Investigating the Factor Structure and Measurement Invariance of Phonological Abilities in a Sufficiently Transparent Language

Timothy C. Papadopoulos
University of Cyprus

Panayiota Kendeou
Neapolis University Pafos

George Spanoudis
University of Cyprus

Theory-driven conceptualizations of phonological abilities in a sufficiently transparent language (Greek) were examined in children ages 5 years 8 months to 7 years 7 months, by comparing a set of a priori models. Specifically, the fit of 9 different models was evaluated, as defined by the Number of Factors (1 to 3; represented by rhymes, syllables, and phonemes) × Relationships Between the Latent Variables (orthogonal vs. oblique) × Level (first vs. second order). In addition, the invariant sequence of phonological abilities was examined through longitudinal factorial invariance. We administered a set of 10 phonological tasks that differed in linguistic complexity to 280 Greek-Cypriot children. Reading fluency was also assessed to externally validate the conceptualization of phonological sensitivity. The results provided evidence for the model that depicts phonological sensitivity as a unified construct that develops in an invariant structure across time and significantly predicts a child’s reading performance.

Keywords: phonological ability, dimensionality, transparent languages

Although phonological sensitivity underlies successful reading acquisition in all languages (e.g., Frith, Wimmer, & Landerl, 1998; Goswami, Gombert, & De Barrera, 1998; Ho & Bryant, 1997; Papadopoulos, Georgiou, & Kendeou, 2009), the nature and conceptualization of phonological processing differs across languages (Loizou & Stuart, 2003; Ziegler, Perry, Jacobs, & Braun, 2001). According to the psycholinguistic grain size theory (Ziegler & Goswami, 2005), in transparent languages in which the mapping of graphemes onto phonemes is relatively unambiguous, phonological recoding operates at a smaller grain size (e.g., phoneme). In nontransparent languages in which the mapping of graphemes onto phonemes is relatively ambiguous, phonological recoding operates at a larger grain size (e.g., syllable). The purpose of the present study was to test whether these language differences have implications for the conceptualization of phonological abilities across development in a sufficiently transparent language (i.e., Greek; Protopapas & Vlahou, 2009). To address this aim, we evaluated the fit of several theory-driven models of phonological abilities that have been proposed in the relevant literature and explore the factor structure and measurement stability across time or between adjacent grades.

With respect to the conceptualization of phonological sensitivity, one set of models that has been proposed depicts phonological sensitivity as a set of distinct abilities that includes phoneme level and subsyllabic skills (Treiman, 1985), word segmentation skills (Morais, 1991b), or only phoneme level skills (Morais, 1991a). Another set of models depicts phonological skills as a single ability representing sensitivity in a continuum from shallow to deep (Goswami & Bryant, 1990; Stanovich, 1992) or from narrow to broad (Skowronek & Marx, 1989). A point of agreement of these two sets of models has been the premise that there are multiple phonological skills that are distinguished by linguistic complexity and the load of the cognitive processing involved. A point of disagreement has been whether these multiple phonological skills are distinct abilities or the same construct.

On the one hand, evidence for the distinct abilities hypothesis comes from a series of studies demonstrating that different phonological skills relate to reading in different ways. With respect to this hypothesis, two different models have been suggested: (a) a two-factor orthogonal (i.e., uncorrelated factors) model representing separate shallow (or supraphonemic sensitivity) and deep (or phonemic sensitivity) distinct abilities in both nontransparent (e.g., in English; Muter, Hulme, Snowling, & Taylor, 1997) and transparent languages (e.g., in Italian; Cossu, Shankweiler, Liberman, Katz, & Tola, 1988) and (b) a three-factor orthogonal model

1 Protopapas and Vlahou (2009) have reported that the overall consistency for Greek at the grapheme-phoneme level is 95.1% in thefeedforward and 80.3% in the feedback direction. On the basis of these findings, they supported that Greek is a sufficiently transparent language.
representing rhyme, syllabic, and phonemic sensitivity distinct abilities in transparent languages (e.g., in Norwegian; Høien, Lundberg, Stanovich, & Bjaalid, 1995). On the other hand, evidence for the unitary hypothesis comes from a series of correlational studies demonstrating high relations among different phonological abilities and from factor analyses suggesting the existence of a single construct. With respect to the unitary hypothesis, three models have been proposed: (a) a one-factor model representing a single ability in nontransparent (e.g., in English; Anthony & Lonigan, 2004; Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999; Stahl & Murray, 1994) and transparent (e.g., in Dutch; Vloedgraven & Verhoeven, 2007) languages, (b) a two-factor oblique (i.e., correlated factors) model representing separate supraphonemic and phonemic sensitivity correlated abilities, in transparent languages (in German; Skowronek & Marx, 1989; in Turkish; Durgunoglu & Oney, 1999), and (c) a three-factor oblique model representing rhyme, syllabic, and phonemic sensitivity correlated abilities (Anthony et al., 2002). We identified a fourth model that has been hypothesized but not tested (Yopp, 1988): a second-order factor model representing a single ability conceptualization from two latent variables, each representing supraphonemic and phonemic sensitivity abilities.

We agree that the evaluation of different theory-based models of phonological sensitivity is warranted and has theoretical as well as practical implications. For this reason, in the present study we selected and tested a number of theory-driven models with strong consonance (Mueller & Hancock, 2010). To date, these models have been proposed and tested mostly in isolation. We believe that the articulation, justification, and testing of competing alternative models in a single study is necessary so that we can provide a more complete picture of the conceptualization of the phonological sensitivity construct. Furthermore, we propose that testing the unitary versus distinct hypotheses can benefit from the application of models allowing the concurrent examination of a general factor and any specific ability factors, such as those that have been successfully tested in intelligence research, namely, nested factor modeling (Gignac, 2005; Gustafsson & Balke, 1993; Mulaik & Quartetti, 1997). In these models, the general factor is indexed by all manifest variables and specific factors, which are orthogonal to both the general factor and to each other. In the present study, we also examined the plausible conceptualization of an orthogonal first-order solution of broadly used phonological tests, in comparison with the aforementioned hierarchical models, aiming to directly test whether such a model could provide a competing—if not better—conceptualization of the structure of phonological abilities compared with the models that have been tested to date. It is important to note that relevant research indicates that the differences in the two approaches may be smaller in practice than in theory. Although hierarchical and nested-factor models differ considerably at a theoretical level, they are rarely distinguishable on the basis of fit (Mulaik & Quartetti, 1997). Thus, if the hierarchical models examined here have significantly poorer fit than the nested-factor model, then that would indicate potential problems with the model specification, that is, the structure of phonological abilities as has been proposed and tested to date.

It is important to note that the examination of the factor structure of phonological abilities on its own is not sufficient. It is necessary to examine the degree to which the suggested factor structure remains constant across different measures and across time. To date, there are only a handful of studies addressing this issue, applying different methodologies in data analysis. For example, Anthony and Lonigan (2004) used confirmatory factor analysis on pulled data from four independent studies and reported high-stability correlations among two latent constructs (an Onset/Rime Sensitivity factor and a Phoneme Awareness factor) for kindergartens to first grade and for first grade to second grade. Similar findings have been reported in a series of cross-sectional, correlational studies in kindergarten and early elementary school children by Wagner and colleagues (Wagner, Torgesen, & Rashotte, 1994; Wagner et al., 1997). In contrast, Schatschneider et al. (1999), using item response theory analysis with data from kindergarten through Grade 2, concluded that although the seven tasks of their phonological battery were well-represented as a unitary construct, the construct was variant across time. Specifically, the construct, as a whole, estimated more accurately children’s phonological abilities in kindergarten and the middle of first grade than at the end of first grade and in second grade. The contradictory nature of the findings by Schatschneider et al. (1999) and those by Anthony and Lonigan (2004) and Wagner et al. (1994, 1997) may be due in part to the different methodological approaches and designs. Particularly, addressing the important issue of invariance in a longitudinal design is a critical component that is currently missing from the relevant literature. In the present study, we directly addressed this limitation by testing whether the various theoretical models depicting different conceptualizations of phonological abilities remained constant across different measures and development. To do so, we used an advanced technique within structural equation modeling (SEM), that of longitudinal factorial invariance. This technique can provide strong support for the consistency of the latent constructs across different phonological sensitivity measures (e.g., rhyme, syllabic, and phonemic sensitivity; see Kline, 2011) and across development.

In summary, in the present study we evaluated the fit of the aforementioned theory-driven models supporting either the distinct abilities hypothesis or the unitary hypothesis using the nested factor modeling approach, and we explored the measurement stability across measures and time.

**The Present Study**

The present study is the second phase of a research program examining the relationship between phonological abilities and reading development in Greek language. In the first phase (Papadopoulos, Spanoudis, & Kendeou, 2009), we established the reliability and construct validity of the first comprehensive phonological battery developed to measure phonological abilities in Greek. The various items that were included in the battery assessed many of the dimensions of phonological ability described in relevant research, ranging from rhyming to syllabic to phonemic sensitivity. Specifically, using item response theory modeling, we demonstrated that the various items and types of tasks conformed to a measurement model that fitted the data well. This was true for analyses where the associations between the difficulty parameters were estimated from the entire sample and from two different gender samples. In addition, reliability and correlation analyses yielded high internal consistency for all tasks, with most of those being also significantly intercorrelated.
Establishing issues relating to scaling, validity, and reliability of the test items of a comprehensive battery is a prerequisite for testing alternative theoretical conceptualizations of the construct of interest. This procedure allows the detection of major threats of construct validity, construct underrepresentation, and construct-irrelevant variance. Subsequently, the researchers can safely examine the latent constructs that are being measured to understand the relationship between those latent constructs and to study the test structure across groups or over time (Tate, 2003). With the measurement issues of the aforementioned battery established in our previous research (Papadopoulos, Spanoudis, & Kendeou, 2009), the present study aims to evaluate the fit of theory-driven models supporting either the distinct abilities hypothesis or the unitary hypothesis using a nested factor modeling approach and to explore the measurement stability from kindergarten through Grade 2 using (partial) factorial invariance within SEM.

The examination of the structure of phonological abilities in Greek is important considering the highly regular and transparent nature of Greek language in which both syllable and phoneme level skills are equally essential in predicting successful early reading, regardless of unit length, position, and stress effects (Aidinis & Nunes, 2001; Protopapas & Vlachou, 2009). In addition, phonological skills have been strongly implicated as significant predictors of reading development regardless of the type of literacy instruction children receive (Papadopoulos, 2001) and regardless of reading difficulties (Papadopoulos, Georgiou, & Kendeou, 2009; Porpodas, 1999). There is also evidence indicating that phonological abilities—along with orthographic processing—not only contribute uniquely to reading ability in the first 2 years of schooling (Georgiou, Parrila, & Papadopoulos, 2008), but also have differential importance in Greek than in English, particularly with respect to their effect on word decoding. Specifically, comparing Greek and English children longitudinally in Grades 1 and 2, Georgiou et al. (2008) concluded that, in word decoding, Greek-speaking children rely on small grain size units as indicated by the significant effect of phonological abilities. In contrast, in reading fluency tasks, Greek-speaking children rely on large grain size units as indicated by the significant effect of orthographic processing tasks. Thus, Greek-speaking children may adjust the grain size units to match the task demands. Taken together, these findings speak for cross-linguistic differences in the development of phonological skills (and reading) with respect to the grain size unit of analysis.

To summarize, the findings from research in Greek are consistent with findings from research in other transparent and nontransparent languages with regard to the role of phonological abilities in reading development (de Jong & van der Leij, 1999; McBride-Chang & Kail, 2002; Parrila, Kirby, & McQuarrie, 2004). What is not clear, however, is the conceptualization and measurement stability across development of phonological skills in Greek. Particularly, we are interested in investigating the fit of the various theory-driven models proposed to explain dimensionality of phonological abilities in the literature and whether these models remain invariant across time.

In the present study, we tested the factor structure of phonological sensitivity in a sample of children across three consecutive waves, from age 5 years 8 months to age 7 years 7 months. To do so, first we compared alternative, theory-driven models representing either distinct or unitary structure of phonological sensitivity to select the best-fitting model or models in all three waves. Then, we compared the winning model or models with the nested-factor model of phonological abilities. Second, we tested factor invariance of the adopted model aiming at reaching a consensus about a baseline or configural model of phonological sensitivity in all three waves. This model was then used to test whether there was metric invariance across time and between adjacent grades (from kindergarten to Grade 1 and from Grade 1 to Grade 2). Because phonological skills tend to change as a result of reading experience in the first years of schooling (e.g., de Jong & van der Leij, 1999; Georgiou et al., 2008; Papadopoulos, Georgiou, & Kendeou, 2009; Parrila et al., 2004; Verhagen, Aarnoutse, & van Leeuwe, 2008), it was deemed necessary to examine whether the tests that we used measured the same constructs (same factor structure) and demonstrated equivalent relationships to these constructs (equal factor loadings) as a function of time. Finally, we conducted linear regression analyses within SEM, testing the significant predictive contribution of the latent phonological constructs to participants’ word reading skills across time to provide converging evidence for the factor structure of the phonological abilities in Greek.

The present study, therefore, systematically addresses the issues of conceptualization and structural invariance in a longitudinal data set. Our approach is more comprehensive than previous efforts (see, e.g., Anthony & Lonigan, 2004; Schatschneider et al., 1999) for at least three reasons: First, we selected and tested the fit of several theory-driven models that have often been studied in isolation; this was made possible by using an innovative approach, that of nested factor modeling. Second, we tested configural and partial measurement invariance while including a broader range of measures within the primary phase of phonological development. Third, we systematically covered three age intervals across preschool and early school years during which phonological skills develop rapidly. Thus, the findings of the present study contribute significantly to the theoretical understanding of the structure, measurement, and development of phonological abilities in a language with a transparent orthography and have direct educational implications with regard to the instruction of these skills in early years.

**Method**

**Participants**

A total of 280 Greek-Cypriot children (141 male and 139 female) participated in the study. The children were native Greek speakers with no reported history of speech, language, or hearing difficulties. The mean age of the group in the initial assessment (Wave 1) was 5 years 8 months ($SD = 31$ months). A year later (Wave 2), the mean age was 6 years 6 months ($SD = 31$ months), and in the final assessment (Wave 3), the mean age was 7 years and 7 months ($SD = 32$ months). Sample size is considered good for the type of analyses performed (Tabachnick & Fidell, 2007). The group’s verbal abilities (the Similarities and Vocabulary sub-scales of the Wechsler Intelligence Scale for Children—Third Edition—Revised; Wechsler, 1992) and nonverbal abilities (the Matrixes subscale of the Cognitive Assessment System; Naglieri & Das, 1997) were assessed in Wave 2, all yielding average performance on the basis of normative scores in Greek (see Georgas, Paraskevopoulos, Bezevegis, & Giannitsas, 1997, and Papadopoulos, Georgiou, Kendeou, & Spanoudis, 2007, for the Wechsler Intelligence
Scale for Children and the Cognitive Assessment System norms, respectively). Almost half of the parents of the participating group were college or university graduates (45%), and the remaining were high school graduates (55%), consistent with the numbers provided by the annual survey of the Statistical Service of Cyprus (2006; 45.3% and 54.7%, respectively). With regard to the community settings, approximately 62% of the participants were from urban communities, and 38% were from rural communities. These values are also in accordance with the composition of the Greek-Cypriot population, with 68.4% residing in urban settings and 31.6% residing in rural settings. These data indicate that the participating sample was representative of children in the Greek-Cypriot population in terms of community settings and parental education attainment. School consent and parental consent for participation in the study were obtained prior to testing.

In kindergarten, the children attended a program including mostly social activities and games with semiformal cognitive or linguistic training. In relation to language and literacy development specifically, children became aware of the range of books and tapes/CDs available for them while teachers were enhancing the children’s opportunities for learning and pleasure from reading books. The development of listening skills was also of major importance while the children were read stories. Children were also constantly introduced to new vocabulary while they were trying to record what they learned or found out. Drawing and prewriting activities were also included on a daily basis. Finally, rhyming and odd-out-word activities in which the child had to identify the word that differed from two or three others in its first or ending syllable (or sound) were occasionally practiced. Generally, the program concentrated on aiding children in learning to be more aware of print around them and to enjoy participating in routine literacy activities.

In Grade 1, the children were receiving formal reading and spelling instruction in a basal reading series that emphasized primarily word recognition, reading comprehension, and incidentally, word decoding and letter–sound correspondences through syllable-splitting activities. Phonological processing skills, in turn, were fostered through segmentation and blending activities as the key strategies. It is important to note that in the Cypriot educational system, the language books used are designed by the Greek Ministry of Education. The language taught is Modern Greek.

In Grade 2, emphasis was placed on reading and demonstrating an understanding of a variety of literary and informational texts, using a range of strategies to construct meaning. Also, children were encouraged to use knowledge of words and cueing systems to read fluently. Comprehension strategies, such as activating prior knowledge to ask questions or make predictions about the theme of a story, were reinforced by identifying and using different strategies before, during, and after reading to understand texts. Children were taught to read and understand high-frequency familiar words automatically and to predict the meaning of and quickly decipher low-frequency or unfamiliar words through different types of cues, such as phonological, grammatical, syntactic (language structure), and semantic (meaning). Children were also taught spelling of familiar and unfamiliar words through a variety of strategies that involve understanding word structures, word meanings, and sound–symbol relationships. Finally, children were prompted to read various texts at a sufficient rate and with sufficient expression to convey the meaning of the text to the reader and to an audience.

Measures

**Phonological ability measures.** Participants’ phonological skills were assessed with 10 tasks that have undergone extensive validation in previous work (Papadopoulos, Spanoudis, & Kendeou, 2009). Six of these tasks measured phonological ability at the syllabic level: Rhyme Oddity, Rhyme Generation, Syllable Segmentation, Syllable Completion, Final Syllable Oddity, and Initial Syllable Oddity tasks.² The remaining four tasks tapped phonological ability at the phonemic level: Initial Sound Oddity, Sound Isolation, Phoneme Elision, and Phoneme Blending tasks. The Rhyme Generation and Final Syllable Oddity tasks consisted of 10 testing items. All other tasks were made up of 15 testing items. Testing preparation, for all 10 tasks, included two sample items with feedback regarding the correctness of participant answers to ensure that all participants knew what was expected of them. All tasks were discontinued after four consecutive failures. In all instances, a participant’s score was the total number of correct responses. The selection and development of this set of tasks relied on the level of difficulty and linguistic complexity used in the relevant research with English-speaking populations (e.g., Anthony et al., 2002, 2003; Muter et al., 1997; Wagner et al., 1997) or in other transparent languages (e.g., de Jong & van der Leij, 1999; Vloeckgaven & Verhoeven, 2007) or with those that have been previously used in research studying the development of phonological skills in Greek in early years (e.g., Aidinis & Nunes, 2001; Loizou & Stuart, 2003) in relation to typical or atypical reading development (Papadopoulos, Georgiou, & Kendeou, 2009; Poroopoulos, 1999; Tafa & Manolitsis, 2008) and in cross-linguistic comparisons (e.g., Georgiou et al., 2008). The words for all tasks were sampled from the language books used in the Greek educational system after a systematic analysis of the available corpus (Papadopoulos & Loizou, 2007), so that a difficulty progression within each task was maintained.

**Rhyme Oddity.** This task was adapted from the work of Bradley and Bryant (1985). The child was required to listen to three words presented orally and to identify the one that ended with a different rhyme (final vowel–consonant–vowel [VCV]) compared with the other two (e.g., μπάλα/έλασσα/γάλα/; [ball, horse, milk]).

**Rhyme Generation.** In this task, the children were asked to produce words that rhymed with a target word (e.g., καλάθι → αγκαθή; /kala/thi/ → /anga/thi/; [basket → thorn]; final VCV). Both real words and pseudowords were considered as permissible responses. The maximum time allowed to generate a word for each target word was 30 s (see also Muter et al., 1997).

**Syllable Segmentation.** In this task, participants were explicitly directed to tap the number of syllables in the spoken words, aiming at revealing participants’ intuitive notions of syllabic units. The task was adapted from Mann and Liberman (1984) and included words that varied in length, containing one to six syllables. The first two testing items had a simple structure (CVC or VCV: πως; /pos/; [how]; έλα; /ela/; [come]). All the other words were

---

² Rhyme Oddity and Rhyme Generation were included in the syllabic level phonological measures because of the structure of rhyming tasks in Greek; adequate performance on these tasks depends on successful manipulation of the final syllable and the vowel preceding it.
made up of syllables with relatively higher complexity as defined by the number of consonants preceding a vowel or the use of diphthongs (example of a bisyllabic word: μπούκλα; /bukla/; [curl]).

**Syllable Completion.** In this task, the experimenter pronounced the first syllable of a bisyllabic word, and the participants were asked to provide the second syllable to complete the word (see also Loizou & Stuart, 2003). All words contained open syllables as target syllables. These syllables ended with a vowel and were structured as CV (e.g., γά-κο: /γα-κα:/ [cat]), CCV (e.g., κά-δρο: /κα-δρο:/ [frame]) or CCCV (e.g., δέ-τρο: /δε-τρο:/ [tree]). A set of pictures, each depicting the matching familiar object, was used following the administration procedure of this task in its original version.

**Final Syllable Oddity.** This task measured children’s awareness of final syllables. The children were given triads of bi- and trisyllabic spoken words and were asked to select from the triad the odd word that ended differently or did not alliterate with the other two. The words in each triad had the same stress position, which is a critical feature of the identity of words in languages such as Greek (Protopapas & Gerakaki, 2009). Given that Greek has predominantly an open-syllable structure, the alliteration was based on the initial consonant of the target cluster (in contrast to similar tasks in English where the alliteration is based on the rime; see Bradley & Bryant, 1985). The items were split into three groups: those that ended with a CV syllable (e.g., γάλα/γάλα/ /μπότα: /γαλα/γαλα/μπάτα:/ [milk, cat, boot]), those that ended with a CCV syllable (e.g., κορέκλα/κορέκλα/ /πέπλα: /καρέκλα/καρέκλα/ /πέπλα:/ [chair, doll, veils]), and those that ended with a CCCV syllable (e.g., καστρό/καστρό/ /μοίρα: /καστρα/καστρα/ /μοίρα:/ [castles, trees, yard]), with the initial consonant being the only sound that was different in the odd-out word in all instances.

**Initial Syllable Oddity.** In this task, participants were asked to pay attention to initial syllables and select the member of each three-item set that began with a different syllable than the other two. This task was also adapted from Bradley and Bryant (1985). There were three different groups of items: those that began with CV (e.g., μούμι/μόρι/ /μένο: /μαμά/μαμά/ /μένο:/ [mom, day, stay]), those that began with CCV (e.g., κρατώ/κρατώ/ /κρασί: /κατώ/κατώ/ /κρασί:/ [hold, hang, wine]), and those that began with CCCV (e.g., στρατό/στρατό/ /στράτες: /στρατα/στρατα/ /στράτες:/ [army, plot, street]). With the exception of only three item sets, which were introduced to participate in the task, the odd word out was contrasted to the other two on the basis of the syllable’s vowel.

**Initial Sound Oddity.** In this task, the child had to indicate which word of a series of three words started with a different sound (e.g., λόπα/λόπα/ /θαμμί: /λάμπα/λάμπα/ /ποιμή:/ [lamp, pound, bread]). This kind of task has been widely used to assess children’s phonemic awareness (e.g., Byrne & Fielding-Barnsley, 1989, 1990; de Jong, Seveke, & van Veen, 2000). The items used in this task consisted primarily of bisyllabic and high-frequency words that are typically acquired by Grade 1 children. The first testing item was relatively easy as only the odd-out word started with a consonant. Half of the remaining items were made up of words that could be contrasted on the basis of the initial phoneme with relative ease, because none of them shared the same vowel in the first syllable. The other half were somewhat more difficult because the target word shared the same vowel with one of the other words (e.g., μελα/μαρο/ /θέλες: /μελι/μορό/ /θελί:/ [honey, baby, wants]).

**Sound Isolation.** This task was a Greek adaptation of the work of Wagner, Torgesen, Laughon, Simmons, and Rashotte (1993) in which they compared alternative models of young readers’ phonological processing abilities. In this test, children were asked to repeat the first, last, or middle sound in a word (e.g., Which is the middle sound in the word θεα: /θεά/; [view]?). Testing items consisted of three- and four-phoneme one- and two-syllable words.

**Phoneme Elision.** This task was also an adaptation of the work by Wagner et al. (1993). In this task, children were asked to repeat a word after deleting an identified phoneme. The targeted phonemes were either vowels or consonants, and their position varied across items. After deleting the target phoneme, the remaining phonemes formed a word (e.g., Say the word τώρα: /τόρα:/ [now], after deleting the sound τ/ → ιώρα: /ιόρα:/ [time]).

**Phoneme Blending.** This task was designed to assess phoneme blending skills. Audio prompts presented the sounds of two- to six-sound words separately, and the child was asked to orally blend them into a word. The child’s response was recorded as correct when he or she reproduced all the sounds in the final word. Word complexity was progressively more difficult. The first four words consisted of two- to four-phoneme segments that were of CV or CVC structure (e.g., φως: /fós:/ [light]). The more difficult items contained more complex phoneme segments, such as CCCV (e.g., στόμα: /στόμα:/ [mouth]). The component sounds of each word were spoken at 500-ms intervals.

**Reading processing measures.**

**Word reading measures.** Two standardized measures were used to assess participants’ word reading ability, namely, a real word reading task and pseudoword reading task (Papadopoulos, Spanoudis, & Kendeou, 2008) in Waves 2 and 3. In both tasks, the reading speed (fluency) score, that is, the number of words read correctly within 60 s, was recorded for each participant. We used the fluency scores because previous studies in Greek (e.g., Papadopoulos, 2001; Porpodas, 1999) and also cross-language reading studies have shown that Grade 1 children learning to read in Greek achieve a very high accuracy rate (almost 98% for real word reading and 92% for pseudoword reading; Seymour, Aro, & Erskine, 2003). This means that in reading a regular writing system like Greek, even beginning readers with reading difficulties manage to decode almost any letter array successfully. However, this cognitive processing deteriorates when a time frame is set and the child is required to read as many words as possible within it (Georgiou et al., 2008). Both the real word and the nonword lists were preceded by a practice list to familiarize children with the list-reading procedure and with nonwords.
The task started with bisyllabic words and ended with five-syllable words. Cronbach’s alpha for this task is .92 in Grade 1 and .70 in Grade 2.

Procedure

Participants were tested individually in a session lasting approximately 40 min, between February and April from the first (Wave 1; age 5) through the third assessment (Wave 3; age 8). The presentation of the tasks was counterbalanced across the participants within each wave. Graduate students carried out all testing during school hours in quiet rooms at the schools. Before the beginning of the study, the graduate students were trained in test administration and data recording.

Results

Preliminary Analysis

First, we examined the raw scores of all measures in each age level. We found mild departures from normality but no obvious outliers. As expected, the mean performance on all tasks increased from Wave 1 to Wave 3. The mean scores in Wave 1 were at floor level in three out of the 10 tasks: Rhyme Generation, Phoneme Elision, and Blending. These results suggest that these tests were too difficult for the participants in the first assessment. Rhyme Generation was still difficult to perform even a year later, in Wave 2, a finding that is consistent with previous work in both English (e.g., Anthony & Lonigan, 2004) and Greek (e.g., Papadopoulos, Spanoudis, & Kendeou, 2009). Additionally, performances on Final Syllable Oddity, Initial Syllable Oddity, and Initial Sound Oddity were close to chance level in Wave 1. Indices of skewness and kurtosis showed that the vast majority of the measures were within reasonable limits, although some departed from normality. Kurtosis was positively large in the case of three tasks in Wave 1, namely, Rhyme Generation ($k = 6.18$), Phoneme Elision ($k = 5.68$), and Blending ($k = 6.64$), and in two tasks in Wave 3, which were at ceiling, namely, Syllable Completion and Sound Isolation ($k = 11.35$ and 6.69, respectively). As a result, we used a robust maximum likelihood estimator (Satorra–Bentler robust technique) instead of the ordinary maximum likelihood estimator (Bentler, 2006; Byrne, 2006) in subsequent analyses. Table 1 shows the mean scores for each wave, followed by Cronbach’s alpha reliability coefficients and chance levels of performance for all the phonological measures in all years.

Model Comparison Using SEM

To test the factor structure of phonological sensitivity, we compared alternative, theory-driven models representing either distinct or unitary structure of phonological sensitivity. Our aim was to select the best-fitting model across all three waves.

Two different models were tested in support of the distinct abilities hypothesis across ages: the two-factor orthogonal model (with separate supraphonemic and phonemic sensitivity uncorrelated abilities; Model 3 in Table 2) and the three-factor orthogonal model (with separate rhyme, syllabic, and phonemic sensitivity uncorrelated abilities; Model 6). Four different models were tested in support of the unitary abilities hypothesis across ages: the two-factor oblique model (with separate supraphonemic and phonemic sensitivity correlated abilities; Model 2), the three-factor oblique model (with separate rhyme, syllabic, and phonemic sensitivity correlated abilities; Model 5), the one-factor model (a single ability as equivalent to the two- and three-factor oblique counterparts; Models 1 and 4, respectively), and the second-order factor model (a single ability conceptualization from two latent variables; Model 7).

These models were nested in that they could be derived by imposing constraints on the oblique models, in the case of the two- and three-factor models. For example, constraining the correlations between factors in the two-factor oblique model to 0 yielded a one-factor model, whereas constraining the correlations to zero (0) led to a two-factor orthogonal model (Anthony & Lonigan, 2004). A similar approach was used in the case of the three-factor model. In addition, to avoid underidentification in the three-factor models, the variables Syllable Completion and Syllable Segmentation were constrained to be equal. In the case of the second-order model, and because it had only two first-order factors, the following restrictions were applied to the calculation of the model: the variance of the second-order factor was fixed to 1, and an equality constraint was placed on the first-factor loadings, yielding the one-factor equivalent model (Model 7). This con-

### Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Wave 1</th>
<th></th>
<th>Wave 2</th>
<th></th>
<th>Wave 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chance</td>
<td>M</td>
<td>SD</td>
<td>α</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Rhyme Oddity</td>
<td>4.9</td>
<td>6.39</td>
<td>4.86</td>
<td>.91</td>
<td>9.21</td>
<td>4.38</td>
</tr>
<tr>
<td>Rhyme Generation</td>
<td>0.0</td>
<td>0.88</td>
<td>1.24</td>
<td>.77</td>
<td>1.15</td>
<td>1.45</td>
</tr>
<tr>
<td>Syllable Segmentation</td>
<td>1.0</td>
<td>8.70</td>
<td>5.05</td>
<td>.93</td>
<td>12.15</td>
<td>3.22</td>
</tr>
<tr>
<td>Syllable Completion</td>
<td>0.7</td>
<td>9.53</td>
<td>4.83</td>
<td>.94</td>
<td>13.03</td>
<td>1.89</td>
</tr>
<tr>
<td>Final Syllable Oddity</td>
<td>3.3</td>
<td>2.93</td>
<td>2.02</td>
<td>.66</td>
<td>3.42</td>
<td>2.25</td>
</tr>
<tr>
<td>Initial Syllable Oddity</td>
<td>4.9</td>
<td>4.06</td>
<td>3.08</td>
<td>.82</td>
<td>5.25</td>
<td>3.20</td>
</tr>
<tr>
<td>Initial Sound Oddity</td>
<td>4.9</td>
<td>4.32</td>
<td>3.40</td>
<td>.83</td>
<td>7.95</td>
<td>4.85</td>
</tr>
<tr>
<td>Sound Isolation</td>
<td>0.0</td>
<td>3.90</td>
<td>4.32</td>
<td>.91</td>
<td>12.72</td>
<td>3.18</td>
</tr>
<tr>
<td>Phoneme Elision</td>
<td>0.0</td>
<td>1.62</td>
<td>3.24</td>
<td>.93</td>
<td>8.87</td>
<td>4.86</td>
</tr>
<tr>
<td>Blending</td>
<td>1.0</td>
<td>1.61</td>
<td>2.94</td>
<td>.91</td>
<td>8.70</td>
<td>4.13</td>
</tr>
</tbody>
</table>

Note. $n = 280$ in all waves.

* These tasks consisted of 10 testing items; all other tasks were made up of 15 testing items.
strained second-order model tests properly the one-factor (second-order) solution compared with the second-order model that does not include the constraint, which artificially uses two parameters (loadings) to explain the one-factor intercorrelation of the two-factor oblique model.

These nested models were directly compared using a chi-square difference test, which, in turn, allowed for the selection of the most parsimonious, best-fitting model. The difference between chi-square for nested models is itself distributed as chi-square with \( k \) degrees of freedom, where \( k \) equals the degrees of freedom for the more constrained model minus the degrees of freedom for the less constrained model. This means that it is possible to test directly whether more constrained models have a significantly poorer fit than less constrained models. However, it is important to note that with the Satorra–Bentler (S-B) chi-square statistics being used as an index of fit rather than the standard chi-square statistic, the difference between S-B chi-square for nested models is typically not distributed as chi-square. For this reason, to calculate the S-B chi-square from the nested models, we used a scaled difference chi-square test statistic that has been recently developed by Satorra and Bentler (2001) for this purpose.

In addition, to compare the nonnested models, we used fit indexes that take parsimony into account, namely, Akaike information criterion (AIC) and expected cross-validation index (ECVI). Although AIC is more broadly used and accepted as an index reflecting the discrepancy between model-implied and observed covariance matrices (Browne & Cudeck, 1992), we also chose to use ECVI as a cross-validation index. ECVI penalizes for number of free parameters and, therefore, is considered as a more robust index of model comparison. Lower AIC and ECVI indicate a better fit (Byrne, 2006; Hu & Bentler, 1999). In addition, we adhered to the following criteria for evaluating good model fit: comparative fit indexes (CFIs) and Tucker–Lewis indexes (TLIs) greater than .95 and root-mean-square errors of approximation (RMSEAs) below .06 (Byrne, 2006; Hu & Bentler, 1999).

**Distinct abilities hypothesis.** For the two-factor orthogonal model (Model 3), indicators for the supraphonemic sensitivity were Rhyme Oddity, Rhyme Generation, Syllable Segmentation, Syllable Completion, Final Syllable Oddity, and Initial Syllable Oddity, whereas indicators for the phonemic sensitivity were Initial Sound Oddity, Sound Isolation, Phoneme Elision, and Blending. For the three-factor orthogonal model (Model 6), indicators for the rhyme phonological abilities were Rhyme Oddity, Rhyme Generation, Final Syllable Oddity, and Initial Syllable Oddity. Indicators for the syllabic phonological abilities were Syllable Segmentation and Syllable Completion. Indicators for phonemic sensitivity abilities were the same as those in the two-factor model. Both models had a poor fit to the data in all three age groups. Table 2 displays the fit indices for the phonological ability models across ages.

**Unitary hypothesis.** The indicators for the two-factor (Model 2) and the three-factor oblique (Model 5) models were the same as those of the orthogonal counterpart models. The one-factor models (Models 1 and 4) included all tasks as indicators. Finally, the one-factor second-order model included as indicators the latent constructs of the two-factor model (Model 7). The second-order model was based on the two-factor oblique model rather than the three-factor oblique model because the two-factor oblique model was more parsimonious and had a better fit than the three-factor oblique model. In essence, the one-factor second-order model was identical to the two-factor oblique, with the exception that the covariation between supraphonemic and phonemic sensitivity factors was modeled as a second-order factor of a general phonological ability. As can be seen in Table 2, the one-factor (two and three parameters) models (Models 1 and 4) had a poor fit to the data in all waves. In contrast, the two-factor oblique (Model 2), the one-factor second-order (Model 7), and the three-factor-oblique models (Model 5) had a fairly good fit in all waves. Notably, the fit indices of the two-factor oblique model were the same as those of the one-factor equivalent model, suggesting that they could be seen as reparameterizations of each other. Importantly, all three models provide support for a unifying structure of phonological abilities.

It is also interesting to note that in the one-factor second-order model, the coefficients to the general phonological ability factor were statistically significant and ranged in size from .88 to .98 from the supraphonemic sensitivity factor and from .73 to .98 from the phonemic sensitivity factor for Wave 1 to Wave 3, respectively (all \( ps < .001 \)). The estimated intercorrelations of the two factors of supraphonemic and phonemic sensitivity skills in the two-oblique model ranged in size from .71 to .97 for Wave 1 to Wave 3, respectively (all \( ps < .001 \)). Similarly, the estimated intercorrelations of the three factors of rhyme phonological abilities, syllabic phonological abilities, and phonemic sensitivity abilities in the three-oblique model ranged in size from .60 (for Wave 1, between syllabic and phonemic sensitivity factors) to .95 (for...
Wave 3, between rhyme and phonemic sensitivity factor; all p < .001. The high intercorrelations in both of these models provide further support for the conceptualization of phonological ability as a unitary abilities construct.

Next, we examined whether the solution of nested factor models would have a better fit to the data or provide a better conceptualization of the unified structure of phonological abilities compared with the three models already discussed. Two nested factors models were examined that were parametrized to allow simultaneous indications of (a) the general phonological ability factor and the two specific factors of supraphonemic and phonemic sensitivity as an alternative to the two-factor oblique and one-factor second-order model (Model 8) and (b) the general phonological ability factor and the three specific factors of rhyme phonological abilities, syllabic phonological abilities, and phonemic sensitivity abilities as an alternative model for the three-factor oblique model (Model 9). It was hypothesized that the nested-factor models would account for the intersubtests covariation of the phonological abilities better than the oblique and higher order factor models as they would allow for concurrent examination of which of the above latent scores accounted for the largest portion of variance for all subtests. Thus, Model 8 was a nested model, with all 10 phonological tasks specified to weight on a first-order general phonological abilities factor. The six rhyme and syllabic ability tasks were specified to weight on a first-order supraphonemic factor, and the four phonemic awareness tasks were specified to weight on a first-order phonemic factor. In turn, Model 9 was a nested model, with all 10 phonological tasks specified to weight on a first-order general phonological abilities factor: The six rhyme and syllabic ability factors were specified to weight on a first-order supraphonemic factor, the two syllabic tasks were specified to weight on a first-order syllabic phonological abilities factor, and the four phonemic awareness tasks were specified to weight on a first-order phonemic factor. Also, as in the previous analyses, we allowed pairs of residual (error) variances to covary in order to improve the fit of the model in each wave. These correlations were suggested by EQS modification indices and represent measurement error held in common by some of the measures we used. Specifically, in Model 8 Initial Sound Oddity and Sound Isolation, and Phoneme Elision and Blending were allowed to covary in Wave 1. Initial Sound Oddity and Sound Isolation, and Rhyme Oddity and Rhyme Generation were allowed to covary in Wave 2. Finally, Initial Sound Oddity and Initial Syllable Oddity, and Syllable Completion and Sound Isolation were allowed to covary in Wave 3. In turn, in

Table 2
Fit Indices for Models of Participants’ Phonological Abilities at All Three Ages

<table>
<thead>
<tr>
<th>Wave and model</th>
<th>S-Bχ²</th>
<th>df</th>
<th>CFI</th>
<th>NFI</th>
<th>RMSEA</th>
<th>90% CI</th>
<th>AIC</th>
<th>ECVI</th>
<th>S-BΔχ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1: One factor (two parameters)</td>
<td>96.42***</td>
<td>33</td>
<td>.90</td>
<td>.85</td>
<td>.08</td>
<td>[.06, .10]</td>
<td>140.42</td>
<td>0.503</td>
<td></td>
</tr>
<tr>
<td>M2: Two factors, oblique</td>
<td>71.43***</td>
<td>32</td>
<td>.94</td>
<td>.90</td>
<td>.06</td>
<td>[.04, .08]</td>
<td>117.43</td>
<td>0.421</td>
<td></td>
</tr>
<tr>
<td>M3: Two factors, orthogonal</td>
<td>163.70***</td>
<td>33</td>
<td>.79</td>
<td>.75</td>
<td>.12</td>
<td>[.10, .14]</td>
<td>207.70</td>
<td>0.744</td>
<td></td>
</tr>
<tr>
<td>M4: One factor (three parameters)</td>
<td>156.98***</td>
<td>34</td>
<td>.80</td>
<td>.76</td>
<td>.11</td>
<td>[.09, .13]</td>
<td>198.98</td>
<td>0.713</td>
<td></td>
</tr>
<tr>
<td>M5: Three factors, oblique</td>
<td>64.23**</td>
<td>31</td>
<td>.95</td>
<td>.91</td>
<td>.06</td>
<td>[.04, .08]</td>
<td>112.23</td>
<td>0.402</td>
<td></td>
</tr>
<tr>
<td>M6: Three factors, orthogonal</td>
<td>247.72***</td>
<td>34</td>
<td>.65</td>
<td>.63</td>
<td>.15</td>
<td>[.13, .17]</td>
<td>209.72</td>
<td>1.038</td>
<td></td>
</tr>
<tr>
<td>M7: One factor (second order)</td>
<td>71.43***</td>
<td>32</td>
<td>.94</td>
<td>.90</td>
<td>.06</td>
<td>[.04, .08]</td>
<td>117.43</td>
<td>0.421</td>
<td></td>
</tr>
<tr>
<td>M8: Nested factor (two parameters)</td>
<td>47.72***</td>
<td>23</td>
<td>.96</td>
<td>.93</td>
<td>.06</td>
<td>[.03, .08]</td>
<td>117.72</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td>M9: Nested factor (three parameters)</td>
<td>45.38***</td>
<td>23</td>
<td>.96</td>
<td>.93</td>
<td>.06</td>
<td>[.03, .08]</td>
<td>109.38</td>
<td>0.392</td>
<td></td>
</tr>
<tr>
<td>Wave 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1: One factor (two parameters)</td>
<td>157.01***</td>
<td>33</td>
<td>.84</td>
<td>.81</td>
<td>.12</td>
<td>[.10, .13]</td>
<td>201.01</td>
<td>0.720</td>
<td></td>
</tr>
<tr>
<td>M2: Two factors, oblique</td>
<td>59.44**</td>
<td>32</td>
<td>.97</td>
<td>.93</td>
<td>.05</td>
<td>[.03, .08]</td>
<td>105.44</td>
<td>0.378</td>
<td></td>
</tr>
<tr>
<td>M3: Two factors, orthogonal</td>
<td>190.87***</td>
<td>33</td>
<td>.80</td>
<td>.77</td>
<td>.13</td>
<td>[.11, .15]</td>
<td>234.87</td>
<td>0.842</td>
<td></td>
</tr>
<tr>
<td>M4: One factor (three parameters)</td>
<td>138.86***</td>
<td>34</td>
<td>.87</td>
<td>.83</td>
<td>.10</td>
<td>[.09, .12]</td>
<td>180.86</td>
<td>0.648</td>
<td></td>
</tr>
<tr>
<td>M5: Three factors, oblique</td>
<td>55.23***</td>
<td>31</td>
<td>.97</td>
<td>.93</td>
<td>.05</td>
<td>[.03, .08]</td>
<td>103.23</td>
<td>0.370</td>
<td></td>
</tr>
<tr>
<td>M6: Three factors, orthogonal</td>
<td>241.64***</td>
<td>34</td>
<td>.74</td>
<td>.71</td>
<td>.15</td>
<td>[.13, .16]</td>
<td>283.64</td>
<td>1.017</td>
<td></td>
</tr>
<tr>
<td>M7: One factor (second order)</td>
<td>59.44***</td>
<td>32</td>
<td>.97</td>
<td>.93</td>
<td>.05</td>
<td>[.03, .08]</td>
<td>105.44</td>
<td>0.378</td>
<td></td>
</tr>
<tr>
<td>M8: Nested factor (two parameters)</td>
<td>49.84***</td>
<td>23</td>
<td>.97</td>
<td>.94</td>
<td>.06</td>
<td>[.04, .09]</td>
<td>113.84</td>
<td>0.408</td>
<td></td>
</tr>
<tr>
<td>M9: Nested factor (three parameters)</td>
<td>45.54***</td>
<td>23</td>
<td>.97</td>
<td>.95</td>
<td>.06</td>
<td>[.03, .08]</td>
<td>109.54</td>
<td>0.393</td>
<td></td>
</tr>
<tr>
<td>Wave 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1: One factor (two parameters)</td>
<td>199.60***</td>
<td>33</td>
<td>.72</td>
<td>.68</td>
<td>.13</td>
<td>[.12, .15]</td>
<td>243.60</td>
<td>0.873</td>
<td></td>
</tr>
<tr>
<td>M2: Two factors, oblique</td>
<td>66.70***</td>
<td>32</td>
<td>.94</td>
<td>.89</td>
<td>.06</td>
<td>[.04, .08]</td>
<td>112.70</td>
<td>0.404</td>
<td></td>
</tr>
<tr>
<td>M3: Two factors, orthogonal</td>
<td>386.58***</td>
<td>33</td>
<td>.40</td>
<td>.39</td>
<td>.20</td>
<td>[.18, .21]</td>
<td>430.58</td>
<td>1.543</td>
<td></td>
</tr>
<tr>
<td>M4: One factor (three parameters)</td>
<td>261.13***</td>
<td>34</td>
<td>.61</td>
<td>.59</td>
<td>.16</td>
<td>[.14, .17]</td>
<td>303.13</td>
<td>1.086</td>
<td></td>
</tr>
<tr>
<td>M5: Three factors, oblique</td>
<td>71.83***</td>
<td>31</td>
<td>.93</td>
<td>.89</td>
<td>.07</td>
<td>[.05, .09]</td>
<td>119.83</td>
<td>0.429</td>
<td></td>
</tr>
<tr>
<td>M6: Three factors, orthogonal</td>
<td>457.04***</td>
<td>34</td>
<td>.28</td>
<td>.28</td>
<td>.21</td>
<td>[.19, .23]</td>
<td>499.04</td>
<td>1.789</td>
<td></td>
</tr>
<tr>
<td>M7: One factor (second order)</td>
<td>66.70***</td>
<td>32</td>
<td>.94</td>
<td>.89</td>
<td>.06</td>
<td>[.04, .08]</td>
<td>112.70</td>
<td>0.404</td>
<td></td>
</tr>
<tr>
<td>M8: Nested factor (two parameters)</td>
<td>39.84***</td>
<td>23</td>
<td>.97</td>
<td>.94</td>
<td>.06</td>
<td>[.02, .07]</td>
<td>103.84</td>
<td>0.372</td>
<td></td>
</tr>
<tr>
<td>M9: Nested factor (three parameters)</td>
<td>41.19***</td>
<td>23</td>
<td>.97</td>
<td>.93</td>
<td>.06</td>
<td>[.03, .08]</td>
<td>105.19</td>
<td>0.377</td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 280. Goodness-of-fit statistics were calculated using Satorra–Bentler (S-B) robust technique. Wave 1 mean age = 5.8 (SD = 0.31); Wave 2 mean age = 6.6 (SD = 0.31); Wave 3 mean age = 7.7 (SD = 0.32). Chi-square difference tests are comparisons to the one-factor model; models are presented in the specific order to enhance visibility of the differences. CFI = comparative fit index; NFI = normed fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval; AIC = Akaike index criterion; ECVI = expected cross-validation index; M = Model. a Model 8 is compared with Model 7 in all three waves. b Model 9 is compared with Model 5 in all three waves. * p < .05. ** p < .01. *** p < .001.
Model 9, Syllable Completion and Syllable Segmentation were constrained to be equal.

As shown in Table 2, the nested-factor model (Model 8) produced an S-Bχ² that had a statistically significantly better fit to the data than the two-factor oblique and the one-factor second-order models, at least in Waves 1 and 3 (p < .05). The model indices indicated that the nested-factor model was good fitting in Wave 1, S-Bχ²(23, N = 280) = 47.72, p < .001; CFI = .96; NFI = .93; and RMSEA = .06 (90% CI [.03, .08]); in Wave 2, S-Bχ²(23, N = 280) = 49.84, p < .001; CFI = .96; NFI = .94; and RMSEA = .06 (90% CI [.04, .09]); and in Wave 3, S-Bχ²(23, N = 280) = 39.84, p < .001; CFI = .97; NFI = .94; and RMSEA = .05 (90% CI [.02, .07]). Moreover, a careful look at the factor loadings suggests that the general phonological factor accounted for the largest portion of variance for almost all the subtests in all three waves, with these factor loadings on the general factor being significant in all instances, as opposed to the factor loadings on the specific phonological ability factors.

The results for Model 9, as shown in Table 2, were similar to those of Model 8. The nested-factor model with three parameters (Model 9) produced an S-Bχ² that had a statistically significant better fit to the data than the three-factor oblique model in Waves 1 and 3 (p < .01). The model indices indicated that the nested-factor model was good fitting in Wave 1, S-Bχ²(23, N = 280) = 45.38, p < .001; CFI = .96; NFI = .93; and RMSEA = .06 (90% CI [.03, .08]); in Wave 2, S-Bχ²(23, N = 280) = 45.54, p < .001; CFI = .97; NFI = .95; and RMSEA = .06 (90% CI [.03, .08]); and in Wave 3, S-Bχ²(23, N = 280) = 41.19, p < .001; CFI = .97; NFI = .93; and RMSEA = .06 (90% CI [.03, .08]). Similarly, as suggested by the factor loadings, the general phonological factor accounted again for the largest portion of variance for almost all the subtests in all 3 years.

In short, the nested-factor models (Models 8 and 9) had a better fit to the data and better represented an overall unified construct of phonological ability compared with the two-factor oblique and the one-factor second-order and three-factor models, respectively. However, these two models were similar to each other on the basis of overall goodness of fit, in spite of their associated lower AIC and ECVI values that seem to favor Model 9 over Model 8 in Waves 1 and 2 and Model 8 over Model 9 in Wave 3. It is more likely that these negligible differences are better explained as a result of the complexity of each of the models (two vs. three parameters) rather than as a result of a statistical comparison of the two models (Brown, 2006). For this reason, we adopted the simpler (Model 8; nested-factor model with two parameters) and better identifiable model (with at least four indicators per parameter) for all subsequent analyses. Parameter values for the nested-factor two-parameter model applied to each age group are shown in Figures 1, 2, and 3.

**Longitudinal Factorial Invariance**

To investigate the factor structure and measurement invariance of phonological abilities across time, we conducted factorial invariance and partial factorial or metric invariance (Byrne, Shavelson, & Muthén, 1989) across and between adjacent waves (Meredith, 1993). Given that the dimensionality of phonological abilities could be established through the first set of analyses, the objective in this set of analyses was to test the degree to which the obtained constructs are the same across time, meaning yielding

---

**Figure 1.** Nested-factor model of phonological abilities in Wave 1. The squares represent observed variables. The circles represent the latent phonological ability variables of supraphonemic and phonemic sensitivity skills and of the general phonological ability factor at first-order. Coefficients are presented in standardized form. The parameter values for Wave 1 are as follows: S-Bχ²(23, N = 280) = 47.72, p < .001; CFI = .96; NFI = .93; and RMSEA = .06 (90% CI [.03, .08]); AIC = 117.72. SPS = supraphonemic sensitivity; PA = phonological ability; PS = phonemic sensitivity; RO = Rhyme Oddity; RG = Rhyme Generation; SS = Syllable Segmentation; SC = Syllable Completion; FSO = Final Syllable Oddity; ISY = Initial Syllable Oddity; ISO = Initial Sound Oddity; SI = Sound Isolation; PE = Phoneme Elision; BL = Blending.
identical factor structure. To do so, the nested factor model with two parameters (Model 8) was used as a baseline model for the factorial invariance analysis.

In general, factorial invariance routine involves various levels from weaker forms of configural invariance to strict full metric invariance (Horn & McArdle, 1992). It involves also testing and comparing nested models that impose successive restrictions on model parameters (Vandenberg & Lance, 2000). Four hierarchical steps of measurement invariance are commonly tested, from less to more constrained: configural, weak (or metric), strong (or scalar), and strict (Meredith, 1993). Configural invariance is the first step of a measurement invariance procedure and is satisfied when the same pattern of fixed and free factor loadings (and other parameters) is invariant across time. It is important to know that the same factor structure is present at all testing occasions; thus, configural invariance must be established for subsequent steps to be meaningful. If configural invariance is not maintained across time, then it is likely that developmental processes may have produced changes in ability structure. In weak or metric invariance, all factor coefficients are constrained to be equal across time, whereas construct variances and covariances are free to vary (Widaman & Reise, 1997). At this step, therefore, the equality of factor loadings is tested. Metric invariance requires not only that measures have their loadings on the same ability construct, but also that the magnitude of the loadings can be constrained equally across time or between adjacent groups. However, in the case of phonological skills, it seems reasonable to hypothesize that even if configural invariance can be established, changes in development in conjunction with reading instruction, could cause differences in the magnitude of the factor loadings for some of the phonological measures. That is, it may not be possible to obtain complete metric invariance in some of the tasks, although the tasks tap the same ability factor across time. If this hypothesis turns out to be true in our data, the results would be in agreement with the notion of heterotypic continuity of phonological abilities, as initially suggested by Anthony et al. (2003). This would simply mean that a measure may be expected to be a better index of a latent phonological sensitivity at one point in development than at another. To test strong factorial invariance, it is required to add equality constraints to the intercepts across time, testing the equality of the indicator intercepts. Finally, in the strict variance, the equality of indicator residual variances is examined (Brown, 2006).

To test configural invariance of the phonological abilities, we analyzed data by fitting the three waves of data with the nested-factor two-parameters model. This initial baseline model provided the basis for the comparison with the three subsequent models in the invariance hierarchy, testing, in turn, for weak, strong, and strict invariance. Model fit was evaluated using chi-square difference tests and change in relative fit indexes. Tests of invariance were applied using the robust maximum likelihood estimator (S-B robust technique; Byrne, 2006).

Table 3 presents the hierarchy of models describing the tests of factor and measurement invariance and providing evidence for differences and similarities across the three assessments (Waves 1 to 3): (a) same pattern of fixed and free loadings for each wave (configural invariance or baseline model), (b) factor loadings invariant across waves, (c) indicators’ intercepts invariant, and (d) equality of indicator residual variances. Evaluation of fit indexes among the configural, weak, strong, and strict models revealed a statistically significant difference in chi-square values and relative

---

**Figure 2.** Nested-factor model of phonological abilities in Wave 2. The squares represent observed variables. The circles represent the latent phonological ability variables of supraphonemic and phonemic sensitivity skills and of the general phonological ability factor at first-order. Coefficients are presented in standardized form.

The parameter values for Wave 2 are as follows: S-By²(23, N = 280) = 49.84, p < .001; CFI = .96; NFI = .94; and RMSEA = .06 (90% CI [.04, .09]); AIC = 113.84. SPS = supraphonemic sensitivity; PA = phonological ability; PS = phonemic sensitivity; RO = Rhyme Oddity; RG = Rhyme Generation; SS = Syllable Segmentation; SC = Syllable Completion; FSO = Final Syllable Oddity; ISY = Initial Syllable Oddity; ISO = Initial Sound Oddity; SI = Sound Isolation; PE = Phoneme Elision; BL = Blending.

---

**Table 3.** Hierarchy of models describing the tests of factor and measurement invariance.
fit indexes between each model, suggesting that the configural model is the only acceptable model, on the basis of the comparisons among the models and on the basis of robust criteria for model fit: S-B $\chi^2(23, N = 280) = 622.36$, $p < .001$; CFI = .98; NFI = .96; RMSEA = .05 (90% CI [.05 to .06]). This finding provides evidence that the obtained constructs are the same across time.

Next, partial metric invariance was conducted because weak invariance produced a significant increase in model, $\chi^2 (S-B \Delta \chi^2 = 75.17, p < .01)$, indicating that some of the factor loadings were not equal across time. This means that, although the results suggest that for all three waves the data were fairly well described by the general phonological ability, supraphonemic sensitivity, and phonemic sensitivity factors, they do not necessarily imply that

![Diagram of nested-factor model of phonological abilities in Wave 3.](image)

**Figure 3.** Nested-factor model of phonological abilities in Wave 3. The squares represent observed variables. The circles represent the latent phonological ability variables of supraphonemic and phonemic sensitivity skills and of the general phonological ability factor at first-order. Coefficients are presented in standardized form. The parameter values for Wave 3 are as follows: S-B $\chi^2(23, N = 280) = 39.84$, $p < .001$; CFI = .97; NFI = .94; and RMSEA = .05 (90% CI [.02, .07]); AIC = 103.84. SPS = supraphonemic sensitivity; PA = phonological ability; PS = phonemic sensitivity; RO = Rhyme Oddity; RG = Rhyme Generation; SS = Syllable Segmentation; SC = Syllable Completion; FSO = Final Syllable Oddity; ISO = Initial Syllable Oddity; SI = Sound Isolation; PE = Phoneme Elision; BL = Blending.

### Table 3
**Fit Indexes for the Second-Order Model in the Invariance Sequence**

<table>
<thead>
<tr>
<th>Model</th>
<th>Versus</th>
<th>S-B $\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>NFI</th>
<th>RMSEA</th>
<th>90% CI</th>
<th>S-B $\Delta \chi^2$</th>
<th>S-B $\Delta$df</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Configural, nested-factor model</td>
<td>—</td>
<td>622.36</td>
<td>340</td>
<td>.98</td>
<td>.96</td>
<td>.05</td>
<td>[.05, .06]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. Weak (metric)</td>
<td>Model 1</td>
<td>713.66</td>
<td>357</td>
<td>.92</td>
<td>.84</td>
<td>.06</td>
<td>[.05, .07]</td>
<td>75.17***</td>
<td>17</td>
</tr>
<tr>
<td>3. Strong</td>
<td>Model 2</td>
<td>2,690.11</td>
<td>375</td>
<td>.80</td>
<td>.77</td>
<td>.13</td>
<td>[.13, .14]</td>
<td>142.17***</td>
<td>18</td>
</tr>
<tr>
<td>4. Strict</td>
<td>Model 3</td>
<td>3,027.79</td>
<td>394</td>
<td>.80</td>
<td>.77</td>
<td>.13</td>
<td>[.13, .14]</td>
<td>94.46***</td>
<td>19</td>
</tr>
</tbody>
</table>

Partial metric invariance: Nested models (between adjacent grades)

| 5. Baseline between Waves 1 to 2   |                             | 215.78        | 134 | .99   | .98   | .05         | [.04, .06]       | —                  | —               |
| Waves 1 to 2 with equality constraints on all factor loadings | Model 5                   | 239.36        | 144 | .99   | .98   | .05         | [.04, .06]       | 22.48*             | 10              |
| Waves 1 to 2 with equality constraints on invariant lambda   | Model 5                   | 230.07        | 142 | .99   | .98   | .05         | [.04, .06]       | 14.03              | 8               |
| 6. Baseline between Waves 2 to 3   |                             | 259.84        | 134 | .99   | .98   | .06         | [.05, .07]       | —                  | —               |
| Waves 2 to 3 with equality constraints on all factor loadings | Model 6                   | 307.69        | 144 | .99   | .98   | .06         | [.05, .07]       | 47.93***           | 10              |
| Waves 2 to 3 with equality constraints on invariant lambda   | Model 6                   | 266.25        | 137 | .99   | .98   | .06         | [.05, .07]       | 6.46               | 3               |

Note. Chi-square difference tests are indicated with the model numbers in the second column. S-B = Satorra–Bentler; CFI = comparative fit index; NFI = normed fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval. 

*p < .05. **p < .01. ***p < .001.
the actual factor loadings are the same across waves. Thus, the hypothesis of the equivalency of factor loadings across time or of the parameters being invariant across time was tested by imposing equality constraints on lambda between adjacent years. Specifically, two types of analyses were performed to test further the parsimony of the nested-factor model between adjacent years (from Wave 1 to Wave 2 and from Wave 2 to Wave 3): (a) in the first set of analyses, we held equality constraints on lambda on all factor loadings between adjacent years, and (b) in the second set of analyses, we released equality constraints on lambda for the invariant loadings based on model modification indices. In essence, partial invariance evaluation of the model fit was used as a post hoc procedure so we could provide a substantively compelling account for the parameters or the sources of noninvariance (Byrne et al., 1989). Also, given the nonindependence of the tests, we decided not to explore partial measurement invariance with alternative series of tests, but rather to test initially the hypothesis of an invariant pattern of factor loadings by constraining all lambda parameters to be equal. All four nested models (two models between Waves 1 and 2 and two models between Waves 2 and 3) were compared with their equivalent baseline nested-factor model. Table 3 shows the fit indexes for all these nested models with constrained factor loadings to equality.

Both analyses between adjacent years, in which equality constraints on lambda were held on all factor loadings resulted in better model fits compared with the initial models with no imposed parameter constraints. This means that certain variables caused the missfit in each of the adjacent years. Specifically, the results indicated that Initial Sound Oddity ($\chi^2 = 5.91, p < .05$) and Syllable Completion ($\chi^2 = 5.33, p < .05$) were inconsistent from Wave 1 to Wave 2. Similarly, Blending ($\chi^2 = 8.45, p < .01$), Final Sound Oddity ($\chi^2 = 7.33, p < .05$), and Syllable Completion ($\chi^2 = 7.46, p < .01$) were inconsistent from Wave 2 to Wave 3. In contrast, the increase observed in the model chi-square in the case of the second set of analyses between adjacent grades where equality constraints were released for the noninvariant lambda was insignificant for both Waves 1 to 2 and Waves 2 to 3 comparisons. These results indicated that the remaining eight variables constrained to be equal from Waves 1 to 2 and those seven variables constrained to be equal from Waves 2 to 3 were invariant across the adjacent years they were tested.

**Linear Regression Analyses Within SEM**

Our findings raised an interesting question. What are the roles of the specific and phonological ability factors in Greek and their predictive validity to reading outcomes, such as word reading fluency? To address this issue, we conducted linear regression analyses within SEM to test the significant predictive contribution of the latent specific phonological constructs derived from these models, namely, the supraphonemic and phonemic sensitivity, to participants’ word reading skills, concurrently (in each of Waves 2 and 3).

These analyses provide converging evidence for the factor structure of the phonological abilities in Greek. Evidence for the differential contribution of the latent phonological sensitivity abilities (and, thus, for the distinct abilities hypothesis) would be provided if these two factors account for a unique amount of variance regardless of the order entered in the regression equation. Evidence for the equal contribution of the latent phonological sensitivity abilities (and, thus, for the unitary abilities hypothesis) would be provided if either of those two factors does not account for any additional variance when entered in the regression equation after controlling for the effects of the other.

To perform these analyses, the factor scores of the latent constructs were extracted from the SEM model. Also, a composite score, expressed in z-score units, of word reading fluency for both Word Identification (tapping word recognition) and Word Attack (tapping decoding) was calculated. This latter score was entered as the dependent variable in the regression analyses. The phonological ability factor scores, as yielded from Waves 2 and 3, were entered separately and interchangeably as predictors in each of the analyses to estimate their unique effect on word reading fluency. This means that the analysis was done twice, first with the supraphonemic sensitivity and phonemic sensitivity and then with the order of entry of the predictors being reversed. The total amount of explained variance also was obtained (total $R^2$) in each case.

The correlation analysis showed that the phonological ability factors were significantly correlated with word reading fluency in both waves ($rs$ ranged from .30 to .41, all at $p < .001$). The results of the hierarchical regression analyses showed that in all instances, both supraphonemic and phonemic sensitivity factors accounted for unique variance in word reading fluency measures only when they were entered first in the equation (Table 4). This means that when any one of these factors was entered in the regression equation after controlling for the effects of the other, it did not account for any variance in word reading fluency, suggesting that these two factors share the same predictive variance. It is also worth noting that the total amount of explained variance in all instances was equal to the variance explained by any of the two factors. These results provide converging evidence for the conceptualization of phonological abilities as a unitary construct, as the contribution of any of the two latent factors to the measured reading skill becomes insignificant after controlling for the effects of the other factor.

**Table 4**

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>$\beta$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>SPS factor</td>
<td>.398</td>
<td>.16***</td>
</tr>
<tr>
<td>2</td>
<td>PS factor</td>
<td>-.302</td>
<td>.00</td>
</tr>
<tr>
<td>1</td>
<td>PS factor</td>
<td>.403</td>
<td>.16***</td>
</tr>
<tr>
<td>2</td>
<td>SPS factor</td>
<td>-.602</td>
<td>.00</td>
</tr>
<tr>
<td>Total $R^2$</td>
<td></td>
<td></td>
<td>.16***</td>
</tr>
<tr>
<td>Wave 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>SPS factor</td>
<td>.409</td>
<td>.17***</td>
</tr>
<tr>
<td>2</td>
<td>PS factor</td>
<td>.507</td>
<td>.00</td>
</tr>
<tr>
<td>1</td>
<td>PS factor</td>
<td>.410</td>
<td>.17***</td>
</tr>
<tr>
<td>2</td>
<td>SPS factor</td>
<td>-.097</td>
<td>.00</td>
</tr>
<tr>
<td>Total $R^2$</td>
<td></td>
<td></td>
<td>.17***</td>
</tr>
</tbody>
</table>

*Note. n = 280. SPS factor = Supraphonemic Sensitivity factor; PS factor = Phonemic Sensitivity factor.*** $p < .001$.**
The primary aim of the present study was to examine the theoretically appropriate conceptualization and measurement invariance of phonological abilities in a (sufficiently) transparent language. The findings showed that the conceptualization of phonological abilities was captured most accurately as a nested-factor model, which consisted of a first-order general factor, a first-order supraphonemic factor, and a first-order phonemic factor, representing a unified single construct. The important aspect of this finding is that it was obtained after testing and comparing across time a number of alternative, theory-driven models that have been proposed for the conceptualization of phonological abilities in different languages, albeit in isolation.

In what was termed as the best fitting model, there were six unique indicators of supraphonemic sensitivity, namely, Rhyme Oddity, Rhyme Generation, Syllable Segmentation, Syllable Completion, Initial Syllable Oddity, and Initial Syllable Oddity. Similarly, there were four unique indicators of phonemic sensitivity, namely, Initial Sound Oddity, Sound Isolation, Phoneme Elision, and Blending. There was also a robust general factor on which all 10 phonological tasks were specified to weight. Furthermore, all the coefficients on the general phonological factor were statistically significant, accounting for the largest portion of variance for the vast majority of the subtests in all three waves, as opposed to the coefficients on the specific phonological ability factors. In conjunction with the results from the factorial structure invariance analysis and the linear regression analysis, the answer to the conceptualization question is clear and robust supporting the presence of a single phonological sensitivity construct.

Three aspects of the present findings significantly contribute to the existing literature. First, the conceptualization of phonological sensitivity is supported as a unitary construct in Greek language, a language with transparent orthography, strengthening existing evidence for the universality of the construct as unitary in transparent and nontransparent languages. It is, indeed, important to see the results of the present study converge with the robust evidence that has been accumulated particularly in English (Anthony & Lonigan, 2004; Schatschneider et al., 1999; Stahl & Murray, 1994; Wagner et al., 1997). Even though the distance between Greek and English language is still under investigation, the aforementioned convergence allows some degree of generalizability across languages. In that respect, we reinstate Anthony and Lonigan’s (2004) call for the need to recognize the unitary structure of phonological ability in theories of phonological sensitivity development and its relations with literacy acquisition.

Second, the unitary construct conceptualization remains invariant across time in early years. Specifically, using an advanced technique, that of longitudinal factorial invariance within SEM, the findings of the present study resolved the contradictory findings by Schatschneider et al. (1999) and those by Anthony and Lonigan (2004). We were able to directly test measurement invariance and detect possible sources of invariance (in relation to certain tasks). Recall that even though the findings of the aforementioned two studies agreed on the conceptualization of phonological sensitivity as a unitary construct, they disagreed on whether phonological sensitivity was invariant across time and measures. Our findings confirm that the aforementioned contradictory findings were most likely due to the different methodological approaches and analyses used, which, in turn, did not allow for a direct test of measurement invariance longitudinally.

The third aspect of the present findings that contributes to the literature relates to our methodological approach. Specifically, the finding that a significant improvement in model fit can be achieved by modeling the various subtests as a nested- and first-order factor demonstrates the importance of this approach for the relevant research. Despite the well-documented advantages of modeling factors as completely first-order orthogonal models in the area of intelligence research (e.g., Gustafsson & Balke, 1993), this model had never been used, to date, to examine the plausible conceptualization of orthogonal first-order solutions of phonological tests. Rather, the literature revolved around the more traditional hierarchical, oblique, or orthogonal models supporting either the distinct abilities hypothesis or the unitary hypothesis. To the best of our knowledge, this is the first study that examined the plausibility of this type of model in reading research and demonstrates its appropriateness for answering the unitary versus distinct construct type of questions.

Indeed, the nested-factor model that eventually emerged as metrically the most parsimonious was also theoretically superior to the two- and three-factor oblique and the one-factor second-order models for three important reasons. First, the two- or three-factor oblique models only indirectly account for a possible general phonological ability factor. Second, by setting the model with a general factor indexed by all indicator variables and phonological sensitivity factors that are orthogonal to both the general factor and each other, the constraint to refer to a second-order latent phonological ability variable defined by only two or three indicators was also circumvented. Finally, with the use and acceptance of a nested-factor model, it was also possible to elucidate the relation of the phonological tasks to the general phonological ability independently of their relation to the specific latent factors. Instead, all indicators assume a direct relation with the general phonological factor apart from the strong relation to their specific latent factor. For all of these reasons, the use of the nested-factor modeling provided an innovative framework for the conceptualization of phonological abilities as a unidimensional construct. By accepting the nested-factor model as the most parsimonious, we do not ignore that the other three models (the two- and three-factor oblique and the one-factor second-order models) that also represent phonological abilities as a unidimensional construct were more parsimonious than their orthogonal counterparts. At the same time, we do not ignore that the general phonological ability factor may depend to some extent on a general ability factor. In an effort to better understand the dependency of this phonological ability factor on general ability, we regressed on it the nonverbal ability scores of the participants in all three waves. Results showed that there is a moderate relation between phonological ability factor and ability scores ($r = .31$ to .41; $p$ values deriving from model comparisons among the new models with the best-fitting models in each wave being nonsignificant), suggesting that the general phonological ability factor reflects what it really measures, namely, a single phonological ability (see also Anthony & Lonigan, 2004).

As we mentioned previously, determining whether the best-fitting model was invariant across time would add value to the search for the best conceptualization of the phonological sensitivity construct. This nested-factor model with two parameters, the best-fitting model, exhibited configural or structural invariance because the magnitude of the relations among the variables to the constructs was the same across time. This finding was important because it indicated that the latent constructs (both the general and the specific phonological factors) represented what was common
among the constituent variables, and this representation was not different across age.

Equally important was to determine the extent to which all tasks measured the phonological ability construct to the same degree across the age groups, indicating metric invariance. For this purpose, the earlier analysis was followed by partial invariance analysis across time. The findings provided evidence for partial metric invariance. This means that although the structure of the nested-factor model was the same across waves, this was not true for all tasks measuring the single phonological ability construct. Instead, it is more appropriate to say that the analysis identified a subset of parameters in the model that were invariant and another subset of parameters that varied between adjacent groups. However, it is important to note that this noninvariant subset constituted only a small portion of the model (only two variables from Wave 1 to Wave 2 and three variables from Wave 2 to Wave 3); thus, meaningful comparisons across groups could still be made (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). Our data also fulfilled two additional criteria for the noninvariant items (see Cheung & Rensvold, 1998, for a detailed description of these criteria): (a) all the noninvariant items related strongly and meaningfully to the constructs in all waves, and (b) configural invariance held, that is, the noninvariant items loaded on the same factors in all waves. Thus, there was a reasonable degree of congruence between the three waves concerning the operationalization of the phonological ability construct. In fact, our data showed that Syllable Completion and Initial Sound Oddity were the only noninvariant parameters in the analysis from Wave 1 to Wave 2. Similarly, Syllable Completion, Blending, and Final Syllable Oddity were the only noninvariant parameters from Wave 2 to Wave 3. All these variables met the criteria for the noninvariant items.

Why the aforementioned tasks showed metric partial invariance is an interesting question raised by the present findings. As far as the Syllable Completion task is concerned, previous research in both Greek (Loizou & Stuart, 2003; Papadopoulos, Spanoudis, & Kendeou, 2009) and English (Muter & Diethelm, 2001) has shown that it is one of the relatively easier tasks that may approach ceiling early on, even in kindergarten. Similarly, with regard to the Initial Sound Oddity, Aidinis and Nunes (2001) have reported that Greek preschoolers and first graders perform better in the initial sound tasks than in the final sound tasks. This may be partly attributed to the role of the vowel acting as a syllable at the beginning of the word, making some of the vowel–syllable tasks easier to perform (Papadopoulos, Spanoudis, & Kendeou, 2009). As far as Blending is concerned, it has been found to be a rather demanding task for young children in Greek (Papadopoulos, Spanoudis, & Kendeou, 2009), probably because it is dependent on letter knowledge more than any other aspect of phonological ability (Manolitsis & Tafa, 2011). As far as Final Syllable Oddity is concerned, it appears to be a difficult task at the syllabic level because adequate performance on this task requires primarily the recognition of the final syllable structure and then the individual phoneme that differs in the final syllable (i.e., the consonant in the type of the task used in the present study). This process requires higher planning control at the syllabic level from the child’s part (Botinis, Fournakis, & Prinou, 1999). In conclusion, in a longitudinal study in which phonological skills are developing, a lack of invariance over time is reasonably to be expected, as a result of increasing reading instruction and experience. These findings are also in agreement with the notion of heterotypic continuity of phonological abilities, as initially suggested by Anthony et al. (2003). Phonological abilities appeared to develop in a continuous way, with a structural stability, as a single ability across the 3 years of assessment. Its heterotypic nature is witnessed by the various tasks used to assess it and as a function of development of some of the tasks. At the very least, to the degree that these findings are replicable, a future direction of the relevant research could focus on addressing the issue of partial invariance of the phonological abilities in different age groups and languages.

From a practical point of view, the current set of findings provides a conceptual framework for implementing appropriate teaching practices that promote the development of phonological skills in early years that is different from the development of these skills in English. Specifically, the development of phonological sensitivity skills in Greek does not parallel the linguistic onset-rime models according to which development occurs from syllables, to onsets and rimes, to phonemes (Papadopoulos, Spanoudis, & Kendeou, 2009). Rather, access to the higher and larger onset-rime units (supraphonemic sensitivity) develops naturally and in a way similar to the access of the lower level of phonemes (phonemic sensitivity) as the present study has shown. In this context, the presence of a unitary construct suggests that phonological skills do not consist of a set of discrete skills that can be taught in isolation of each other, but rather a family of skills that are interrelated and interdependent; thus, this set of skills can be taught concurrently. Furthermore, and in conjunction with the item analysis performed on this battery of tasks in our previous work, there is a developmental progression that can be taken into account during instruction by considering three different levels that are integral components in each task: (a) grain size (supraphonemic vs. phonemic sensitivity); (b) position (initial vs. final or middle); and (c) task demands (analysis or synthesis). Our findings suggest that tasks placing emphasis on supraphonemic sensitivity level, final position of the target, and analysis processing are easier to start working with than tasks focusing on phonemic sensitivity level, initial position, and synthesis processing.

What this study has offered to the ongoing discussion about the nature of phonological abilities is a different angle to the unitary-or-distinct dilemma. Perhaps, the main question should not be whether phonological skills develop differently in languages with different level of transparency. Rather, the future of the research on phonological sensitivity ought to lie in its ability to break the boundaries of the various paradigms and build new universal theories. Only then, the various models and stances could be integrated into an overarching scheme that would be able to accommodate the architecture of phonological skills, their development, and their use in different contexts. For the universality and specificity of the concept to be fully understood, the nature of phonological ability ought to be studied systematically and longitudinally across languages.

References


Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological...


Muter, V., & Diethelm, K. (2001). The contribution of phonological skills and letter knowledge to early reading development in a multilingual


Papadopoulos, T. C., & Loizou, M. (2007). Corpus of contemporary Greek word and word frequency lists for ages 5 to 12. Nicosia, Cyprus: Department of Psychology, University of Cyprus.


Received November 5, 2010
Revision received October 14, 2011
Accepted October 22, 2011

336 PAPADOPOULOS, KENDEOU, AND SPANOUDIS