), appeared to comprise a distinct factor across age groups. Phonological awareness is another metalinguistic skill that have been found to be unidimensional as it is measured by aggregated tasks similar to those used in this study (Anthony & Lonigan, 2004; Anthony & Francis, 2005; Papadopoulos, Kendeou & Spanoudis, 2008; Papadopoulos, Spanoudis & Kendeou, 2009; Schatschneider, Francis, Foorman, Fletcher & Mehta, 1999). Despite the considerable correlation ($r = .48$-$.58$) between the morphological and phonological factors that has been found in all age groups, both factors are better conceptualized separately. This finding clearly suggests a qualitative difference in the ability for conscious manipulations of spoken word segments that relates to implicit knowledge of grammatical rules. Furthermore, separation of both morphological and phonological factors from semantic, discourse and pragmatic factors, specifies the finer distinction in the observed variance of word-related knowledge (word-formation from word-meaning and word-using skills). The found distinction between metalinguistic and the other linguistic factors since the preschool period consists an index of the emerging awareness for language use and structure and clarifies the predictive relation of the respective skills to later reading abilities. Both morphological and phonological constructs are associated with future reading skills and it has been shown that they share variance in their substantial contribution to literacy outcomes (Diamanti, et al, 2017; Manolitsis, 2006).

Results substantiating the distinction of the *pragmatics* factor is another significant finding which indicates that pragmatic language aspects should also be examined in investigations of language development. Early in development children participate in conversations, practicing social uses of language (e.g., turn-taking) and managing relevant speech acts (e.g., requesting) in order to maintain social interactions with adults and other children (Snow, Pan, Imbens-Bailey & Herman, 1996). Pragmatic competence is achieved through orchestration of an extensive set of cognitive, linguistic and social skills including conversation initiation and topic management, coherent production of utterances, efficient use of contextual cues, understanding of non-literal language etc. (Matthews, Biney & Abbott-Smith, 2018). Narrative construction ability has also been examined as an index of pragmatic ability, even though it actually extends outside the scope of pragmatics. In sum, pragmatic skills are affected by an array of factors that are often difficult to measure as standardized tests are somewhat insensitive to children's problems in real-life communication.

This study utilized a task based on well-established notions regarding the nature of pragmatic skills evaluating both linguistic and social pragmatic aspects of language such as interpretation, intent, interaction, informativity, request, politeness, etc. This effort aimed to assess associated skills and to provide some useful insight on how young children master particular aspects of pragmatic language as well as the distinction of the respective factor across age groups. According to model fit criteria, separation of the pragmatic factor was substantiated for all children providing further support for language dimensionality in this age group. The strong association between the discourse and pragmatics factors was expected as the discourse factor relies heavily upon children's narrative production that was evaluated for thematic coherence and linguistic cohesion. Retelling and narration skills constitute crucial elements of communication ability and are thus closely related to pragmatic skills (de Villiers, 2004). The distinction of the pragmatic factor, however, is clearly supported for all children with increasing independent variance as children grow and become more competent in fine-tuning their language according to perceived demands of the social and communicative context. Therefore, evaluation of pragmatic skills should be included in language assessments from early on. Tasks that elicit language production through proper pictorial materials could prove more useful than simple narrative samples (De Villiers, 2004). Material could combine linguistic with socio-cognitive aspects (Adams 2002) including (a) questions related to the comprehension and interpretation of a communicative situation; (b) vignettes representing interactional skills related to the contextual variation, as well as (c) items that can highlight the child's intention/ ability to communicate. Implementation of an interviewing approach for parents and caregivers could also provide samples of child's pragmatic language skills in various communicative contexts that could not be observed otherwise (Dewart & Summers, 1995).

Overall, the current findings support the multidimensional structure of Greek language comprising psychometrically distinct dimensions that correspond to phonological, semantic, morphological, pragmatic and discourse skills and are evident from an early age. Similar evidence has been provided by only one other study with

Spanish speaking students (LARRC, Yeoman-Maldonado, Bengochea & Mesa, 2018) highlighting language morphology as a distinguishing linguistic ability as early as prekindergarten. This finding should be further explored in non-Indo-European languages such as Cantonese, Mandarin and Korean that have very limited inflectional morphemes and derivational processes while having a more extensive compounding morphology (McBride-Chang, Tardif, Cho, Shu, Fletcher, Stokes, Wong & Leung, 2007). It seems that learners of languages with prevalent morphological forms (such as Greek and Spanish) grasp grammatical features earlier than English learners without being depended upon parallel lexical growth (Devescovi, Caselli, Marchione, Pasqualetti, Reilly & Bates, 2005). This multidimensional view of language is in agreement with existing linguistic notions and assessment practices followed in clinical and educational settings in order to identify areas of weaknesses and plan proper interventions.

Limitations and Future Directions

Several limitations of this study would be considered. First, the sample size of 800 participants is likely associated with excessive power levels for the omnibus chi-square tests of model fit. Consequently, some tests of omnibus but also relative model fit (e.g, difference chi-square tests) were likely artifacts of Type-I errors. Second decisions on unidimensionality and distinctiveness were based on one methodology (i.e., AVE) although several other alternatives are available (e.g., DIMTEST, Gorsuch's protocol, etc).

The small number of items in some tasks may have affected the sensitivity of certain tasks to measure the intended constructs. It should be noted however that rigorous pilot testing of a more extensive version of Logometro indicated that the majority of eliminated items had largely overlapping variances with items preserved.

Finally, results of this study extend previous investigations that have examined language dimensionality in young children. However, it is very important to examine the same relationships in longitudinal studies that follow the same group of children in order to strengthen the above assumptions.

In conclusion, our findings regarding multidimensionality of oral language ability at this phase of development have significant practical implications as they support the current practice for evaluating different language abilities and highlight the need for making finer distinctions within the language domain. Differentiating between language factors will continue enabling us to fully appreciate the onset of development and progression of the critical skills related to children’s language and literacy development. Moreover, it may allow early identification of risk factors related to reading and spelling difficulties and enable educators to enhance those skills at a very young age.

References

Adams, C. (2002). Practioner Review: The assessment of language pragmatics. *Journal of Child Psychology and Psychiatry*, *43*(8), 973-987.

Afthanorhan, W.M.A.B.W, Ahmad, S. & Mamat, I. (2014). Pooled Confirmatory Factor Analysis (PCFA) Using Structural Equation Modeling on Volunteerism Program: A Step By Step Approach. *International Journal of Asian Social Science*, *4*(5), 642-653.

Angsuchoti, S., Wijitwanna, S., & Pinyopanuwat, R. (2008). Statistics for social science and behavioral science: How to use LISREL. Bangkok, Thailand: Mission Media [in Thai].

Anthony, J. L., & Lonigan, C. J. (2004). The Nature of Phonological Awareness: Converging Evidence from four Studies of Preschool and Early Grade School Children. *Journal of Educational Psychology*, *96*(1), 43-55. http://dx.doi.org/10.1037/0022-0663.96.1.43.

Anthony, J. & Francis, D. (2005). Development of Phonological Awareness. Current Directions in Psychological Science. *Current Directions in Psychological Science, 14*, 255-259. http://dx.doi.org/10.1111/j.0963-7214.2005.00376.x.

Barrett, P. (2007). Structural Equation Modelling: Adjusting Model Fit. *Personality and Individual Differences, 42*(5), 815-824.

Bates, E., & Goodman, J. C. (2001). On the inseparability of grammar and the lexicon: Evidence from acquisition. In M. Tomasello & E. Bates (Eds.), *Essential readings* *in developmental psychology*. *Language development: The essential readings* (p. 134–162). Blackwell Publishing.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

Bentler, P. M. (2006). *EQS 6, Structural Equations Program Manual*. Encino, CA: Multivariate Software Inc.

Best, W., (2005). Investigation of a new intervention for children with word-finding problems. *Journal of Language Communication Disorders, 40*, 279-318.

Brian, J.D. (1990). *Morphology and Universals in Syntactic Change. Evidence from Medieval and Modern Greek*. New York/London: Garland Publishing, Inc.

Bowey, J. A. (1994). Phonological sensitivity in novice readers and nonreaders. *Journal of Experimental Child Psychology*, *58*, 134-159.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York, NY: Guilford.

Browne, M. W. & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research, 21*(2), 230-258.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Burnham, K. & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research, 33*, 261-304.

Byrne, B. M. (2012). Structural equation modeling with Mplus. *Basic concepts, applications, and programming*. New York: Routledge.

Campisi, L., Serbin, L. A., Stack, D. M., Schwartzman, A. E., & Ledingham, J. E. (2009). Precursors of language ability and academic performance: An intergenerational, longitudinal study of at-risk children. *Infant and Child Development*, *18*(5), 377-403. http://dx.doi.org/10.1002/icd.628.

Catts, H. W., Fey, M. E., Zhang, X. & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*(4), 331-361. http://dx.doi.org/10.1207/s1532799xssr0304_2.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance,. *Structural Equation Modeling, 14*(3), 464-504.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes*, *16*(2-3), 113-142. http://dx.doi.org/10.1080/01690960042000012.

Conti-Ramsden, G. & Durkin, K. (2012). Language Development and Assessment in the Preschool Period. *Neuropsychology Review, 22*, 384-401.

Devescovi, A., Caselli, M., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language, 32*, 759–786.

De Villiers, J. (2004). Assessing Pragmatic Skill in Elicited Production. *Seminars in speech and language, 25*, 57-71.

Dewart, H. & Summers, S. (1995). *The pragmatics profile of everyday communication skills in children*. NFER-Nelson.

Diakogiorgi, K., Baris, T. & Valmas T. (2005). Ικανότητα χρήσης μορφολογικών στρατηγικών στην ορθογραφημένη γραφή από μαθητές της Α′τάξης του δημοτικού [Ability to use morphological strategies in spelling by children in the first elementary grade*]*. *Psychology, 12,* 568–586.

Diamanti, V., Benaki, A., Mouzaki, A., Ralli, A., Antoniou, F., Papaioannou, S. & Protopapas, A. (2017). Development of early morphological awareness in Greek: Epilinguistic versus metalinguistic and inflectional versus derivational awareness. *Applied Psycholinguistics, 39*(3), 545-567.

Diamantopoulos, A. & Siguaw, A. D. (2000). *Introducing LISREL: A guide for the uninitiated*. London: Sage Publications.

Dickinson, D. K., McCabe, A., Anastopoulos, L., Peisner-Feinberg, E. S. & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95,* 465–481.

Dockrell, J. (2001). Assessing language skills in Preschool children. *Child Psychology & Psychiatry Review*, *6*, 74-85.

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

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., Francis, D. J., Fowler, A. E. & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*, 1-23.

Fletcher, J. M., Foorman, B. R., Boudousquie, A., Barnes, M. A., Schatschneider, C., & Francis, D. J. (2002). Assessment of reading and learning disabilities: A research-based intervention-oriented approach. *Journal of School Psychology, 40*(1), 27-63. http://dx.doi.org/10.1016/S0022-4405(01)00093-0.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A. & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*(1), 3-17. http://dx.doi.org/10.1037/0022-0663.88.1.3.

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

Fornell, C. & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research, 18*(1), 39-50.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research, 11*, 572-580.

Hair J. F., Blanck., W. C., Babin., B. J., Anderson, R. E. & Tatham., R. (2006). *Multivariate Data Analysis*. New Jersey: Pearson International Edition.

Hair, J., Blanck, W. C., Babin, B. J. & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper saddle River, New Jersey: Pearson Education International.

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman Publishing.

Holton, D., Mackridge, P., Philippaki-Warburton, I., & Spyropoulos,V. (2012). *Greek: A Comprehensive Grammar* – 2nd Edition. London and New York: Routledge.

Hooper, D., Coughlan, J. & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hipfner-Boucher, K. & Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J. & Girolametto, L. (2014). Relationships between preschoolers' oral language and phonological awareness, *First Language, 34*, 178-197, http://dx.doi.org/10.1177/0142723714525945.

Hogan, T., Bridges, M. S., Justice, L. M. & Cain, K. (2011). Increasing Higher Level Language Skills to Improve Reading Comprehension. *Focus on Exceptional Children. 44*. http://dx.doi.org/10.17161/fec.v44i3.6688.

Holton, D., Mackridge, P. & Philippaki-Warburton, I. (2012). *Greek: A Comprehensive Grammar* – 2nd Edition. London and New York: Routledge.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53–60.

Hu, L.-T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Kanellou, M., Korvesi, E., Ralli, A., Mouzaki, A., Antoniou, F., Diamanti, V., & Papaioannou, S. (2016). Narrative skills in preschool and first grade children. *Preschool and Primary Education*, *4*, 35-67. (In Greek)

Kim, N. (2015). Tests Based on Skewness and Kurtosis for Multivariate Normality. *Communications for Statistical Applications and Methods, 22,* 361-375.

Kim, Y.-S., Al Otaiba, S., Folsom, J., Greulich, L. & Puranik, C. (2014). Evaluating the Dimensionality of First-Grade Written Composition. *Journal of speech, language, and hearing research, 57*, 199-211. http://dx.doi.org /10.1044/1092-4388 (2013/12-0152).

Klem, M., Gustafsson, J.E. & Hagtvet. B. (2015) The Dimensionality of Language Ability in Four-Year-Olds: Construct Validation of a Language Screening Tool. *Scandinavian Journal of Educational Research, 59*, 195-213.

Kline, R. B. (1998). *Methodology in the social sciences. Principles and practice of structural equation modeling*. New York, NY, US: Guildford Press.

Kline, R. B. (2011). *Principles and practice of structural equation modelling*. New York: The Guildford Press.

Kline, R. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher, C. Schatschneider, & D. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 171-207). New York, NY: Rootledge.

Kwak, S. K. & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. *Korean Journal of Anesthesiology*, *70*(4), 407-411.

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

Language and Reading Research Consortium (LARRC) (2015b). The dimensionality of Spanish in young Spanish-English dual-language learners. *Journal of Speech, Language, and Hearing Research, 58*, 754-766.

Language and Reading Research Consortium (LARRC) (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language, and Hearing Research*, *60*(5), 1273-1284.

LARRC, Yeoman-Maldonado, G., Bengochea, A. & Mesa, C. (2018). The dimensionality of oral language in Kindergarten Spanish-English dual language learners. *Journal of Speech, Language, and Hearing Research, 61*, 2779-2795.

Li, C-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936-949.

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in 2- to 5-year-old children. *Journal of Educational Psychology*, *90*(2), 294-311. http://dx.doi.org/10.1037/0022-0663.90.2.294.

Lonigan, C. J., Burgess, S. R., Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*, 596–613.

Lonigan, J. C. & Milburn, T. F. (2017). Identifying the Dimensionality of Oral Language Skills of Typically Developing Preschool through Fifth-Grade Children. *Journal of Speech Language and Hearing Research, 60*. http://dx.doi.org/10.1044/2017_JSLHR-L-15-0402.

Looney, S. W. (1995). How to use tests for univariate normality to assess multivariate normality. *American Statistician*, *49*(1), 64-70.

Mandler, J. & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall, *Cognitive Psychology*. *9*, 111-151, http://dx.doi.org/10.1016/0010-0285(77)90006-8.

Manolitsis, G. (2006). "Η σχέση της μορφολογικής επίγνωσης με την αναγνωστική ικανότητα κατά τα πρώτα στάδια ανάπτυξής της. [The relationship of morphological awareness with early reading ability]," in *Proceedings of the 26th annual meeting of the Department of Linguistics. Faculty of Philosophy: Studies in Greek linguistics*, eds D. Koutsogianis, A. Haralabopoulos, and N. Liosis (Thessaloniki: Aristotle University of Thessaloniki), 282–293.

Mardia, K.V. (1995). How to use tests for univariate normality to assess multivariate normality, *American Statistician*, *49*(1), 64-70. http://dx.doi.org/10.1080/00031305.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrica*, *57*(3), 519-530. http://dx.doi.org/10.2307/2334770.

Mastropavlou, M (2006). *The role of phonological salience and feature interpretability in the grammar of typically developing and language impaired children*. Unpublished doctoral dissertation, Aristotle University of Thessaloniki.

Matthews, D., Biney, H. & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition and exceutive functions. *Language Learning and Development*, *14*(3), 186-223. http://dx.doi.org/10.1080/15475441.2018.1455584.

McBride-Chang, C., Tardif, T., Cho, J.-R., Shu, H., Fletcher, P., Stokes, S.F., Wong, A. & Leung, K. (2007). What's in a word? Morphological awareness and vocabulary knowledge in three languages. *Applied Psycholinguistics*, *29*, 437-462.

McCabe, A., & Rollins, P. R. (1994). Assessment of preschool narrative skills. *American Journal of Speech-Language Pathology*, *3*(1), 45–56.

McGregor, K.K., Berns, A.J., Owen, A.J., Michels, S.A., Duff, D., Bahnsen, A.J. & Lloyd, M. (2012). Associations between syntax and lexicon among children with or without ASD and language impairment. *Journal of Autism and Developmental Disorders*, *42*, 35-47. http://dx.doi.org/10.1007/s10803-011-1210-4.

McGuire, B. & Tinsley, H. E. A. (1981). A Contribution to the Construct Validity of the Tennessee Self-Concept Scale: A Confirmatory Factor Analysis. *Applied Psychological Measurement*, *5*(4), 449-457.

Mouzaki, A., Ralli, A., Antoniou, F., Diamanti, V., Papaioannou, S., (2017). *Logometro*. Athens: Intelearn.

Moran, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 1-10). Charlotte, SC: IAP.

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz. S. E., Lyon, G. R. & Shankweiler, D. P. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology*, *90*, 347-373.

Muthén, L. K., & Muthén, B. O. (2017). Mplus user's guide: Statistical analysis with latent variables. *Statistical analysis with latent variables*. Version. Retrieved from http://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf.

O'Connor, R. E. & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, *3*, 159–197.

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. http://dx.doi.org/10.1007/s11145-008-9159-1.

Papadopoulos, T. C., Spanoudis, G., & Kendeou, P. (2009). The dimensionality of phonological abilities in Greek. *Reading Research Quarterly*, *44*, 127-143.

Park, I., & Schutz, R. W. (2005). An introduction to latent growth models: analysis of repeated measures physical performance data. *Research Quarterly for Exercise and Sport*, *76*(2), 176-192.

Peterson, C., & McCabe, A. (1991). Linking children's connective use and narrative macrostructure. In A. McCabe & C. Peterson (Eds.), *Developing narrative structure* (pp. 29-53). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Pinker, S. (1997). *How the Mind Works*. London: Penguin Books.

Protopapas, A. (2017). Learning to read Greek. In L. T. W. Verhoeven & C. A. Perfetti (Eds.), Learning to read across languages and writing systems (pp. 155–180). Cambridge University Press. doi: 10.1017/9781316155752.008

Protopapas, A., Sideridis, G. D., Simos, P. G. & Mouzaki, A. (2007). The development of lexical mediation in the relationship between text comprehension and word reading skills in Greek. *Scientific Studies of Reading*, *11*, 165–197.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S. & Frith, U. (2003), *Brain*, *126*, 841-865.

Ralli, A. (2003). Morphology in Greek linguistics: The state of the art. *Journal of Greek Linguistics*, *4*, 77–129.

Ralli, A. (2005). *Μορφολογία [Morphology]*. Athens, Greece: Patakis.

Ralli, A., Kazali, E., Kanellou, M., Mouzaki, A., Antoniou, F., Diamanti, V. & Papaioannou, S. (submitted). Exploring the contribution of oral language in narrative ability through retelling during early school years. *Journal of Psycholinguistic Research*.

Rapala, M. M. & Brady, S. (1990). Reading ability and short-term memory: The role of phonological processing. *Reading and Writing, 2*. 1-25. http://dx.doi.org/10.1007/BF00383371.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development 83*, (5), 1762–1774

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrica, 66*, 507-514.

Savalei, V. (2010). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods, 15,* 352-367. https://dx.doi.org/10.1037/a0020143.

Scarborough, H. (1998). Predicting the future achievement of second graders with reading disabilities: contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, *48*(1), 115–136. https://doi.org/10.1007/s11881-998-0006-5.

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91*(3), 439-449, http://dx.doi.org/10.1037/0022-0663.91.3.439 .

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. New Jersey: Lawrence Erlbaum Associates.

Sheskin, D. J. (2004). *Handbook of parametric and nonparametric statistical procedures* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Snow, C. E., Pan, B., Imbens-Bailey, A. & Herman, J. (1996). Learning how to say what one means: A longitudinal study of children's speech act use. *Social Development, 5*, 56– 84.

Stanovich, K. E. & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*(1), 24-53. http://dx.doi.org/10.1037/0022-0663.86.1.24

Stein, N. & Glenn, C. (1979) An Analysis of Story Comprehension in Elementary School Children, in R. Freedle (Ed.) *New Directions in Discourse Processing*, Nor-wood, NJ: Ablex Publishing.

Stein, N. L., & Albro, E. R. (1997). Building complexity and coherence: Children's use of goal-structured knowledge in telling stories. In M. G. W. Bamberg (Ed.), *Narrative development: Six approaches* (pp. 5-44). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Stone, B. & Brady, S. (1995). Evidence for phonological processing skills in less-skilled readers. *Annals of Dyslexia, 45*. 51-78. http://dx.doi.org/10.1007/BF02648212.

Suggate, S., Reese, E., Lenhard, W. & Schneider, W. (2014). The relative contributions of vocabulary, decoding, and phonemic awareness to word reading in English versus German. *Reading and Writing: An Interdisciplinary Journal, 27*(8), 1395-1412. http://dx.doi.org/10.1007/s11145-014-9498-z.

Tabachnick, B. G. & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Boston: Allyn & Bacon.

Thacker, J. W., Fields, M. W. & Tetrick, L. E. (1989). The factor structure of Union Commitment: An application of confirmatory factor analysis. *Journal of Applied Psychology*, *74*(2), 228-232.

Tomarken, A. J. & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology, 112*, 578-598.

Tomblin, J. B. & Zhang, X. (2006). The dimensionality of language ability: An application of item response theory. *Journal of Speech, Language, and Hearing Research, 49,* 1193–1208.

Torgesen, J. K.,Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27*, 276-286.

Torgesen, J. K. (1996). *Phonological awareness: A critical factor in dyslexia*. Baltimore: Orton Dyslexia Society.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology*, *40*(1), 7-26. http://dx.doi.org/10.1016/S0022-4405(01)00092-9.

Tsimpli, I. (2014). Early, late or very late?: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism, 4,* 283–313.

Tsimpli, I. & Hulk, A. (2013). Grammatical gender and the notion of default: Insights from language acquisition. *Lingua, 137,* 128–144.

Yuan, K.-H., & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. *Sociological Methodology, 30*, 165-200.

Varlokosta, S. & Nerantzini, M. (2013). Grammatical gender in Specific Language Impairment: Evidence from determiner-noun contexts in Greek. *Psychology, 20*, 338-357.

Varlokosta, S. & Nerantzini, M. (2015). The acquisition of past tense by Greek-speaking children with Specific Language Impairment: The role of phonological saliency, regularity and frequency. In S. Stavrakaki (Ed.), *Specific Language Impairment: Current trends in research*. pp. 253-286. Amsterdam: John Benjamins Publishing.

Vellutino, F. R., Scanlon, D., M. & Lyon, R. G. (2000). Differentiating Between Difficult-to-Remediate and Readily Remediated Poor Readers More Evidence Against the IQ-Achievement Discrepancy Definition of Reading Disability. *Journal of learning disabilities. 33*, 223-38, http://dx.doi.org/10.1177/002221940003300302.

Venkatraman, N. (1989). Strategic Orientation of Business Enterprises: The Construct, Dimensionality and Measurement. *Management Science, 35*(8), 942-962.

Yuan, K.-H., & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. *Sociological Methodology, 30*, 165-200.

Wagner, R., Torgesen, J., Laughon, P., Simmons, K., & Rashotte, C. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85*, 83–103.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R. & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: a 5-year longitudinal study. *Developmental Psychology*, *33*(3), 468–479.

Wolf, M., Bowers, P. & Biddle, K. (2000). Naming-Speed Processes, Timing, and Reading A Conceptual Review. *Journal of learning disabilities, 33,* 387-407. http://dx.doi.org/10.1177/002221940003300409.

Wood, F., Hill, D., & Meyer, M. (2001). *Predictive Assessment of Reading*. Winston-Salem, NC.

Zhou, M. & Shao, Y. (2014). A Powerful Test for Multivariate Normality. *Journal of Applied Statistics*, *41*(2), 351-363.http://dx.doi.org/10.1080/02664763.2013.839637.

Table 1.
Descriptive Statistics by Measure and Grade (Means, standard deviations, skewness and kurtosis)

	Prekindergarten				Kindergarten				Grade 1			
Factors with Indicators	Total score Mean	Total score SD	Skew	Kurtosis	Total score Mean	Total score SD	Skew	Kurtosis	Total score Mean	Total score SD	Skew	Kurtosis
NAR	7.38	2.78	0.07	0.63	8.76	2.93	- 0.26	0.75	9.72	2.69	- 0.60	0.59
RETEL	5.04	2.74	0.00	- 0.63	5.66	2.81	- 0.23	- 0.23	9.36	3.46	- 0.22	0.56
ST_COMP	7.38	2.43	- 1.00	1.37	8.37	1.96	- 0.77	0.92	9.20	1.66	- 0.56	- 0.02
VOC_DEF	21.06	8.48	- 0.01	- 0.25	26.34	7.88	- 0.19	- 0.09	31.70	7.23	0.13	- 0.31
CONT_INTER	5.43	1.96	- 0.66	- 1.27	6.27	1.90	- 0.49	0.97	6.73	1.90	- 0.01	0.99
CONT_RESP	6.84	3.42	- 0.10	- 0.45	8.59	3.35	- 0.13	- 0.15	9.36	3.46	- 0.30	- 0.33
COM_INT	9.67	4.58	0.20	0.60	11.37	3.92	- 0.24	- 0.09	14.46	4.36	0.32	1.21
MORPH_DER	1.50	0.84	0.59	- 0.61	1.93	0.93	- 0.11	- 1.35	2.53	0.75	- 1.36	0.80
MORPH_INFLV	3.78	3.23	0.01	- 1.67	5.00	2.92	- 0.64	- 1.06	6.72	1.82	- 1.78	2.90
MORPH_INFLN	3.82	2.97	- 0.02	- 1.53	4.73	2.85	- 0.46	- 1.24	6.90	1.65	- 2.15	5.17
LIST_COMP	11.31	2.94	- 0.98	1.68	12.95	2.07	- 0.50	- 0.17	14.02	1.68	-.1.21	2.38
REC_VOC	17.99	6.61	- 0.60	- 0.28	22.93	4.98	- 1.68	4.20	25.60	2.74	- 0.79	0.32
NAM	12.39	3.56	- 0.44	- 0.36	14.45	2.95	- 0.74	0.82	16.34	2.15	- 0.55	0.13
INIT_PHON	3.47	2.03	0.01	- 0.68	4.89	2.04	0.99	- 0.68	6.20	1.23	- 1.78	2.88
BL_PHON	0.80	1.54	2.10	3.37	1.63	2.13	0.15	0.38	4.84	2.06	- 1.02	0.13
PHON_EL	0.72	1.68	2.39	4.65	1.63	2.13	1.25	0.38	4.84	2.06	- 0.85	- 0.72
Multivariate normality Mardia's test (c.r)	3.466				3.153				14.087			

Note. NAR=narrative; RETEL=retelling; ST_COMP=story comprehension; VOC_DEF=vocabulary definition; CONT_INTER=context interpretation; CONT_RESP=context proper response; COM_INT=communicative intent; MORPH_DER=implicit understanding; MORPH_INFLV=implicit understanding of verb inflections; MORPH_INFLN= implicit understanding of noun inflections; LIST_COMP=listening comprehension (sentences); REC_VOC=receptive vocabulary; NAM=naming; INIT_PHON=initial phonemes; BL_PHON=blending phonemes; PHON_EL=phoneme elision.

Table 2.
Description of structure of models estimated

Model	Indicators of each Factor
One-dimensional (Model A)	(INIT_PHON + BL_PHON + PHON_EL + LIST_COMP + REC_VOC + NAM + MORPH_DER + MORPH_INFLV + MORPH_INFLN + CONT_INTER + CONT_RESP +COM_INT + NAR + RETEL + ST_COMP + VOC_DEF)
Two-dimensional (Model B)	(INIT_PHON + BL_PHON + PHON_EL + MORPH_DER + MORPH_INFLV + MORPH_INFLN); (NAR + RETEL + ST_COMP + VOC_DEF + LIST_COMP + REC_VOC + NAM + CONT_INTER + CONT_RESP + COM_INT)
Three-dimensional (Model C)	(INIT_PHON + BL_PHON + PHON_EL); (LIST_COMP + REC_VOC + NAM); (CONT_INTER + CONT_RESP + COM_INT + MORPH_DER + MORPH_INFLV + MORPH_INFLN + NAR +RETEL + ST_COMP + VOC_DEF)
Four-dimensional (Model D)	(INIT_PHON + BL_PHON + PHON_EL); (CONT_INTER + CONT_RESP + COM_INT); (MORPH_DER + MORPH_INFLV + MORPH_INFLN); (NAR +RETEL + ST_COMP + VOC_DEF + LIST_COMP + REC_VOC + NAM)
Five-dimensional (Model E)	(INIT_PHON + BL_PHON + PHON_EL); (CONT_INTER + CONT_RESP +COM_INT); (NAR +RETEL + ST_COMP + VOC_DEF); (LIST_COMP + REC_VOC + NAM) + (MORPH_DER + MORPH_INFLV + MORPH_INFLN)

Note. Indicators enclosed in parentheses and connected by "+" sign were combined to form a factor; indicators separated by ";" were represented as one factor; INIT_PHON = initial phonemes; BL_PHON = blending phonemes; PHON_EL = phoneme elision; LIST_COMP = listening comprehension (sentences); REC_VOC = receptive vocabulary; NAM = naming; MORPH_DER = implicit understanding of derivatives; MORPH_INFLV = implicit understanding of verb inflections; MORPH_INFLN = implicit understanding of noun inflections; CONT_INTER = context interpretation; CONT_RESP = context proper response; COM_INT = communicative intent; NAR = narrative; RETEL = retelling; ST_COMP = story comprehension; VOC_DEF = vocabulary definition.

Table 3.
Results of Confirmatory Models (four- and five-dimensional) by Grade

Model	Y-B χ^2	*df*	MLR scaling factor	RMSEA, *p* close (90% CI)	CFI	TLI	SRMR	AIC	Corrected $\Delta\chi^2$
Prekindergarten (*n* = 180)									
Model (D)	163.404*	98	.98	.061, *p* =.136 [.044, .077]	.956	.946	.052	1,2745.97	
Model (E)	151.465*	94	.97	.058, *p* =.206 [.040, .075]	.961	.950	.050	1,2742.76	12.35[a,]*
Kindergarten (*n*=269)									
Model (D)	205.766*	98	.99	.064, *p* =.034 [.054, .076]	.933	.918	.054	1,9322.96	
Model (E)	154.233*	94	.98	.049, *p* =.541 [.034, .062]	.963	.952	.044	1,9278.37	43.13[b,***]
First-grade (*n*=351)									
Model (D)	225.049*	98	1.03	.061, *p*=.045 [.050, .071]	.914	.895	.058	2,3415.17	
Model (E)	177.260*	94	1.03	.050, *p*=.471 [.039, .061]	.944	.928	.050	2,3372.24	41.34[c,]***

Note. MLR = robust maximum likelihood; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker Lewis index; SRMR = standardized root mean square residual; AIC = Akaike's information criteria.

[a]Four-dimensional versus five-dimensional model in prekindergarten (*df* = 4). [b]Four-dimensional versus five-dimensional model in kindergarten (*df* = 4). [c]Four-dimensional versus five-dimensional model in first grade (*df* = 4).

**p* < .05 ****p*< .001

Table 4.
Convergent and discriminant validity

	Four-dimensional model				Five-dimensional model				
	CR				**CR**				
	SEM_DISCR	PRAGM	MORPH	PHON	DISCR	SEMANT	PRAGM	MORPH	PHON
PK	0.87	0.83	0.80	0.80	0.76	0.83	0.84	0.80	0.80
K	0.81	0.82	0.80	0.77	0.71	0.77	0.82	0.80	0.77
G1	0.73	0.79	0.78	0.75	0.59	0.68	0.78	0.78	0.75
	AVE				**AVE**				
PK	0.50	0.64	0.59	0.57	0.67	0.63	0.64	0.59	0.58
K	0.39	0.61	0.59	0.52	0.39	0.53	0.61	0.59	0.60
G1	0.29	0.55	0.56	0.50	0.28	0.42	0.55	0.56	0.50
	MSV				**MSV**				
PK	0.64	0.64	0.59	0.33	0.83	0.83	0.64	0.59	0.34
K	0.42	0.42	0.30	0.30	0.61	0.61	0.49	0.30	0.38
G1	0.30	0.30	0.23	0.30	0.55	0.55	0.44	0.23	0.31
	ASV				**ASV**				
PK	0.52	0.37	0.40	0.28	0.58	058	0.43	0.43	0.29
K	0.35	0.22	0.17	0.22	0.39	0.38	0.25	0.25	0.23
G1	0.28	0.16	0.16	0.22	0.38	0.26	0.18	0.16	0.22
	√AVE				**√AVE**				
PK	0.71	0.80	0.77	0.75	0.82	0.79	0.80	0.76	0.76
K	0.62	0.78	0.77	0.72	0.62	0.73	0.78	0.77	0.77
G1	0.54	0.74	0.75	0.71	0.77	0.82	0.88	0.88	0.87

Note. PK = Prekindergarten; K = Kindergarten; G1 = First Grade; SEM_DISCR = semantic-discourse; PRAGM = pragmatic; MORPH = morphological; PHON = phonological; SEMANT = semantic; CR = composite reliability AVE = average variance extracted MSV = maximum shared variance; ASV = average shared variance; √AVE = square root of AVE.

Supplemental material

Data handling procedures

Univariate normality for all items was assessed through inspection of skewness and kurtosis criteria (Table 1), histograms and boxplots of the data (not shown). Study variables did not exceed the recommended thresholds (Kline, 2011) for problematic skew (>3) or kurtosis (>10) indicating acceptable univariate normal distribution. Multivariate normality was tested across grades in AMOS 23 using Mardia's normalized estimate of multivariate kurtosis (1970). In the present study critical ratio of Mardia's kurtosis exceeded the recommended cut-off value of |5| (Bentler, 2006; Park & Schutz, 2005) for G1 (c.r = 14.087) suggesting violation of multivariate normality (Table 1). Robust techniques were used for handling legitimate outliers and dealing with multivariate non-normality (Kwak & Kim, 2017). Although there are different testing procedures proposed in the literature to assess multinormality (see Kim, 2015), Mardia's test is commonly used because it is simple and informative regarding non-normality of the data (Zhou & Shao, 2014). Multivariate outliers as indicated by the Mahalanobis distance criterion and its respective p1 and p2 values were not deleted to improve departure from multinormality as this could entail the risk of mistakenly eliminating valid information (Sheskin, 2004, p. 403).

Multivariate collinearity was also examined through the correlation matrix for independent variables (not shown) and the variance inflation factor (VIF). Examination of VIF generated from multiple regression analysis in SPSS 25 by assigning a dummy variable (gender was used) as the dependent variable and task as the independent variable. Given that all correlation coefficients were below <.90 (Hair, Blanck, Babin, & Anderson, 2010) and VIF<10 (Kline, 2011) multicollinearity problems were not addressed further (see Table S2).

Model testing procedures and criteria

The Satorra-Bentler scaled chi square fit index (also known as the Yuan–Bentler T2 statistic [Yuan & Bentler, 2000]) was used to assess overall model fit with a nonsignificant χ^2 value indicating a small discrepancy between expected and observed covariance matrices and thus an acceptable measurement model (Barrett, 2007). Given the sensitivity of χ^2 to sample size (Kline, 1998; Schumacker & Lomax, 2004) we also examined a combination of other fit indices including the Comparative fit index (CFI), the Non-normed fit index (NNFI, also known as the Tucker-Lewis index, TLI), the Root Mean Square Error of Approximation (RMSEA), the Standardized Root Mean Square Residual (SRMR) and the Akaike Information Criterion (AIC). CFI analyzes the extent to which the tested model is superior to an alternative model in reproducing the observed covariance matrix (Bentler, 1990). A cut-off value of CFI $\geq$.95 is presently recognized as indicative of good fit (Brown, 2006; Hu and Bentler, 1999). TLI evaluates the discrepancy between the chi-square value of the hypothesized model and the chi-square of the null model (Bentler & Bonett, 1980) with values $\geq$ .95 indicating acceptable fit (Hu & Bentler, 1999). RMSEA assesses how well optimally chosen parameter estimates fit the population covariance matrix (Hooper, Coughlan, & Mullen, 2008). RMSEA values $\leq$.08 or, more conservatively $\leq$.05, suggest acceptable model fit (Brown, 2006). 90% Confidence intervals for the RMSEA and results of the closeness of fit test which examines the null hypothesis that RMSEA equals .05 (Browne & Cudeck, 1992) were also reported. SRMR is an index of the average of standardized residuals between the observed and the hypothesized covariance matrices (Bentler, 1995, cited in Chen, 2007). SRMR values as high as .08 are deemed acceptable (Hu & Bentler, 1999) and less than $\leq$ .05 indicate a well-fitting model (Byrne, 2012; Diamantopoulos & Siguaw, 2000). AIC, a log likelihood measure of fit useful for comparing typically non-nested models, adjusts the chi-square value for the number of estimated parameters (Burnham & Anderson, 2004). AIC's smaller value indicates better fit (Kline, 2013). CFI incremental change was also used in comparison of nested models, with a value <.01 favoring the more parsimonious one (Moran, Marsh & Nagengast, 2013).

In addition to the above model fit indices, the scaled χ^2 difference test ($\Delta\chi^2$) was estimated based on the procedure described by Satorra and Bentler (2001) to compare pairs of nested models. If the $\Delta\chi^2$ test is not statistically significant, the more restricted model is retained as having model fit no worse than the more complex model. If the difference was statistically significant we accepted the more complex model as more adequate. Descriptive fit indices and results of statistical comparisons were not the only criteria we took into account in order to decide the measurement model with the best fit.

Table S1

Variance Invariance Flator (VIF) across grades

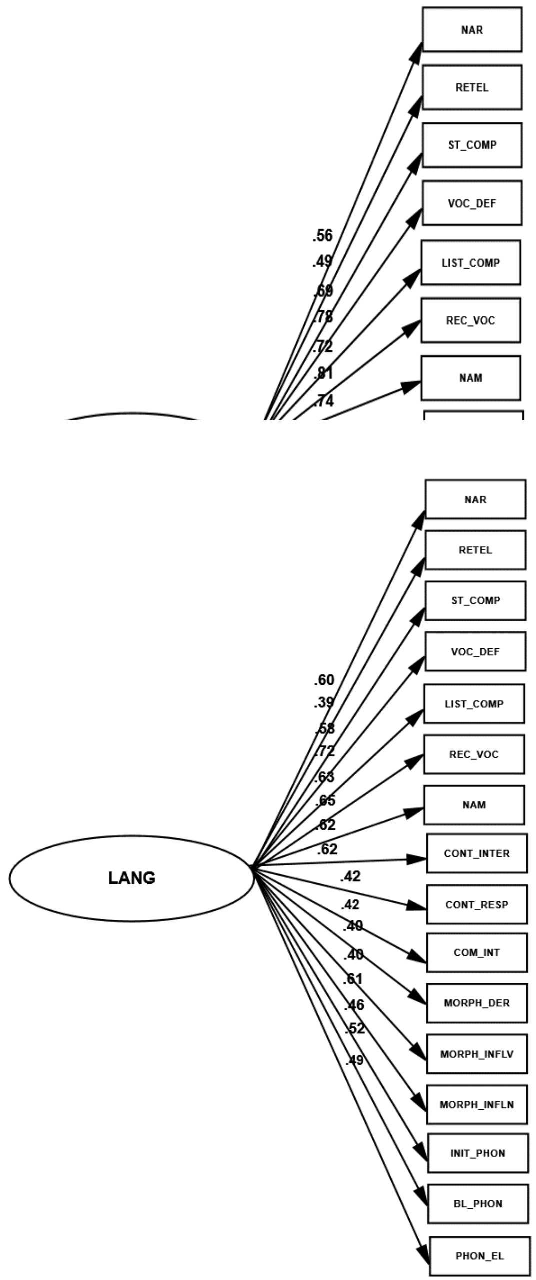
VARIABLE	Prekindergarten	Kindergarten VIF	First grade
NAR	1.59	1.28	1.21
RETEL	1.48	1.44	1.32
ST_COMP	2.04	1.61	1.27
VOC_DEF	2.68	2.08	1.75
CONT_INTER	1.84	1.62	1.83
CONT_RESP	2.85	2.58	1.75
COM_INT	3.16	2.49	1.94
MORPH_DER	1.43	1.37	1.36
MORPH_INFLV	3.61	2.82	2.61
MORPH_INFLN	3.74	2.76	2.18
LIST_COMP	2.20	1.82	1.52
REC_VOC	2.80	1.95	1.64
NAM	2.44	1.82	1.34
INIT_PHON	1.95	1.69	1.51
BL_PHON	2.05	1.83	1.72
PHON_EL	2.39	1.84	1.80

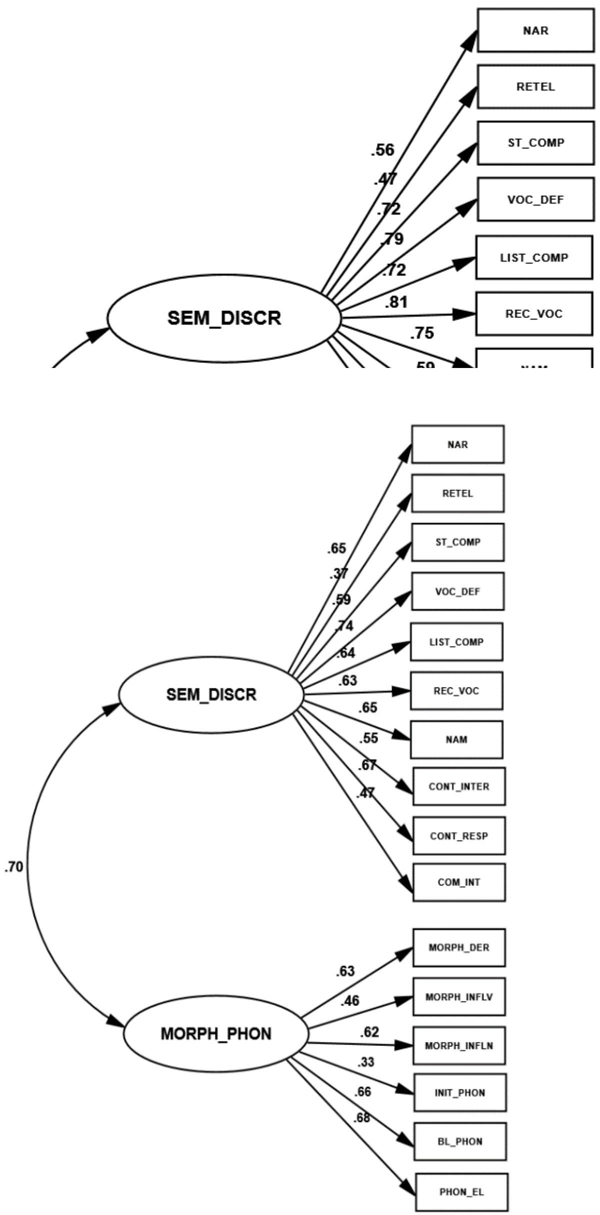
Note. NAR = narrative; RETEL= retelling; ST_COMP = story comprehension; VOC_DEF = vocabulary definition; CONT_INTER = context interpretation; CONT_RESP = context proper response; COM_INT = communicative intent; MORPH_DER = implicit understanding; MORPH_INFLV = implicit understanding of verb inflections; MORPH_INFLN = implicit understanding of noun inflections; LIST_COMP = listening comprehension (sentences); REC_VOC = receptive vocabulary; NAM = naming; INIT_PHON = initial phonemes; BL_PHON = blending phonemes; PHON_EL= phoneme elision.

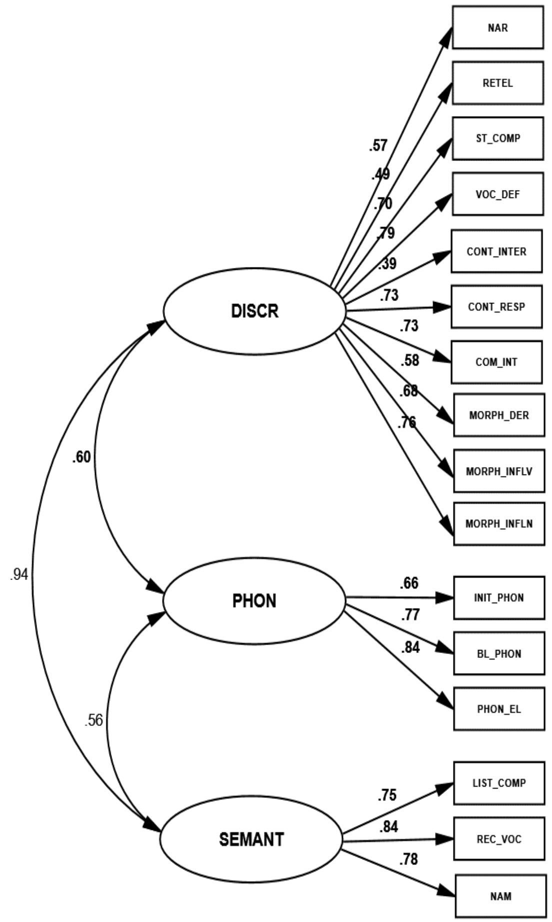
Table S2.
CFA results of the one-, two- and three-dimensional Models by Grade

Model	Y-B χ^2	*df*	MLR scaling factor	RMSEA, *p* close (90% CI)	CFI	TLI	SRMR	AIC
Prekindergarten (*n* = 180)								
Model (A)	438.963*	104	1.00	.134, *p* <.001 [.121, .147]	.772	.737	.080	1,3014.55
Model (B)	337.622*	103	.99	.112, *p* <.001 [.099, .126]	.841	.814	.072	1,2910.59
Model (C)	333.299*	101	.99	.113, *p* <.001 [.100, .127]	.842	.812	.066	1,2909.43
Kindergarten (*n*=269)								
Model (A)	691.196*	104	1.00	.145, *p* <.001 [.135, .155]	.635	.579	.098	1,9799.11
Model (B)	640.734*	103	.99	.139, *p* <.001 [.129, .150]	.666	.610	.102	1,9740.50
Model (C)	648.012*	101	.98	.142, *p* <.001 [.132, .152]	.660	.596	.097	1,9745.28
First-grade (*n*=351)								
Model (A)	735.327*	104	1.03	.132, *p* <.001 [.123, .141]	.572	.506	.095	2,3930.73
Model (B)	651.283*	103	1.03	.123, *p* <.001 [.114, .132]	.629	.567	.096	2,3842.41
Model (C)	692.753*	100	1.01	.129, *p* <.001 [.120, .138]	.599	.524	.091	2,3869.67

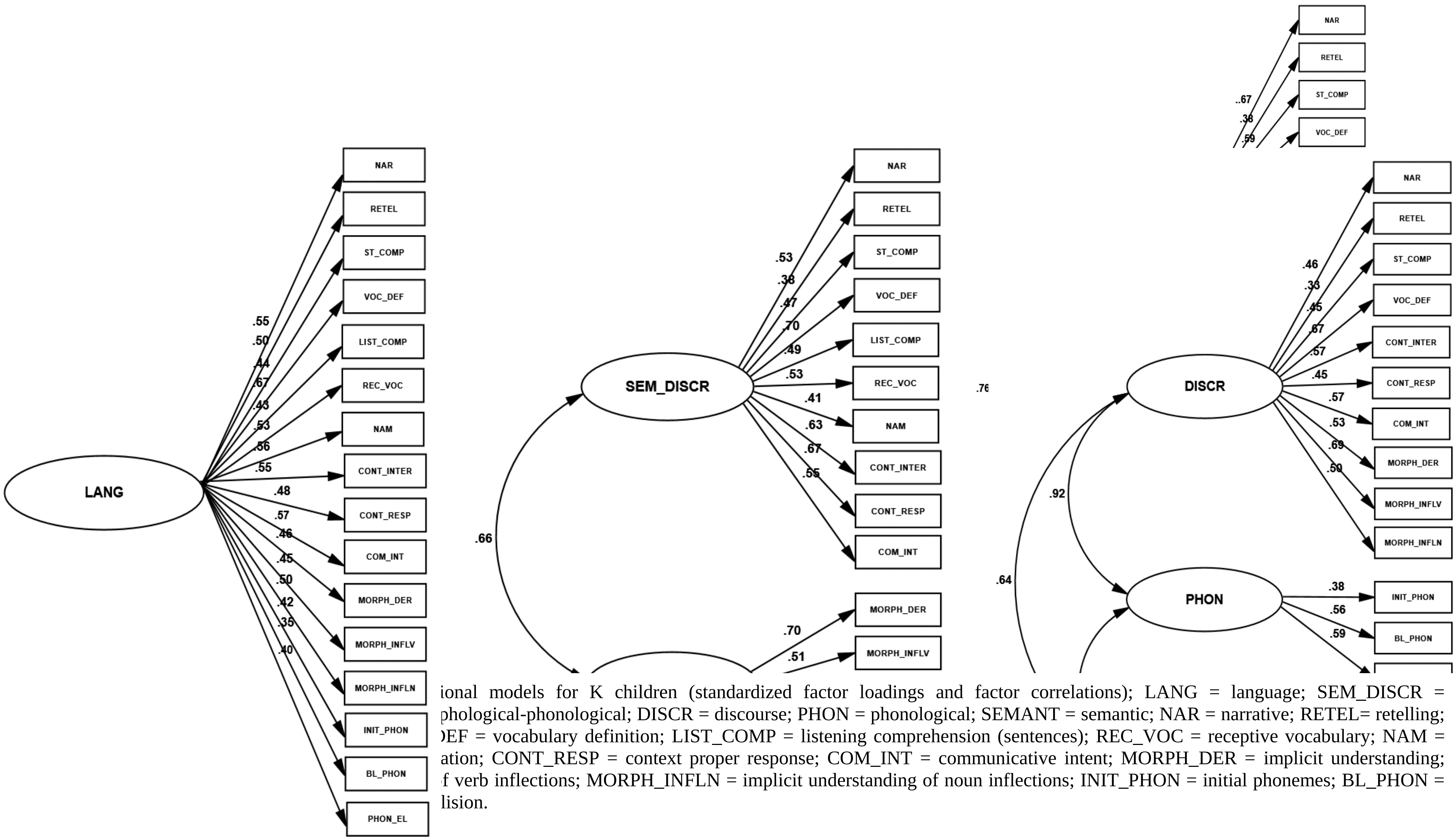
Note. Model (A) = one-dimensional; Model (B) = two-dimensional; Model (C) = three-dimensional; MLR = robust maximum likelihood; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker Lewis index; SRMR = standardized root mean square residual; AIC = Akaike's information criteria.







nd factor correlations); LANG = language; SEM_DISCR =
cal; SEMANT = semantic; NAR = narrative; RETEL= retelling;
nsion (sentences); REC_VOC = receptive vocabulary; NAM =
ommunicative intent; MORPH_DER = implicit understanding;
oun inflections; INIT_PHON = initial phonemes; BL_PHON =



ional models for K children (standardized factor loadings and factor correlations); LANG = language; SEM_DISCR = phological-phonological; DISCR = discourse; PHON = phonological; SEMANT = semantic; NAR = narrative; RETEL= retelling; DEF = vocabulary definition; LIST_COMP = listening comprehension (sentences); REC_VOC = receptive vocabulary; NAM = ation; CONT_RESP = context proper response; COM_INT = communicative intent; MORPH_DER = implicit understanding; f verb inflections; MORPH_INFLN = implicit understanding of noun inflections; INIT_PHON = initial phonemes; BL_PHON = lision.

Figure S3. One, two- and three- dimensional models for G1 children (standardized factor loadings and factor correlations); LANG = language; SEM_DISCR = semantic_discourse; MORPH_PHON = morphological-phonological; DISCR = discourse; PHON = phonological; SEMANT = semantic; NAR = narrative; RETEL= retelling; ST_COMP = story comprehension; VOC_DEF = vocabulary definition; LIST_COMP = listening comprehension (sentences); REC_VOC = receptive vocabulary; NAM = naming; CONT_INTER = context interpretation; CONT_RESP = context proper response; COM_INT = communicative intent; MORPH_DER = implicit understanding; MORPH_INFLV = implicit understanding of verb inflections; MORPH_INFLN = implicit understanding of noun inflections; INIT_PHON = initial phonemes; BL_PHON = blending phonemes; PHON_EL= phoneme elision.