

Research Report

Adapting a receptive vocabulary test for preschool-aged Greek-speaking children

Areti Okalidou[†], Asimina Syrika[‡], Mary E. Beckman[§] and Jan R. Edwards[¶]

[†]Department of Educational and Social Policy, University of Macedonia, Salonika, Greece

[‡]Waisman Center, Department of Communicative Disorders, University of Wisconsin-Madison, Madison, WI, USA

[§]Department of Linguistics, Ohio State University, Columbus, OH, USA

[¶]Department of Communicative Disorders, University of Wisconsin-Madison, Madison, WI, USA

(Received 3 April 2009; accepted 28 January 2010)

Abstract

Background: Receptive vocabulary is an important measure for language evaluations. Therefore, norm-referenced receptive vocabulary tests are widely used in several languages. However, a receptive vocabulary test has not yet been normed for Modern Greek.

Aims: To adapt an American English vocabulary test, the Receptive One-Word Picture Vocabulary Test-II (ROWPVT-II), for Modern Greek for use with Greek-speaking preschool children.

Methods & Procedures: The list of 170 English words on ROWPVT-II was adapted by (1) developing two lists (A and B) of Greek words that would match either the target English word or another concept corresponding to one of the pictured objects in the four-picture array; and (2) determining a developmental order for the chosen Greek words for preschool-aged children. For the first task, adult word frequency measures were used to select the words for the Greek wordlist. For the second task, 427 children, 225 boys and 202 girls, ranging in age from 2;0 years to 5;11 years, were recruited from urban and suburban areas of Greece. A pilot study of the two word lists was performed with the aim of comparing an equal number of list A and list B responses for each age group and deriving a new developmental list order.

Outcomes & Results: The relative difficulty of each Greek word item, that is, its accuracy score, was calculated by taking the average proportion of correct responses across ages for that word. Subsequently, the word accuracy scores in the two lists were compared via regression analysis, which yielded a highly significant relationship ($R^2 = 0.97$; $p < 0.0001$) and a few outlier pairs (via residuals). Further analysis used the original relative ranking order along with the derived ranking order from the average accuracy scores of the two lists in order to determine which word item from the two lists was a better fit. Finally, new starting levels (basals) were established for preschool ages.

Conclusions & Implications: The revised word list can serve as the basis for adapting a receptive vocabulary test for Greek preschool-aged children. Further steps need to be taken when testing larger numbers of 2;0 to 5;11-year-old children on the revised word list for determination of norms. This effort will facilitate early identification and remediation of language disorders in Modern Greek-speaking children.

Keywords: receptive vocabulary, test adaptation, Greek language tests, Greek.

What this paper adds

It contributes a methodology for adapting vocabulary tests to another language.

Address correspondence to: Areti Okalidou, Department of Educational and Social Policy, University of Macedonia, 156 Egnatias Street, PO Box 1591, Salonika 540 06, Greece; e-mail: okalidou@uom.gr

Introduction

Typically developing children produce their first words at about 12 months. Word learning is one of the first signs that a child is acquiring language normally, and a delay in word learning is one of the first signs that a child is having difficulty with language acquisition. Thus, assessment of word knowledge is critical to any diagnostic evaluation of a child who is suspected of having a language disorder. For very young children, the most efficient way to assess whether their vocabulary comprehension is within normal limits is by asking parents whether their child understands a checklist of early acquired words. In English, receptive vocabulary can be assessed via parent report for children as young as 12 months of age with the MacArthur Communicative Development Inventory (CDI; Fenson *et al.* 1993). This is true also for the many other languages for which an adaptation of the CDI has been developed, such as German (for example, Grimm & Doil, 2000); British English (Hamilton, Plunkett, & Schaefer, 2000); Mexican Spanish (Jackson-Maldonado, Thal, Marchman, Newton, Fenson, & Conboy, 2003); Parisian French (Kern & Langué, 2000); Hebrew (Maital, Dromi, Sagi & Bornstein, 2000); Japanese (Ogura, Yamashita, Murase & Dale, 1993).

By the time a child is 2 years or older, vocabulary size can no longer be assessed reliably by parent report, because a typical 2-year-old has too large a receptive vocabulary for a simple checklist. Therefore, clinicians use standardized tests, such as the *Receptive One Word Picture Vocabulary Test—II* (ROWPVT-II; Brownell 2000) or the *Peabody Picture Vocabulary Test—IV* (PPVT-IV; Dunn *et al.* 2006) to evaluate whether a child's receptive vocabulary is age-appropriate. Receptive vocabulary tests share the following characteristics: A first characteristic is an age-graded list of target words, rank-ordered by difficulty. Most vocabulary tests in English use a large number of nouns, along with some verbs (presented in the present progressive form) and adjectives. A second characteristic is that for each target word, the child sees an array of four pictures and is prompted to point to one of them (for example, 'show me *skunk*' or 'point to *barking*'). A third one is that testing starts at the beginning of a block in the list that is specified based on the child's chronological age. A fourth characteristic is that a criterion referring to the number of consecutive correct responses in the starting block establishes the "basal" item for the child, and testing progresses to later items until a "ceiling" is reached, as defined by another criterion, the number of errors in a block. Otherwise, earlier and earlier items are tested until the 'basal' criterion is met. The raw score then is the number of words below the basal word plus

the number of correct responses between the basal and the ceiling items. A fifth characteristic is that these tests are typically standardized on a large cross-section of the population of interest. For example, ROWPVT-II was normed on 3,661 children from ages 2 to 18 years from all areas of the United States, testing an average of 155 children per age group.

Receptive vocabulary is a particularly important measure for language evaluations because it is highly correlated with verbal IQ (for example, Bornstein and Haynes 1998) and is predictive of later academic performance, particularly in the area of reading (for example, Metsala 1999, Nation 2001, Nation and Snowling 1997, Ehri and Snowling 2005, and Vellutino *et al.* 2004). Unfortunately, receptive vocabulary tests are not available for every language and dialect. Norm-referenced receptive vocabulary tests are available for both American English (for example, Brownell 2000, and Dunn *et al.* 2006) and British English (*British Picture Vocabulary Scale*, 2nd edn; Dunn *et al.* 1997), and also for several other major world languages with relatively large speaker populations. These include, among others, Puerto Rican Spanish (Wiener *et al.* 1978), Mexican Spanish (Brownell 1985), French (Theriault-Whalen *et al.* 1993), Japanese (Ueno *et al.* 1991), Korean (Kim *et al.* 1995), and Cantonese (Cheung *et al.* 1997). However, vocabulary tests are not available for the vast majority of the world's languages, even for major languages with medium-sized populations, such as Greek. The purpose of this study was to adapt an American English vocabulary test (ROWPVT-II) for Modern Greek.

It is important to note that this is an adaptation, and not a translation. A number of researchers (Ali 1967, Hymes 1970, Peña 2007, Roca 1955, Thorndike 1973) have discussed the many reasons why simple translations of language-based tests from one language to another inevitably result in lower reliability and lesser validity of the translated version. First, as Roca (1955) and Peña (2007) note, the original word and the word that translates it may be ranked at different levels of difficulty in the two languages. For example, the word *cup* is a high-frequency word in English and is familiar to preschool children, but the translation equivalent ($\varphi\lambda\iota\tau\zeta\acute{\alpha}\nu$ /fli'dzani/) has a much lower frequency in Greek. In addition, as Hymes (1970) notes, a concept may be represented in one culture, but not in another. For example, the concept named by the English *pitching* exists in most English-speaking cultures, but there is no equivalent term relating to baseball or cricket in Greek. Moreover, even if present, a concept that is represented by one word in one language may only be represented by a phrase rather than a single word in the other language. For example, the Greek word /ma'θitria/ would be translated into English as *female*

student of primary/secondary education. Another problem is that a word may represent one meaning in one language but multiple more or less related meanings in another (Hymes 1970, Simon and Joiner 1976). For example, the translation equivalent of the Greek word *κορίδι* /ko'rijos/ in English is *bug* (as in *insect*), but *bug* is also used as a verb in English to mean either *to wiretap* or *to annoy*.

The above examples make it clear that the overall purpose of developing comparative vocabulary tests across two languages is not well met by directly translating a receptive vocabulary test into another language. As Peña (2007) points out, test adaptation needs encompass functional, cultural, and metric equivalence. Functional equivalence aims to elicit the same target behaviour across languages by finding equivalent words in the second language that meet that language's criteria of acceptability in terms of oddity, familiarity, ease or difficulty with grasping meaning and appropriateness of use in context (Hymes 1970, Peña 2007). Cultural equivalence means that the test items should represent culturally valid meanings in each language. Metric equivalence requires that word selection is made according to item difficulty in each language. In adapting word lists from English to Spanish, Tamayo (1987) showed that performance was more comparable across English and Spanish speakers when the two word lists were matched by item difficulty rather than by translation. Item difficulty can be indexed by lexical frequency (referring to frequency of use) or by directly calculating the percentage of participants who correctly respond to each item (Peña 2007).

Because of these considerations, adaptations of language and achievement tests into a second language have generally relied on a combination of direct translation of some words and the substitution of other words, as needed. Also, after words are chosen, adaptation typically means reordering the items in the adapted test, since relative item difficulty may differ for even very good translation equivalents (Roca 1955, Renzulli and Paulus 1969). As Clark (1965) noted (cited in Simon and Joiner 1976), when developing an adapted version of a Spanish test in Portuguese, test equivalence was achieved by matching item reliability and item difficulty, and then by aiming for a good match in the graded relative order of items across the two tests.

These methodological issues have analogues even in developing a new test for a language where a test already exists. For example, despite the similarities noted above for different tests of receptive vocabulary size for English, Channell and Peek (1989) found only moderate correlations in performance among earlier versions of the PPVT (Dunn and Dunn 1981) and the ROWPVT (Gardner 1985), and two other tests, the Picture Vocabulary subtest of the *Test of Language Development—Primary*

(TOLD-P, Hammill and Newcomer 1982) and the *Expressive One Word Picture Vocabulary Test* (EOWPVT; Gardner 1979) at ages 4;0–5;8 (years;months). Even when an item occurs on two tests, a child may not respond in the same way to it. Such differences may be due to sequencing effects (for example, on one of the lists, the preceding picture array may include a target or foil that acts as a prompt for the common item) or to any number of other differences such as the pictures used or the relative difficulty of differentiating a foil item from the target item. All of these factors point to methodological issues that must be addressed in adapting a test to a new language.

The purpose of this study was to adapt an American English receptive vocabulary test for use with Greek-speaking preschool-aged children in Greece. There are a relatively large number of preschool-aged children in Greece (more than 300,000 in 2002, according to <http://www.earthtrends.wri.org>), but no norm-referenced receptive vocabulary tests are available for this population. Therefore, speech–language pathologists must rely entirely on informal clinical assessment, which is not standardized across different clinicians or clinics. This lack of a norm-referenced vocabulary test for preschool children is a problem for two reasons: First, it makes early diagnosis, remediation, and subsequent assessment difficult because there are no norms for vocabulary development in young children. Second, research is hampered because there is no agreed-upon tool for assessing vocabulary size across different studies.

Method

Test components

The test we chose to adapt is the ROWPVT-II. This test consists of two parts: (1) 170 test plates, each plate being an array of four coloured line drawings; and (2) 170 associated target words of English, each one corresponding to a single picture in the four-picture plate.

With one exception, we used the picture arrays as is. The one exception was the picture array for the target word *W* (that is, the letter name /dʌblju/), where we replaced the target letter with the Greek letter ω (for the letter name $\omega\mu\acute{\epsilon}\gamma\alpha$ /o'mega/), which has a similar shape. This replacement was necessary because the pictured letter may not have been familiar to Greek children.

The main tasks in adapting the list, therefore, were: (1) developing an appropriate list of Greek words that could be used with the picture arrays; and (2) determining an appropriate order for the Greek words that we chose. The next subsection describes the procedures for developing the word list. We determined an appropriate order by administering a pilot test using the original

order of the picture arrays, as described in the following three subsections.

Developing the word list

The procedures for choosing Greek words to go with the ROWPVT-II picture arrays were as follows: For each of the 170 picture arrays, the second author, who is a native Greek speaker, chose one or more candidate items in Greek. For many words, at least one of the candidate items was a direct translation equivalent of the English target item (for example, *αντίχειρας* /a'diçiras/ for *thumb*). Other candidate items either named some other aspect of the picture (for example, *γροθιά* /çro'tça/ 'fist' for the picture associated with *thumb*) or named something else that more or less resembled the picture for the target word (for example, *σκίουρος* /'sciuros/ 'squirrel' for the target *skunk*). She also provided at least one Greek word to name each of the three foil pictures in each four-picture array. The candidate items and foils were then used to build two alternative lists of targets, list A and list B, each containing 170 words. The reason for making two lists of targets was to be able to test two items for some target words where there was a problem with the most direct translation. There were four types of problems:

- First, in thirteen cases the closest translation equivalent is not the most common word for the pictured object/action/attribute. For example, the word *thumb* directly translates to *αντίχειρας* /a'diçiras/, which is not a familiar word for preschool children in Greece, who tend to use the term *δάκτυλο* /'ðaktilo/ 'finger' for all digits. In these cases, we put the direct translation on one list¹ and the more familiar word on the other.
- Second, in five cases the pictured object or action is not a familiar concept for Greek children. For example, there are no skunks in Greece, and the Greek translation of *skunk* *κουνάβι* /ku'navi/ actually names the European polecat, which looks different from the American skunk. In these cases, we tried two other words, where one word named a similar concept and the other named a foil picture. For example, for the *skunk* target array, we used *σκίουρος* /'sciuros/ 'squirrel' as an alternative to the name for the target picture on one list and *ζέβρα* /'zevra/ 'zebra', substituting a foil word, on the other list.
- Third, in twenty-nine cases there were several Greek translations of the target English words, and we had no basis for deciding a priori which would be the appropriate one for that place in the list. For example, *happy* (a word in the first set of words presented to 4-year-olds) can be translated as

χαρούμενος /xa'rumentos/, as *κεφάλτος* /çe'fatots/, or as *ενθουσιασμένος* /enthusja'zmentos/. In these cases, we did one of two things, depending on the familiarity of the different translations. When two of the alternate forms were familiar to children (48 cases), we assigned one to list A and one to list B. When only one of the translations was familiar, we chose that form for one list and the name of a foil picture for the other.

- Finally, there was one case where the pictured object could not be familiar to the children, but there was an easy substitution. This was the case of the roman letter 'W' where we could substitute the very similar Greek letter 'ω'. The relative frequency of the roman letter "W" is 2.360% (Lewand, 2000) and the type frequency of the Greek letter "ω" in GreekLex (Ktori *et al.*, 2008) is 9.534, therefore both roughly fall in the low-to-medium range.

In all of the cases where we chose two different words to test in the two lists, we tried to match the familiarity of each of the Greek words that we chose to that of the target English word, using relative word frequency in the Kuçera–Francis corpus for English (Kuçera and Francis 1967) and the ILSP database for Greek (Gavrilidou *et al.* 1999). That is, we used frequency to stand in for any more direct measure of familiarity, since we had familiarity ratings for only some of the English words (Pisoni *et al.* 1985) and had no familiarity ratings at all for the Greek words.

For subsequent coding purposes, target items for the picture arrays were tagged as belonging to one of three distinct types: (1) identical items ('I') for picture arrays where the same Greek word was used in both lists; (2) synonymous items ('S') for arrays where two different synonymous translations were used for the same English target (for example, the words *αμάξι* /a'maksi/ and *αυτοκίνητο* /afto'çinito/ for the target *car*); and (3) different items ('D') for arrays where the word on one list named the target picture and the other word named a foil picture (for example, *έκρηξη* /'ekriksi/ 'eruption' named the target picture and the other Greek word *αστραπή* /astra'pi/ 'lightning' named a foil picture). The two lists contained 76 pairs with identical items, 45 pairs with synonym items and 49 pairs with items from different pictures.

Because Greek has a much richer inflectional morphology than English, one final methodological issue that needed to be addressed was the choice of the morphological shape of each target item. Candidate items that are nouns were presented in the nominative case and candidate items that are verbs were presented in the third person singular, as this was the closest match

Adapting a receptive vocabulary test

to the reduced present progressive form (for example, 'show me *barking*') used in the English ROWPVT. Because the word for child is neuter, most adjectives could be presented in the neuter singular form, to avoid reducing the response set. For example, the pictures for English *happy* are a smiling boy (for the target) and a scowling girl, a fearful girl, and an angry boy (for the foils). Choosing the masculine form for *happy* would have reduced this from a four-alternative forced-choice to a two-alternative forced-choice response. The one exception was the array for the English target *parallel*, where we chose the feminine plural forms *παράλληλος* /pa'raliles/ for the translation equivalent and *κάθετος* /'kaθetes/ 'vertical' to name a foil picture. (All four pictures in the array showed arrangements of two lines, and *γραμμή*/gra'mi/ 'line' is feminine.)

Subjects

Participants were 427 children, 225 boys and 202 girls, ranging in age from 2 years to 5 years. Table 1 shows the distribution of age groups, gender, and list assignment for the participants. We have further subdivided the 2- and 3-year-olds into 'younger' and 'older' age groups. We made a particular effort to ensure that we had represented the entire age range for these two youngest groups of children. (Note that this was the opposite sampling strategy to that used in the ROWPVT-II norming study, for which fewer 2- and 3-year-olds were tested and twice as many 4- and 5-year-olds were tested).

Children were recruited from urban and suburban areas of Northern (Salonika, N = 336), Western (Ioannina, N = 59), and Southern (Crete, N = 32) Greece. All children—with the exception of two 2-year-olds who were tested in their homes—were attending private or public preschools. For each child, the parent or teacher completed a questionnaire regarding parents' occupation, age, educational level, language environment at home, and the child's hearing and communication status. Based on the response to this questionnaire, four children were excluded from participating in this study because they came from bilingual families. That

Table 1. Age groups, gender distribution within age groups, and distribution of children between the two lists

Age group	Range (years;months)	List A	List B	Boys	Girls
Younger 2s	2;0–2;5	13	15	7	21
Older 2s	2;6–2;11	25	19	21	23
Younger 3s	3;0–3;5	37	46	43	40
Older 3s	3;6–3;11	47	36	42	41
4-year-olds	4;0–4;11	68	66	84	50
5-year-olds	5;0–5;11	32	23	28	27

is, the remaining 427 participants listed in Table 1 were all from monolingual Greek-speaking homes.

Administering the pilot test

The procedures for administering the original English ROWPVT were followed. The prompting phrase *Show me* — was translated directly, as *Δείξε μου* — /'ðikse mu —/, and the tester said the target word embedded in this phrase to the child, who was asked to point to the corresponding picture in the array of four.

The testers were four undergraduate students from the University of Macedonia. For each child, the tester pseudo-randomly presented either list A or list B with the aim of getting an equal number of list A and list B response sets for each age group, and equal numbers of boys and girls in each list (Table 1).

The testers followed the standard administration procedures for ROWPVT-II. For each child, it was necessary to obtain a basal group of eight consecutive correct responses. Presentation started at variable places in the lists, depending on the child's age. If the first eight responses were correct, this starting point was the basal. Otherwise, the earlier and earlier blocks were tested until the basal criterion was achieved or the beginning of the list was reached. The basal item ranged from item number 1 (for all the 2-year-olds and 22 of the 3- and 4-year-olds) to item number 35 (for 25 of the 5-year-olds). In addition, if the basal was the starting item, then it was necessary to establish the ceiling. In this case, the tester continued presenting items in subsequent blocks until each child reached a ceiling, defined as six incorrect results in any block of eight items. This ceiling ranged from item number 8 (for a 2-year-old) to item number 138 (for a 5-year-old). The tester recorded each child's response to each item presented by entering the number of the picture they pointed to on the corresponding row of an individual response form and noting whether the response was correct or incorrect.

Data tabulation

For each child in each list group, we made a score sheet in which we entered the child's response to each item on the list as either correct or incorrect. All items below the basal item for the child were scored as correct responses and all items above the ceiling were scored as incorrect.

Results

Raw scores

The first analysis was a general evaluation of the test items as an age-graded list of Greek words. For this

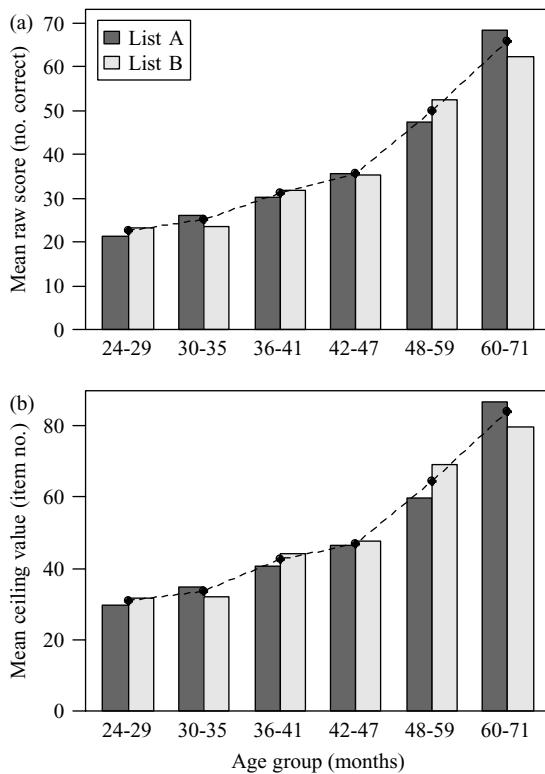


Figure 1. Mean raw score averaged across the children in each age group. The dots and line in each panel track the means averaged over all the children; the bars plot means averaged separately for the children who were administered the two different lists.

purpose, we compared average raw scores across the age groups (Figure 1a). As shown, there was a monotonic increase in the raw scores across the age groups, both overall and for each list separately. Also, the difference between the two lists was generally smaller than the increase across adjacent age groups, except for the two youngest groups, where the differences between list A and list B were larger than the increase from the younger 2-year-olds to the older 2-year-olds in the list B group. (Figure 1b shows an alternative measure, the mean ceiling item reached, averaged across the age groups in the same way. This measure shows exactly the same trends as the average raw score.)

Comparing lists A and B

The next analyses focused more closely on determining whether the children's response to items on lists A and B yielded similar accuracy rates and a similar progression of increasing difficulty, as gauged by relative accuracy, from the beginning to the end of the list. For these analyses, we calculated the weighted proportion of correct responses for each item separately for each list by taking the average of the proportions of correct responses in each of the six age groups. We will call this measure the 'accuracy score'

for that word in that list. Item pairs above 138 are not included in this analysis because none of the children in either list-group responded correctly to these more difficult items. Figure 2a plots the accuracy scores as a function of the item number using black for the list A words and grey for the list B words. The two lines track the accuracy scores for the 76 words that were identical between the two lists, and the dots show the proportion of correct for the 94 items where we used different words for the two lists—with small dots for the 45 items where the two words were synonyms for the same target picture and large dots for the 49 items where one of the two words named the target picture and the other named a foil. It can be observed that the accuracy scores for the identical items are very similar across the two lists. Furthermore, many of the pairs of different words (both synonyms and names of target versus foil pictures) also have fairly similar accuracy scores.

We performed a regression to evaluate this relationship between the percent correct responses by children in each of the two list groups and we found that there was a highly significant relationship between the accuracy scores for the items on the two lists ($R^2 = 0.97$; $p < 0.0001$). Moreover, the coefficients of this regression function were 0.01 for the intercept and 0.96 for the slope, which are very close to the values zero and 1 that would be returned if the two proportions were exactly identical between the paired items on the two lists.

Figure 2b plots the residuals from this regression, with the thick solid line tracking the residuals for the items where the two lists had the same word and the small black and large grey dots showing the two types of items where the words differed across the two lists (Figure 2a). The dashed lines show the maximum difference of 0.06 (that is, 6%) that was obtained for the items where the two lists had identical words. As the distribution of dots shows, most residuals for items where different words were paired fell well within this maximum difference for the accuracy scores for words that were shared between the two lists. We identified 13 word pairs as outliers on this analysis because their accuracy scores on the two lists differed by more than the maximum difference obtained between pairs of identical words. That is, since we plan to use the accuracy rate for each word as a measure of the word's relative difficulty, the difference in accuracy for item pairs that tested the same word on the two lists is a gauge of the measurement error, and these 13 outliers are pairs of words with a reliably large difference in relative difficulty. List numbers are shown for these 13 outliers: four of these paired two different synonyms for the same picture (αντίχειρας /a'diçiras/ 'thumb' δάκτυλο /'ðaktilo/ 'finger' for the picture of *thumb*, στρογγυλό /stro'ji'lo/ 'round' οβάλ /o'val/ 'oval' for the picture *round*, στιβάδα /sti'vaða/ 'stack' στοίβα /'stiva/ 'stack' for

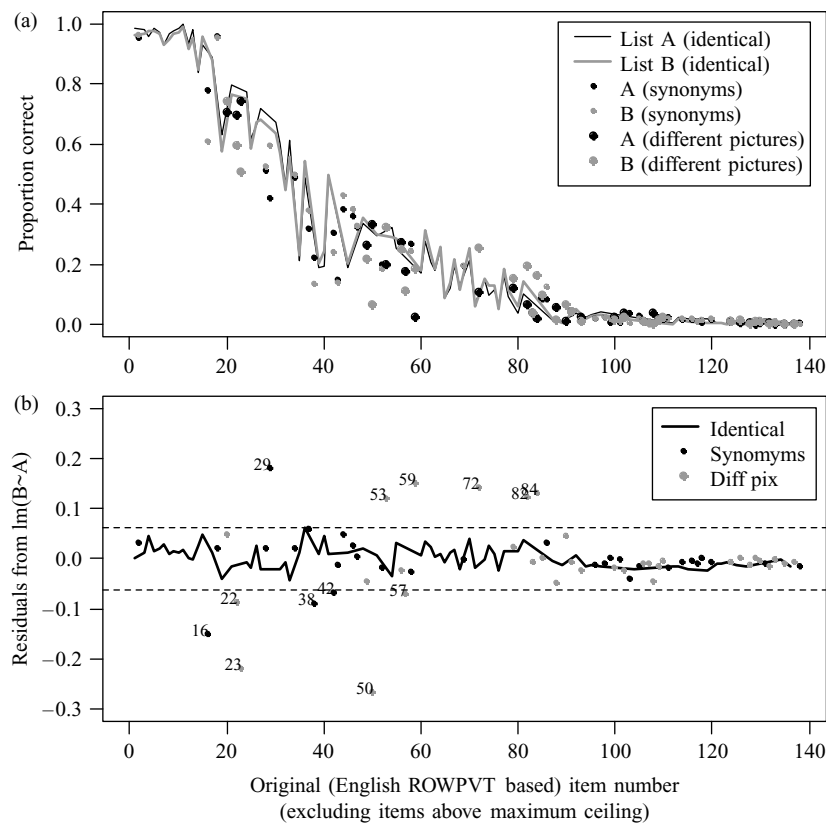


Figure 2. (top) Proportion of correct responses over all six age groups for the three different types of 138 items in the two lists that were identified correctly by at least one child; (bottom) residuals from regressing the proportion of correct responses by children who were assigned to list B against the proportion of correct responses by children who were assigned to list A. The dashed lines demarcate the maximum residual for identical items.

the picture of *stack* and ρίχνει /'rixni/ 'dropping, throwing' πετάει /pe'tai/ 'throwing' for the picture of *pitching*) and nine paired names of two different pictures in the four-picture array.

Figure 3 shows the other comparison that we made to evaluate the differences in the children's performance between the two list groups. Using the accuracy scores as our measure of relative difficulty, we ranked the words on each list in decreasing order by their accuracy rates, keeping the original order in the case of ties. For the identical items, we then regressed the ranks obtained for words in list B against the ranks obtained for words in list A. This relationship was very strong ($R^2 = 0.98$), particularly for words in the first half of the list, where the accuracy rates for a word on the two different lists fall within ten places of each other. Above about item number 85 (the median ceiling for the oldest age group), the ranks begin to fan out away from the $x = y$ line. Here the relative accuracy score is less reliable, since it is based on the responses of only a small number of participants. Data points for the majority of other items show the same pattern. Except for the 13 outlier pairs, most word pairs had similar accuracy ranks between the two lists

in the region where accuracy rates for the identical pairs were consistent between the two groups.

Reordering the list and choosing among item pairs

The next analyses focused on using the item order effect to determine how to reorder the list items and how to choose between those words of the lists that matched synonyms or different picture names. For these analyses, we first calculated a combined accuracy score by averaging across the two lists, excluding the 13 outlier-pairs. Figure 4 plots this combined accuracy score for each item pair as a function of the original (English ROWPVT-II based) item number on the list, separately by age group. It can be observed that there are effects both of age (the older children generally have higher accuracy scores than the younger children) and of item number (higher-numbered items generally have lower accuracy scores than lower-numbered items). Also, the effect of age is consistent across the list, but the effect of item number is not; some items have smaller accuracy scores than would be predicted from their order in the list. In particular, there are substantial deviations from

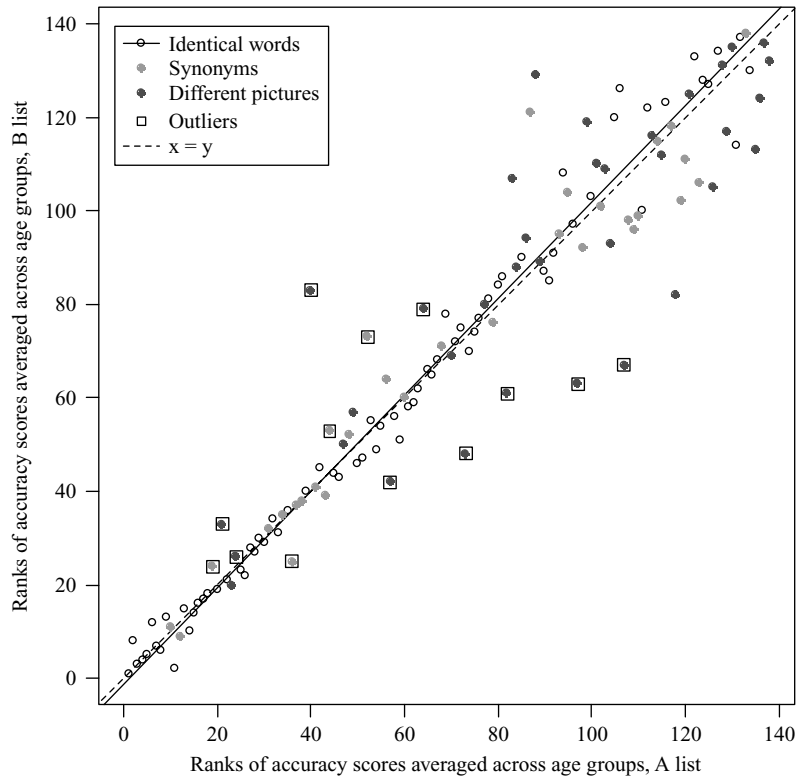


Figure 3. Relative difficulty (ranked accuracy score) of words on list B plotted against relative difficulty of words on list A. The solid line is a regression curve for words that occurred on both lists. The dashed line compares what the line would look like if difficulty for the two lists was exactly equal. Squares pick out the 13 outliers identified in Figure 2b.

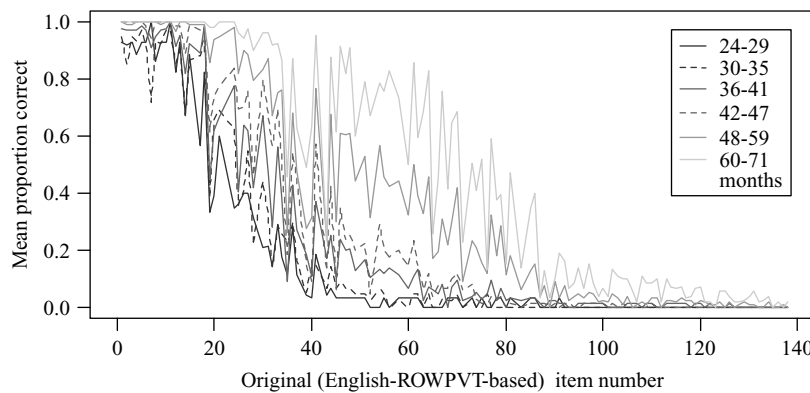


Figure 4. Mean accuracy score by age group as a function of item number, excluding the 13 items where the difference between list A and list B was larger than the largest difference for identical items.

the general trend that seem to be fairly consistent across the age groups between items 35 and 45.

The relationships among the different lines in the right half of Figure 4 also support our interpretation of the spread of points in this region of figure 4. Each line in the graph asymptotes to zero after the median ceiling for the age group, and differences between items are minimized and become completely unreliable. In the regions below these asymptote points, by contrast, the pattern of deviation from monotonically increasing

difficulty, as gauged by the relative accuracy scores across near by items, appears to be consistent across the age groups. Although a word's difficulty as measured by the mean accuracy score necessarily differs across groups (a younger child is less likely to know the word than an older child), the relative difficulty is the same for any word that at least some children in the younger age group know. That is, it appears that the ranks of the words will be consistent across any pair of age groups in regions below the asymptote for the younger group. This

Table 2. Correlations between age groups of ranks of tested items for both age groups

Age groups	Young 2s	Older 2s	Young 3s	Older 3s	4s
Older 2s	0.97				
Young 3s	0.96	0.96			
Older 3s	0.97	0.96	0.98		
4-year-olds	0.94	0.92	0.98	0.98	
5-year-olds	0.91	0.88	0.96	0.95	0.99

appearance is substantiated in Table 2, which gives the correlations between rankings derived from the accuracy rates for each pair of age groups, calculated only over items below the smaller of the maximum ceiling values.

Based on these analyses of the average accuracy scores across the original list order, and on the result showing consistency in ranking among the different age groups, we felt confident in using the ranks determined by the accuracy scores averaged across the six age groups to reorder the items. For the 72 items that tested the same word, the basis score for reordering was the mean accuracy score averaged between the two lists. For the other 66 items that tested different words, we compared the ordinal position of each of the two words in the ranking determined by the relative accuracy scores for

the list to which it belonged to the ordinal position of that item in the original list, and took the word that was positioned closer to its original position in the list. For 42 of these 66 items, the word was taken from list A, and for 24 of these items the word was taken from list B. The basis score for these words was the mean accuracy score just for the children who were tested with that word. We then determined a complete order among all 138 items based on the ordinal position of the basis scores in a ranking from highest accuracy to lowest, with tied items keeping the original relative order. Figure 5 plots the basis accuracy score against the item number in the reordered list. A comparison of figures 4 and 5b shows that there is much less deviation from the monotonic downward trend once the list has been reordered.

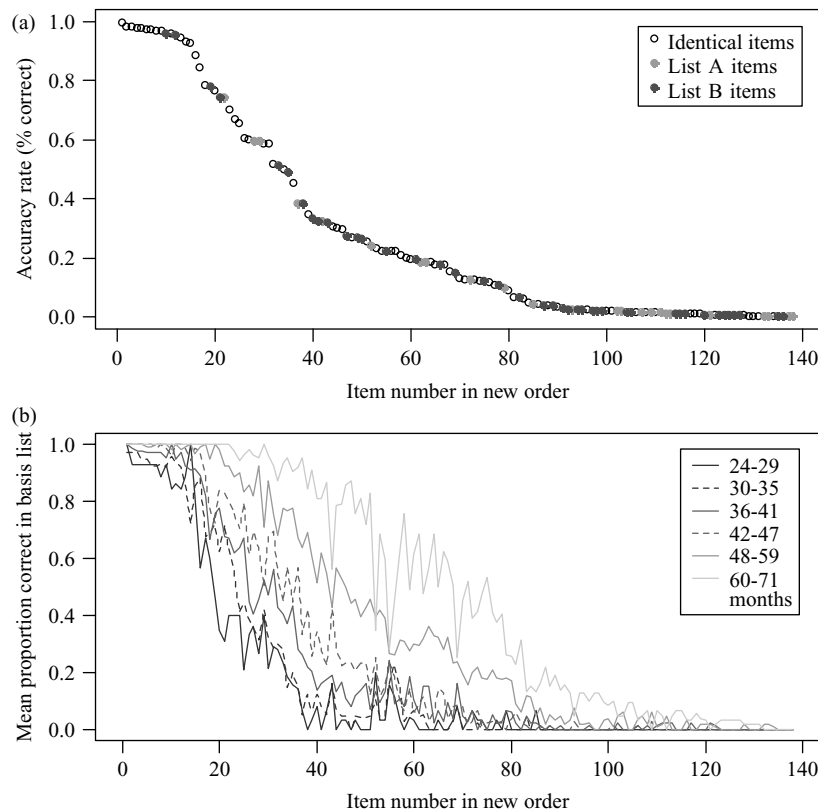


Figure 5. Mean proportion of correct responses as a function of the new order, averaged over all the children for the 77 items that tested the same word twice and averaged over the children who responded in list A or in list B, as appropriate, for the 66 items that tested different words in the two lists averaged across all age groups (top) and separately by age group (bottom).

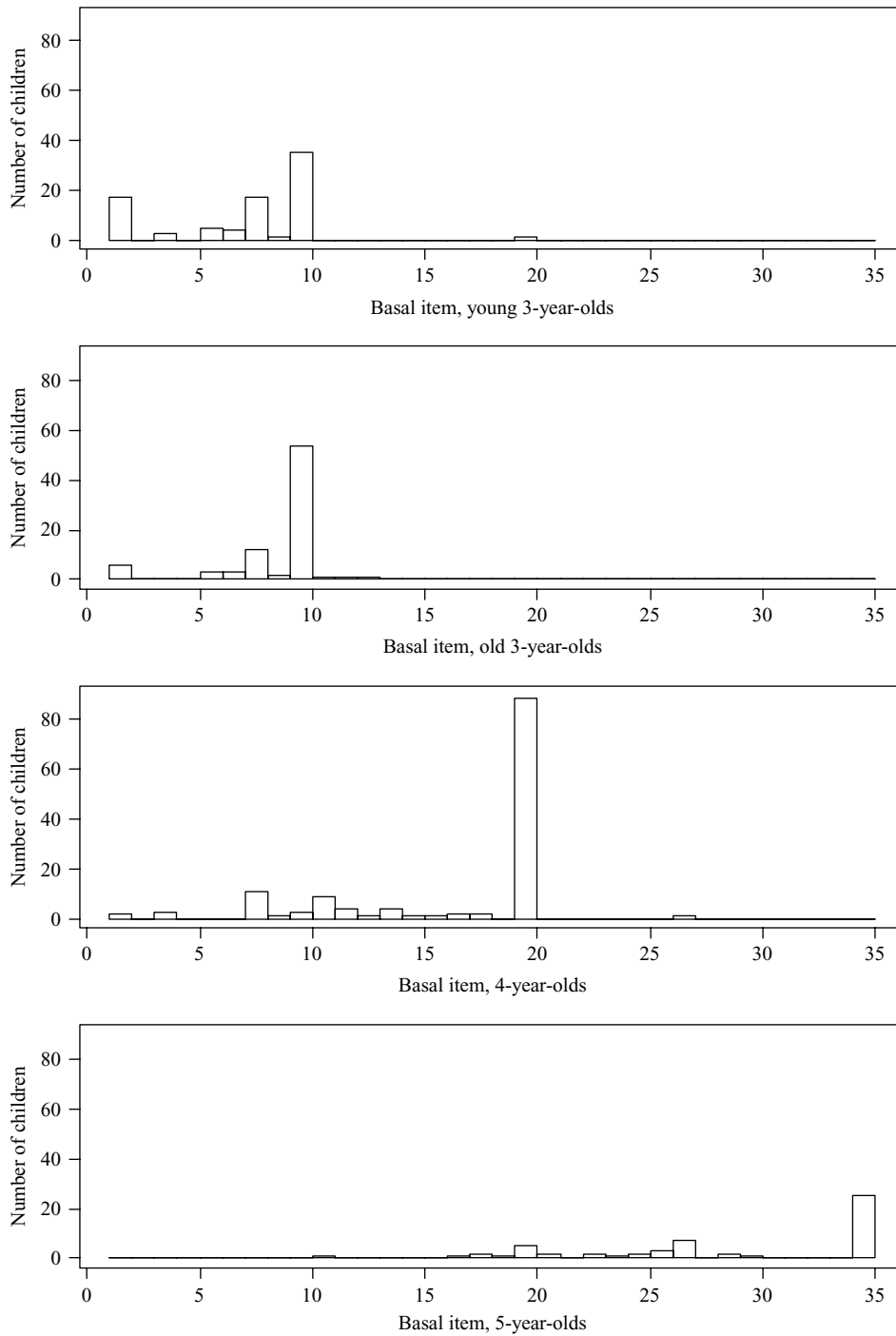


Figure 6. Number of children who obtained a basal score at different item numbers of the original list, by age group.

Comparing basal values across age groups

Finally, we determined new starting levels for the word list. Figure 6 shows histograms by age group for the item number of the starting point for the American English version of the ROWPVT-II (the target word at which the tester began testing and attempted to obtain eight consecutive correct responses). The peak in each plot is

the default starting point for that age group. The bars to the left are for those cases where the examiners had to test earlier items because the child did not achieve a basal score within that test block.

It can be observed that, in order to obtain a basal score, the testers had to test earlier-ranked items for a majority of the younger 3-year-olds and nearly one-third of the older 3-year-olds also required this kind of

Adapting a receptive vocabulary test

back-tracking. For 20 of these children, the testers needed to begin with the original item 1 (*παπούτσι*/pa'putsi/ *shoe*), which would be item number 4 on the reordered word list. Furthermore, the testers had to test earlier-ranked items for nearly one-third of the 4-year-olds (44 out of 134) to obtain a basal score, but only five of these children needed to go back further than the original item number 8, *καρότο*/ka'roto/ *carrot*, which would be item number 10 on the reordered list. Finally, it can be observed that the testers needed to test earlier-ranked items for more than half of the 5-year-olds (30 out of 55), but only five of these children needed to be tested on earlier-ranked items than number 20 *σκίουρος* /'sciuros/ *squirrel*, which would be item number 20 on the reordered list. Hence, the following starting points in administering the adapted version of ROWPVT to monolingual Greek-speaking preschool children are proposed: (one) 2;0–3;11 years, begin at item number 1; (two) 4;0–4;11 years, begin at item number 10; (three) 5;0–5;11 years, begin at item number 20 on the reordered list.

Discussion

The purpose of this study was to adapt an English receptive vocabulary test, the ROWPVT-II, as a basis for developing a receptive vocabulary test for Greek preschool-aged children. There is no receptive vocabulary test for this population currently available in Greece today. In fact, even for school-aged children, the only receptive vocabulary tests that are available are short (20–30-item) subtests embedded in larger tests of intelligence or academic achievement (Georgas *et al.* 1997, Paraskevopoulos *et al.* 1999). Given the importance of early identification and remediation of language disorders, we thought it was important to develop a reliable and valid language test for this group of children. We chose to adapt a receptive vocabulary test because this measure is correlated with IQ and later academic achievement, as well as being indicative of language disorder.

In making this adaptation, we found—as have others—that direct translation could not accomplish the purpose of developing an age-graded list of target words. In some cases, a concept in English (such as *pitching*) was not familiar to Greek-speaking children, while in other cases, a highly familiar word in English (such as *thumb*) was not familiar to Greek children of the same age. Of course, there are mismatches in the other direction as well (that is, Greek-speaking children know concepts and words that English-speaking children do not), and it is possible that we could have found an even better set of Greek words if we had started from scratch rather than adapting the English test.

There are other difficulties with adapting a receptive vocabulary test into another language that we ignored. For example, a hallmark characteristic of Greek is its rich inflectional morphology, such that gender, number, person, verb tense, and case are specified by suffixes on the base form. Surely, this rich morphology helps children to learn and recognize words in context. However, for the purposes of developing this receptive vocabulary test, we chose to use morphological forms, such as the neuter gender, that would preserve an equal probability of all pictures in each array. Again, it is possible that we could have obtained a better word list if we had had the resources to develop picture arrays for sets of target words and foils from scratch. However, the Greek wordlist that we devised by making only one change to any picture array did give relatively high correlations between rankings across age groups and an overall decline in accuracy over the list items tested even before reordering the items.

The results of the pilot experiment were used to choose a final list of 170 words and to order these words based on how many correct responses each item received. This final list appears to be a valid receptive vocabulary test for preschool-aged Greek-speaking children from Greece.

There are two ways to evaluate the validity of receptive vocabulary tests. One is to correlate the results of several different vocabulary tests (for example, Brownell 2000, and Dunn *et al.* 2006). A high correlation between the new test and previous tests indicates that the new test is valid. Unfortunately, we are unable to do this, as there are no other vocabulary tests (or indeed any other language tests) for preschool-aged Greek-speaking children.

A second way to evaluate validity of vocabulary tests is to consider whether the raw scores increase with age, as would be expected, given that older children generally comprehend more words than younger children. Raw scores do increase with age. The correlation between age and raw score is 43% across all of the children in this study, and it is 56% for the children who were tested with the words in list A, which was the source list for two-thirds of the words chosen from the 66 pairs of words that differed between the two lists. We expect that this correlation between age and raw score will be even stronger on the new revised version of the test that we have developed, by reordering the items after choosing the words from the pairs that were not shared between the lists.

The next step toward developing a norm-referenced receptive vocabulary test for Greek-speaking children is to test a large number of 2-through 5-year-olds on the revised word list that we have developed. In this next study, we plan to test a larger and more representative sample of Greek children, taken from more diverse

geographical regions of Greece (suburban and rural as well as urban) and with a wider range of socio-economic status.

Acknowledgements

The authors would like to thank Dimitris Vassilodimitrakis, Virginia Papageorgiou, Savvas Karamixalakis, and Iliana Panou for their assistance with data collection. They thank the children who participated in the task; the parents who gave their consent; and the preschools at which the data were collected. **Declaration of interest:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Note

1. Since direct translation can sometimes be more valid than the adapted version (Ali 1967), direct translation was retained whenever possible.

References

- ALI, M. B., 1967, To establish the feasibility of using translated and adapted versions of an American made mathematics achievement test in East Pakistan. *Dissertation Abstracts*, **28** (1-A), 115–116.
- BORNSTEIN, M. H. and HAYNES, O. M., 1998, Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development*, **69**, 654–671.
- BROWNELL, R., 1985, *Receptive One Word Picture Vocabulary Test—Spanish-Bilingual Edition* (New York, NY: Academic Therapy).
- BROWNELL, R., 2000, *Receptive One Word Picture Vocabulary Test—II* (New York, NY: Academic Therapy).
- CHANNELL, R. W. and PEEK, M. S., 1989, Four measures of vocabulary ability compared in older preschool children. *Language, Speech, and Hearing Services in Schools*, **20**, 407–419.
- CHEUNG, P. S., LEE, Y. S. and LEE, K. Y., 1997, The development of the Cantonese Receptive Vocabulary Test for children aged 2–6 in Hong Kong. *European Journal of Disorders of Communication*, **32**, 127–138.
- DUNN, L. M. and DUNN, L. M., 1981, *Peabody Picture Vocabulary Test—Revised* (Circle Pines, MN: American Guidance Service).
- DUNN, L. M., DUNN, L. M. and DUNN, D. M., 2006, *Peabody Picture Vocabulary Test, Fourth Edition. Examiner's Manual and Norms Booklet* (Circle Pines, MN: American Guidance Service).
- DUNN, L. M., DUNN, L. M., WHETTON, C. and BURLEY, J., 1997, *British Picture Vocabulary Scale*, 2nd edn (Windsor: NFER-Nelson).
- EHRI, L. C. and SNOWLING, M. J., 2005, Developmental variation in word recognition. In C. A. Stone, E. R. Silliman, B. J. Ehren and K. Apel (eds), *Handbook of Language and Literacy: Development and Disorders* (New York, NY: Guilford), pp. 433–460.
- FENSON, L., DALE, P. S., REZNICK, J. S., THAL, D., BATES, E., HARTUNG, J. P., PETHICK, S. and REILY, J. S., 1993, *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual* (Baltimore, MD: Paul H. Brookes).
- GARDNER, M. F., 1979, *Expressive One-Word Picture Vocabulary Test* (Novato, CA: Academic Therapy).
- GARDNER, M. F., 1985, *Receptive One-Word Picture Vocabulary Test* (Novato, CA: Academic Therapy).
- GAVRILIDOU, M., LABROPOULOU, P., MANTZARI, E. and ROUSSOU, S., 1999, Prodiagrafes gia ena ipologistiko morphologiko lexiko tis Neas Ellinikis [Specifications for a computational morphological lexicon of Modern Greek]. In A. Mozer (ed.), *Proceedings of the Third International Conference on the Greek Language* (Athens: Ellinika Grammata), pp. 929–936.
- GEORGAS, D., PARASKEVOPOULOS, I., BESEVEGIS, I. and GIANNITSAS, N. D., 1997, *Greek WISC-III* (Athens: Ellinika Grammata).
- GRIMM, H. and DOIL, H., 2000, *Elternfragebogen für die Früherkennung von Risikokindern. ELFRA-1: Elternfragebogen für einjährige Kinder: Sprache, Gesten, Feinmotorik. ELFRA-2: Elternfragebogen für zweijährige Kinder: Sprache und Kommunikation* (Goettingen: Hogrefe).
- HAMILTON, A., PLUNKETT, K. and SCHAFER, G., 2000, Infant vocabulary development assessed with a British Communicative Development Inventory: lower scores in the UK than the USA. *Journal of Child Language*, **27**, 689–705.
- HAMMILL, D. D. and NEWCOMER, P. L., 1982, *The Test of Language Development—Primary* (Austin, TX: Empiric).
- HYMES, D., 1970, Linguistic aspects of comparative political research. In R. Holt, J. Turner (eds), *The Methodology of Comparative Research* (New York, NY: Free Press), pp. 295–341.
- JACKSON-MALDONADO, D., THAL, D., MARCHMAN, V., NEWTON, T., FENSON, L. and CONBOY, B., 2003, *MacArthur Inventarios del Desarrollo de Habilidades Comunicativas. User's Guide and Technical Manual* (Baltimore, MD: Brookes).
- KERN, S. and LANGUE, J., 2000, Des premiers gestes aux premiers mots: le développement communicatif chez l'enfant de 8 à 30 mois. In *Actes de 3èmes Journées Scientifiques de l'Ecole d'orthophonie de Lyon 24 et 25 novembre* (Lyon: ENS).
- KIM, Y., CHANG, H., LIM, S. and BAK, H., 1995, *Geurim eohyuryeok geomsa* [Picture Vocabulary Test] (Seoul: Seoul Community Rehabilitation Center).
- M., Ktori, VAN HEUVEN, W. J. B. and PITCHFORD, N. J., 2008, GreekLex: A lexical database of Modern Greek. *Behavior Research Methods*, **40** (3), 773–783.
- KUČERA, H. and FRANCIS, W. N., 1967, *Computational Analysis Present-Day American English* (Providence, RI: Brown University Press).
- LEWAND, R. E., 2000, *Cryptological Mathematics*. The Mathematical Association of America, p. 36.
- MAITAL, S. L., DROMI, E., SAGI, A. and BORNSTEIN, M. H., 2000, The Hebrew Communicative Development Inventory: language specific properties and cross-linguistic generalizations. *Journal of Child Language*, **27**, 43–67.
- METSALA, J. L., 1999, Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, **91**, 3–19.
- NATION, I. S. P., 2001, *Learning Vocabulary in Another Language* (Cambridge: Cambridge University Press).
- NATION, K. and SNOWLING, M. J., 1997, Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, **67**, 359–370.
- OGURA, T., YAMASHITA, Y., MURASE, T. and DALE, P. S., 1993, Some findings from the Japanese Early Communicative Development Inventories. *Memoirs of the Faculty of Education, Shimane University*, **27**, 26–38.
- PARASKEVOPOULOS, I., KALATZI-AZIZI, A. and GIANNITSAS, N. D., 1999, *Athena Test of Diagnosis of Learning Difficulties* (Athens: Ellinika Grammata).
- PEÑA, E. D., 2007, Lost in translation: methodological considerations in cross-cultural research. *Child Development*, **78**, 1255–1264.

- PISONI, D. B., NUSBAUM, H., LUCE, P. A. and SLOWIACZEK, L., 1985, Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, **4**, 75–95.
- RENZULLI, J. S. and PAULUS, D. H., 1969, A cross-validation study of the item ordering of the Peabody Picture Vocabulary Test. *Journal of Educational Measurement*, **6**, 15–20.
- ROCA, P., 1955, *Problems of Adapting Intelligence Scales from One Culture to Another* (Haltá Rey, Puerto Rico: Oficina de Evaluación, Departamento de Instrucción Pública).
- SIMON, A. J. and JOINER, L. M., 1976, A Mexican version of the Peabody Picture Vocabulary Test. *Journal of Educational Measurement*, **13**, 137–143.
- TAMAYO, J., 1987, Frequency of use as a measure of word difficulty in bilingual vocabulary test construction and translation. *Educational and Psychological Measurement*, **47**, 893–902.
- THERIAULT-WHALEN, C. M., DUNN, L. M. and DUNN, L. M., 1993, *Echelle de vocabulaire en images Peabody* (Richmond Hill, ON: Psyscan Corporation).
- THORNDIKE, R. L., 1973, Reading as reasoning. *Reading Research Quarterly*, **9** (2), 135–147.
- UENO, K., UTSUO, T. and IINAGA, K., 1991, *PVT kaiga goi hattatsu kensa* [PVT Picture Vocabulary Development Test] (Tokyo: Chiba Test Center).
- VELLUTINO, F. R., FLETCHER, J., SNOWLING, M. J. and SCANLON, D., 2004, Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, **45**, 2–40.
- WIENER, F. D., SIMMOND, A. J. and WEISS, F. L., 1978, *Prueba Ilustrada de vocabulario Espanol* [Spanish Picture Vocabulary Test] (New York, NY: Marymount Manhattan College).