Measurement of Expressive Vocabulary in Multilingual School-Age Children using the Dual-Focus Approach Method: Development of the Lebanese Picture Naming Test (LPNT)

Rama Kanj, MA

American University of Beirut, Department of Education

American University of Beirut, Department of Psychiatry

Karma El-Hassan, PhD[1]

American University of Beirut, Department of Education

American University of Beirut, Office of Institutional Research & Assessment

[1] Corresponding Author: Karma El-Hassan, PhD, Associate Professor of Education, Office of Institutional Research & Assessment (OIRA), American University of Beirut, Bliss St., P.O. Box 11-0236, Beirut 1107 2020, Lebanon E-mail: kelhasan@aub.edu.lb

Author Declaration

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

- Dr. Karma El-Hassan

- Ms. Rama Kanj

Measurement of Expressive Vocabulary in Multilingual School-Age Children using the Dual-Focus Approach Method: Development of the Lebanese Picture Naming Test (LPNT)

Naming is a measure of expressive vocabulary that is fundamental for communication through language. In children, it is the earliest step in linguistic production (Etard et al., 2000). Naming is usually measured through a task of *picture naming* (also called *confrontation naming*) where an individual is shown a picture of an object, an action or a concept and is asked to provide the vocabulary word that accurately corresponds to the picture. Vocabulary acquisition is found to be associated with both internal (biological) and external (environmental) factors (Apiwattanalunggarn & Luster, 2005; Hamadani et al., 2010; Hoff, 2003). Studies show that some biological factors such as older age contribute to higher performance on vocabulary tests (Ilkman, 2015; Spinelli et al., 2005; Vogt, Douglas, & Aussems, 2015), while others such as gender effect on performance seem to be inconsistent (Bornstein, Hahn, & Haynes, 2004; Grabowski, Damasio, Eichhorn, & Tranel, 2003). Environmental factors such as higher socio-economic status (SES) and residence in advantageous neighborhoods were similarly found to positively contribute to vocabulary acquisition in children (Dupere, Leventhal, Crosnoe, & Dion, 2010; Hoff, 2003).

In the clinical and academic setting, picture-naming tests are often administered by teachers, speech and language therapists, educational psychologists, special educators and pediatric neuropsychologist amongst other professionals in the field of childhood development. A body of research shows that screening for naming ability in children is important when evaluating children with dyslexia (van Viersen et al., 2017; Lyytinen, Eklund, & Lyytinen, 2004), autism spectrum disorder (ASD) (Luyster, Kadlec, Carter, & Tager-Flusberg, 2008), social and emotional problems (Horwitz et al., 2003; Tervo, 2007) and cognitive impairments (Oliver, Dale & Plomin, 2004). It also predicts reading and

spelling abilities in children 24 months after traumatic brain injuries (TBI) (Catrroppaa & Anderson, 2004).

With multilingualism in children increasing over the years and the amount of research in this area almost tripling between 1997 and 2005 (Bialystok, 2007), developing cognitive assessment tools requires the consideration of the linguistic make-up of its target population. More research shows that individuals who are bilingual do not function as "two monolinguals in one person" and may perform differently on language assessments than monolinguals (Armon-Lotem, de Jong & Meir, 2015). As a consequence, and in order to preserve the psychometric properties of developed measures, several test construction methods such as the dual-focus approach were developed to account for the multilingual nature of examinees. The dual-focus approach for test development was developed by Erkut, Alarcón, Coll, Tropp, and García in 1999 to allow the development of two or more language versions of a test simultaneously (See Erkut et al., 1999). It aims to create bilingual or multilingual measures where test items are developed in several languages at the same time. In the original study describing the dual-focus approach by Erkut et al. in 1999, the authors describe two features that are specific to this approach. First, test construction involves a team of bilingual and bicultural researchers who are indigenous to the test's target cultures, and second, it is a concept-driven approach rather than a translation-driven approach to attain equivalence. In the present study, the steps involved in test construction were largely based on the dual-focus approach for test development in order to construct items that are culturally and linguistically suitable for our population.

Study Setting: Linguistic Make Up and Type of Schooling in Lebanon

Lebanon's "unique multilingual and multicultural make up" has been a subject of interest to researchers across the years (Bacha & Bahous, 2011; Sayegh, 1965; Shaaban, 1997). Lebanon's official national language is Arabic with the majority of Lebanese

population using Spoken Lebanese Arabic vernacular, part of the Levantine Arabic, as the primary mode of communication in daily conversations. Additionally, most Lebanese people are also bilingual or trilingual. As a consequence of a cultural exchange on Lebanese territories in the 20th century, French and English languages became languages of instruction in Lebanese schools. Children begin to learn a foreign language, French or English and sometimes both, at the Nursery level in private schools and most public schools (Shaaban, 1997). Statistics from the Lebanese Ministry of Education in the year 2018 show that 51.4% of all Lebanese schools offer French as a primary language of instruction and 48.6% offer English as a primary language of instruction (CERD, 2018). French and English language classes are offered on average 8 hours a week at the Elementary level and are also used as the medium of instruction for mathematics, sciences and social studies at all levels. With this being reported, school-age children in Lebanon are all generally bilingual (Arabic-English or Arabic-French) and very often trilingual (Arabic-French-English). An early study describes the common use of loanwords from second languages and translation equivalence sometimes due to filling in words that are difficult to recall in one of the languages (Kotob, 2002) while a more recent one describes that code switching in Lebanese individuals has become more ubiquitous with students switching back and forth between two languages (Bahous, Nabhani & Bacha, 2014). Given the linguistic make-up of Lebanese children and adults, any assessment of language abilities used in a Lebanese context should allow responses in Spoken Lebanese Arabic, French and English to reflect the spoken dialect of Lebanese children to contribute to fair results on test performance.

Another consideration that is particular to the study setting is the type of schooling of Lebanese children. Lebanese schools are referred to as either “public schools” or “private schools” depending on the sector to which they belong. Schools in the public sector are financed by the Ministry of Education and Higher Education (MEHE), whereas schools in the

private sector are generally financed by students' fees. Unfortunately, spending on education in the public sector consistently falls short of spending on education in the private sector (PNUD Report, 2009). Reports show that students enrolled in private schools have a higher success rate in intermediary exams (83.1% compared to 64% in public schools), more qualified teachers and personnel and are less likely to repeat a grade than students enrolled in public schools (PNUD Report, 2009). Reasons for the differences in the quality of education provided in private schools and public schools are also the result of decades of political and sectarian conflicts that are over and beyond the scope of this section despite their implications being drastic on Lebanese students and education till this day. Although there are no reliable reported numbers showing the difference in SES between children enrolled in private schools compared to those enrolled in public schools, it is believed that Lebanon's educational system became to a large extent categorically divided into private sector and public sector, where families belonging to middle to upper SES groups are enrolled in the former and families from lower SES are enrolled in the latter.

Picture-Naming Tests for Lebanese Children

Given the dearth of standardized tests that are developed or adapted to the Lebanese population, practitioners in Lebanon generally continue to rely on picture-naming tests developed and normed in other cultures, mainly of Western and European countries, in order to obtain an objective measure of expressive vocabulary (Simhairi, 2010). As a result, two issues arise from administering a picture-naming test on a population that is different from the population to whom the test is intended. The first issue pertains to the cultural relevance of the test stimuli while the second issue pertains to sample incomparability. To our knowledge, one test, *Evaluation du Langage Oral chez l'enfant Libanais* (ELO-L; Assessment of Oral Language in the Lebanese Child) (Zebib, Henry, Khomsi, Messarra, & Hreich, 2017), is an adaptation of the *Evaluation du Language Oral* by Khomsi (2001), that was recently

developed for Lebanese children between the ages of three and seven years old. The measure includes subtests that assess aspects of oral language including a expressive vocabulary task where children are presented with black and white line drawings. The fact that there exists only one study on a measure of expressive vocabulary for Lebanese children indicates that there is a need to develop tools that are culturally and linguistically suitable for our population. Considering the unique linguistic make-up of Lebanese children, our study aimed to develop, rather than adapt, a test using steps from the dual-focus approach for development of multilingual measures. Although test adaptation could resolve some of the issues related to test bias, some test items are not transferrable across cultures and have no equivalent in another language (Peña, 2007). Moreover, we aimed to compile colored pictures that cover a wide range of semantic categories with known psycholinguistic variables in order to increase the construct validity of the measure. Following test development, we provided preliminary evidence of the measure's validity and reliability and explored variability in language acquisition across age, gender and type of schooling. We referred to this measure as the Lebanese Picture Naming Test (LPNT).

Method

The current study consisted of two phases: the development of the LPNT and data collection. The study in its entirety was conducted under the approval of the Institutional Board of Research (IRB) of the authors affiliated university in Beirut. Ethical considerations also included obtained permission to visit public schools in Beirut from the Lebanese MEHE, in addition to obtained parental consent, school principals' consent, and student oral assent prior data collection.

Participants

The total sample consisted of 74 Lebanese males and females between the ages of 3 to 9 years, enrolled in 3 regular private schools and 2 regular public schools in Beirut (Table 1).

The schools were selected randomly from a list of schools found on the MEHE's website and the ratio of students from private schools to students from public schools was representative of the student distribution across types of schools in Lebanon. According to their homeroom teachers, all students were at least functionally bilingual, meaning their first language was Lebanese Arabic and their language of instruction at school, was either French or English. 42% of the student received classroom instructions across all subjects in English (e.g. math, science, social studies) whereas 58% received classroom instructions in French. Teachers assisted the researcher in selecting students that were easy to reach during break time or before the start of the school day. Nonetheless, efforts were made to select equal proportions of males to females, and children from all ages between 3 and 9 years old. According to teacher and administrators' report, none of the students had a pre-existing diagnosis of neurodevelopmental disorders, intellectual disabilities, learning disability and physical disability affecting sensory modalities and none were receiving special education support, speech and language therapy or any other educational support in or outside school. Parental consent and child oral assent were received from each participant prior test administration. One student (75th participant) interrupted test administration after item 15 and was excluded from the sample.

<Insert Table 1>

Procedures

Phase 1: Development of the Lebanese Picture Naming Test Using Steps from the Dual-Focus Approach. The current study's design was inspired at large by the dual-focus approach for test development to allow the development of test items in multiple languages simultaneously. Some of the features adopted from the dual-focus approach are the collaboration of a research team of bilingual and trilingual experts on the content of the test and the creation of test items in different languages simultaneously.

Step 1: Selecting the Expert Committee Members. Eight experts in fields relevant to early childhood education, language and speech development, teaching or special education, test construction or neuropsychological assessments who have all had direct experience working with Lebanese children between the ages of 3 to 9 years old formed the committee of experts (Table 2).

<Insert Table 2>

Step 2: Materials Development. Picture items were selected from the MultiPic Databank (Dunabeita et al., 2017), a recently published picture set of 750 digitally developed, colored line drawings, covering a wide array of semantic categories (animals, food, clothing, natural elements, furniture, vehicles, etc.) (figure 1). Each picture in the picture set was already assigned a name in English by the authors of the MultiPic Databank. The original number of pictures in the MultiPic Databank (N=750) had to undergo a preliminary reduction to match the number of test items in other widely used picture-naming tests, which typically ranges between 50 to 230 pictures (e.g. Boston Naming Test, Expressive One Word Picture Vocabulary Test-4, and Peabody Picture Vocabulary Test-5). To reduce the number of pictures, we decided to adopt a random selection procedure by selecting every third picture in a randomly ordered list in order to avoid introducing selection bias by the researcher. The method of random selection was carried out twice until the pictures were reduced to almost a little less than a third of the original pool's size and reached an arbitrary number of N = 219. Nonetheless, we checked that the 219 pictures covered diverse semantic categories of similar proportions spanning nouns of animals, natural elements, fruits, furniture, instruments and desk items to name a few (Table 3).

<Insert Figure 1>

< Insert Table 3>

Step 3: Picture Rating and Item Pool Development. The purpose of the rating process was twofold: first, assign target names to the pictures in Lebanese Spoken Arabic, French and English, and second, rate the pictures on the following psycholinguistic variables: *age of acquisition* (AoA) (the age at which individuals usually learns a given word expressively), *name agreement* (the extent to which individuals agree on a single name to refer to a picture), *word frequency* (the frequency with which a word occurs in a given language across modes of communication) and *cultural familiarity* (the extent to which we come in contact with or think about the concept in our everyday life in our culture). The experts' ratings on cultural familiarity and name agreement will be used in the picture selection process, while the measures of word frequency and age of acquisition will be used in arranging the test items in a logical order. The researcher met with the experts individually and provided instructions on the rating process and a description of the psycholinguistic variables. We provided the experts with booklets containing the 219 pictures (sized 10 cm x 10 cm) listed in random order accompanied by Likert-scale ratings for each of the psycholinguistic variables and age ranges between 3 and 9 (6 months ranges for ages below 5 and 1 year ranges for ages between 5 and 9 to match the rate of vocabulary acquisition across development). The experts were also asked to provide any suggestion or remark regarding the quality and clarity of the pictures in order to modify, in future revisions, any test item that may result in incorrect responses due to defective recognition.

Step 4: Systematic Item Selection. After computing the average ratings of cultural familiarity, word frequency and age of acquisition, we carried a systematic selection process that aimed to discard pictures that are low in cultural familiarity and word agreement (figure 2). The first selection criterion relied on the picture's average cultural familiarity (CF) rating. Pictures that obtained an average rating on CF of 3.0/5 and higher were included in the first subset. Out of the 219 pictures in the picture pool, 184 pictures met the first selection

criterion. The second selection criterion was based on name agreement among the committee members in Lebanese Spoken Arabic. Name agreement among committee members refers to the extent to which they agree on a specific name for the picture. In the process of selecting words based on name agreement, polysemic words (words that could have several correct synonyms) were also eliminated. Modal names (most frequently provided name) were derived for each picture. Out of the 184 pictures remaining, 157 pictures were given the same name in Lebanese Spoken Arabic at least 80% of the time across the committee members. All modal responses provided by the members in English matched the names provided by the authors of the MultiPic Databank, which meant we were able to keep the picture names in English as assigned in the original study. After discarding items based on low cultural familiarity and low name agreement, 157 items remained. The 157 items were all assigned during the rating process an Age of Acquisition below 9 years old and had a mean word frequency of 3.7/5 with a standard deviation of 0.9, which indicates varying frequency of use in daily life.

<Insert Figure 2>

Phase 2: Test Administration. In the second phase, the test was administered on the 74 participants to obtain preliminary results on the test's validity and reliability. Test administration and data collection were carried out exclusively by the researcher over a period of 4 weeks in school classrooms under standardized testing conditions: noise-free, well lit, containing only the researcher and the participant sitting facing each other. Test administrator built rapport with every child by introducing herself, describing the task and receiving oral assent from the child to carry out test administration. The administrator consistently provided the following instructions in English, French or Arabic, depending on the child's preferred language as mentioned by the child:

> *I will show you pictures and I will ask you to name each picture using one word. You can say the name in any language you choose. Just say one word to name each picture. If you cannot think of a name, we can skip the picture and try another one! I will be writing down your answers the same way you say them and no answer is a wrong answer. Do you have any questions?*

Participants were generally compliant during the task and showed enthusiasm by smiling, responding well to verbal praise and sustaining their attention throughout the entire task duration. The test was administered in full to all 74 children of all ages. There was no discontinuing rule or ceiling. Participants generally responded promptly to each stimulus and said "pass" or "I don't know" when they did not have a name for the picture. Average duration of test administration was around 20 minutes per child. Spontaneously elicited names provided by students were recorded verbatim in writing during administration and were initially scored as neither correct nor incorrect. The examiner provided a prompt for the first two items if the participant did not seem to understand test instructions (examples of prompts: *what do you call this?* or *This is a…?).* Semantic and phonemic cues were not provided. If the child seemed to be attending to an irrelevant or different aspect of the picture than what is expected, the examiner provided these types of cues, which were adopted from the EOWPVT-4 (Martin & Brownell, 2011):

- *"What kind?"* was used when the response was too general (e.g., child says fruits for apple).
- *"What else is it called?"* was used when the response was too specific (e.g., child says Mercedes for car).
- *"What is this?"* while pointing at the picture was used when the response described a verb or only a part of the picture was named (e.g., child says flying for helicopter, or, camera for photographer).

Other variables. Demographics collected for each participant were age, gender and type of schooling (private or public). In the context of this study, type of school in Lebanon is considered an indicator of the child's SES with students enrolled in public schools coming from a low SES and those enrolled in private schools from middle to high SES. Other records pertaining to the students' academic performance were inaccessible for review and were therefore not collected.

Data Entry and Preprocessing. Responses of each participant were entered verbatim on Microsoft Excel 2016. Responses in Spoken Lebanese Arabic were entered using the Arabic Chat Alphabet of the "Arabizi" where Arabic words are encoded using Latin script and numbers (For more on Arabizi, refer to Yeghan, 2008). After response entry, responses underwent a thorough and meticulous cleaning process in order to code responses as correct or incorrect. Data preprocessing is described below:

- Long responses were collapsed so that any response that includes the target word and additional details was reduced to only the target word.
- Basic variants of the target word in terms of pronunciation due to regional variations in dialect were changed to the most frequently provided target word by the committee members.
- English or French words that are "Arabized" (meaning they were transformed over the years to resemble words in the Arabic dialect) were changed back into their original language.
- Plural forms of the words were changed to singular.

After collapsing and transforming the responses, the answers were coded as correct and incorrect. Responses that match the target word verbatim in Lebanese Arabic, English or French were coded as correct. All other responses were coded as incorrect and those include responses that are semantically related to the target word but conceptually distinct, responses that are phonetically related to the target word but semantically distinct, responses that are

visually related to the target word but semantically distinct, and responses that are unrelated to the target word. The final score, referred to as "total score", was tabulated by adding all correct answers.

Data Analysis

Descriptive Statistical Analysis

Data analysis investigated performance across age, gender and type of schools in addition to a measure of internal reliability. Although at this point, our sample is not large enough to draw conclusions on differences between group performances, a description of performance across gender, age, and type of schooling can be carried out to provide us with preliminary indicators of variance in test performance. All coded data were entered into an SPSS database version 24. Means and standard deviation of total score were calculated for age, gender groups and type of schooling groups. To allow for better descriptive analysis, we grouped participants in three age categories: 3:0 years to 5:11 years, 6:0 years to 7:11 years and 8:0 years to 9:11 years.

Between Groups Analysis. One-way independent analysis of variance (ANOVA) was conducted to compare means of test performance across age groups. Two-tailed group comparison was carried between males and females in the total sample, and one-tailed group comparison was carried between children from private schools and children from public schools. Logistic Regression using Enter method was carried with the variables age, gender and type of schooling to check if the model significantly accounts for variance in the total score.

Within Test Items Analysis. At the scale level, measures of internal reliability were conducted by calculating Cronbach's alpha in each age group and in the total sample.

Results

Sample Characteristics and Descriptive Results

The total sample consisted of 74 participants including 33 Males (44.6%) and 41 Females (55.4%). The mean age of the participants was 6.57 years (*SD* = 2.13) ranging between 3 years and 9 years. Forty-seven (47) students attended private schools (63.5%) and 27 students attended public schools (36.5%). Although 42% of the student are Arabic-English bilingual while 58% are Arabic-French bilingual, answers provided were consistently more prevalent in English. Table 4 shows the distribution of answers across languages on the first 10 items. Mean performance on the picture-naming test across age groups is presented in Table 5.

<Insert Table 4>

<Insert Table 5>

Preliminary Indicators of Construct Validity: Data Analysis at the Sample Level

Comparing Means across Age Groups. We conducted a one-way ANOVA to compare mean test score across the three age groups. Given that the data violates the assumption of homogeneity of variance, Welch's F-test are reported. Results indicate that there was a significant difference between mean scores across the three different age groups: F (2, 37.87) = 16.89, $p < .001$. Games-Howell post-hoc tests revealed significant differences between the first age group (3-5) and both the second (6-7), $p < .001$, and third age group (8-9), $p < .001$, however, there was no significant difference between the second group and the third group, $p = .723$, *ns*.

An item analysis examining item difficulty index (DIF) showed that out of the 157 items administered, 68 items, 104 items, and 111 items, were found to have a DIF of .61 and above (meaning that 61% of the participants answered this item correctly) when administered respectively to the first group (3 to 5 years), second group (6 to 7 years) and third group (8 to 8 years). This likely explains the small difference in test performance

between the two older age groups as almost all test items that were "easy" for those who are 6 and 7 years old were also considered "easy" by those who are 8 and 9 years old.

Comparing Means across Gender and Type of Schooling. A Mann-Whitney-*U*-test was carried out to compare performance of female participants (N = 41, M = 107.71, SD = 24.41) to male participants (N = 33, M = 98.21, SD = 22.02) across all age groups. Females scored significantly higher than males on the test with U= 492.50, z = -2.00, p = .045 and a small-to-medium effect size r = .23. Additionally, the total number of students enrolled in private schools (N = 47, M = 105.80, SD = 25.99) scored significantly higher than the total number of children enrolled in public schools (N = 27, M = 99.40, SD = 18.81) with U = 478.5, z = -1.75, p = .040 (*one-tailed*) and a small-to-medium effect size of r = .20.

Predictive Analysis. Next, we conducted a linear regression analysis predicting total score using age, gender and type of schooling as predictors. Table 6 shows that the variable Age (b = 7.80, β = .70, t (70) = 7.70, p < .001) significantly predicted total correct responses. The beta value indicates that as age increases, total score increases. Similarly, type of schooling (b = -18.76, β = -.38, t (70) = 4.24, p < .001) significantly predicted total correct responses. The beta value indicates that if a student belongs to a private school, they are more likely to receive higher total score on the test. On the other hand, gender did not significantly predict total correct responses (b = 5.73, β = .12, t (70) = 1.4, p = .17 *ns*).

<Insert Table 6>

Reliability: Data Analysis at the Scale Level

Cronbach's alpha coefficients were computed by age groups (3-5; 6-7; 8-9) and for the total sample (N= 74). These coefficients shown in Table 7 are high ranging from 0.90 to 0.95 for the various age groups, which indicates a large-scale homogeneity of the test items.

<Insert Table 7>

Discussion

This study described the development of a multilingual measure of expressive vocabulary using the dual-focus approach for test development and words with known psycholinguistic variables (figure 3). The measure compiled 157 colored images that belong to various semantic categories, are rated as culturally familiar to the target population, have a high rate of name agreement, have an AoA between the ages of 3 and 9 years old and a varying frequency of use in daily life.

<Insert Figure 3>

The dearth of locally developed picture-naming tests or measures of vocabulary that are culturally and linguistically suitable for Lebanese children makes it difficult to compare our study's results to previous local findings. However, the results obtained concerning group performances across age, gender and type of schooling can be linked to previous literature on vocabulary acquisition and provide evidence of validity to the test results. First, age was found to be a contributing factor to test performance on the LPNT. Many studies attest to the fact that age is a major variable in vocabulary gain (Basilio, Puccini, Silva, & Pedromónico, 2005; Bates, Dale, & Thal, 1995; Vogt, Douglas, & Aussems, 2015). This was evident in our study as older children performed better on the picture-naming test and age was found to be the highest contributing factor to an increase in total score. However, there was no significant difference in test performance between the two upper age groups (6-7 and 8-9). Second, the analysis of variance across the entire sample showed that females performed significantly higher than males. However, when gender was added to the model along with age and type of school, it did not show to be a significant contributing factor to vocabulary gain. Gender is not consistently reported as a contributing factor to improved naming ability (Zec, Burkettm, Markwell & Larsen, 2007) and our results showed similar findings. Third, results also show that children enrolled in private schools performed significantly higher than children enrolled

in public schools. As mentioned, we assume that students enrolled in private schools generally have the financial means to afford this type of education and usually belong to a higher SES than those enrolled in public schools. Studies on the relationship between socio-economic background and vocabulary show that children from lower SES build their vocabularies at a slower rate than children from high SES (Feldman et al., 2000; Rescorla & Alley, 2001). Predictive analysis of an increase in test performance on the developed measure revealed that age was the strongest contributing factor, followed by type of schooling, with gender showing no predictive power. Analysis at the item level showed that the 157 test items had an excellent internal consistency across the three age groups and in the total sample indicating a high homogeneity between the test items.

Limitations

A major limitation relates to the Lebanese dialect and the regional variations of names in terms of pronunciations, plural forms, and use. For example, one-word *spoon* had at least three different responses provided in Arabic ("*ma3l2a*", "*mal3a2a*", and "*mal3aqa*"). Reducing the three responses into one of the provided answers dismissed several other variations when administering the test in different areas in Lebanon. This highlights the importance of having the test administrator familiar with the Lebanese culture and dialect in order to differentiate name variations due to dialectical differences from mispronunciations or incorrect responses. This limitation may be difficult to account for as some words in Arabic can have tens of different pronunciations some of which only differ by one diacritic (accent on a letter). Another limitation relates to the sample. Efforts were made to render the sample as representative as possible of the Lebanese population however, this was easier to achieve on the level of the schooling variable, and more difficult across age ranges and gender. It would have been ideal to have both genders equally represented in the sample.

We suggest that our study findings serve as preliminary evidence of validity and reliability to the LPNT. Our hope is that the procedure adopted in the construction of this test deems useful to researchers who seek to develop naming measures culturally and linguistically suitable to multilingual populations.

References

Apiwattanalunggarn, K. L., & Luster, T. (2005). Individual differences in the school performance of 2nd-grade children born to low-income adolescent mothers. Journal of Research in Childhood Education, 19, 314 – 332. doi:10.1080/02568540509595074

Armon-Lotem, S., de Jong, J., & Meir, N. (2015). Methods for assessing multilingual children: disentangling bilingualism from Language Impairment: Multilingual Matters.

Bacha, N. N., & Bahous, R. (2011). Foreign Language Education in Lebanon: A Context of Cultural and Curricular Complexities. Journal of Language Teaching and Research, 2(6), 1320-1328. http://dx.doi.org/10.4304/jltr.2.6.1320-1328

Bahous, R. N., Nabhani, M. B., & Bacha, N. N. (2014). Code-switching in higher education in a multilingual environment: a Lebanese exploratory study. *Language Awareness*, *23*(4), 353-368.

Basílio, C. S., Puccini, R. F., Silva, E. M., & Pedromônico, M. R. (2005). Living conditions and receptive vocabulary of children aged two to five years. *Revista de saude publica*, *39*, 725-730.

Bates, E., Dale, P. & Thal, D. 1995. "Individual differences and their implications for theories of language development". *In Handbook of child language*, Edited by: Fletcher, P. and MacWhinney, B. 96–151. Oxford: Basil Blackwell.

Bialystok, E. (2007). Acquisition of literacy in bilingual children: A framework for research. *Language Learning*, *57*, 45-77.

Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. First Language, 24(3), 267–305.

Catroppa, C., & Anderson, V. (2004). Recovery and predictors of language skills two years following pediatric traumatic brain injury. *Brain and Language*, *88*(1), 68-78.

Center for Educational Research and Development. (2018). Retrieved from https://www.crdp.org/stat-details?id=26000&la=en

Champion, T. B., Hyter, Y. D., McCabe, A., & Bland-Stewart, L. M. (2003). "A matter of vocabulary" performances of low-income African American head start children on the Peabody Picture Vocabulary Test—III. *Communication Disorders Quarterly*, *24*(3), 121-127.

Crystal, D. (2004). The Past, Present, and Future of World English. In: A. Gardt and B. Hüppauf (eds) *Globalization and the Future of German*, 27-45. Berlin/New York: Mouton de Gruyter.

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. Quarterly Journal of Experimental Psychology, 71(4), 808-816.

Dupere, V., Leventhal, T., Crosnoe, R., & Dion, E. (2010). Understanding the positive role of neighborhood socioeconomic advantage in achievement: The contribution of home, child care, and school environments. Developmental Psychology, 46(5), 1227–1244.

Erkut, S., Alarcón, O., Coll, C. G., Tropp, L. R., & García, H. A. V. (1999). The dual-focus approach to creating bilingual measures. *Journal of Cross-Cultural Psychology*, *30*(2), 206-218.

Etard, O., Mellet, E., Papathanassiou, D., Benali, K., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2000). Picture naming without Broca's and Wernicke's area. *Neuroreport*, *11*(3), 617-622.

Feldman, H. M., Dollaghan, C. A., Campbell, T. E, Kurs- Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communi- cative Development Inventory at ages one and two years. Child Development, 71, 310-322

Giddan, J. J., & Milling, L. (1999). Comorbidity of psychiatric and communication disorders in children. *Child and adolescent psychiatric clinics of North America*, *8*(1), 19-36.

Grabowski, T. J., Damasio, H., Eichhorn, G. R. & Tranel, D. Effects of gender on blood flow correlates of naming concrete entities. Neuroimage 20, 940–954 (2003).

Hamadani, J., Tofail, F., Hilaly, A., Huda, S., Engel, P., & Grantham-McGregor, S. (2010). Use of Family Care Indicators and their relationship with child development in bangladesh. Journal of Health Population and Nutrition 28(1):23–33.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. Child Development, 74, 1368–1378.

Horwitz, S.M., Irwin, J.R., Briggs-Gowan, M.J., Bosson Heenan, J.M., Mendoza, J. & Carter AS. Language delay in a community cohort of young children. Journal of the American Academy of Child & Adolescent Psychiatry 2003;42:932–940.

İlkman, S. (2015). *Reading Acquisition in Primary School-age Children* (Master's thesis, Eastern Mediterranean University (EMU)-DoğuAkdenizÜniversitesi (DAÜ).

Khomsi, A. (2001). Evaluation du langage oral (ELO). ECPA : Paris.

Kotob, H. (2002). Bilingue malgré lui ! In: Université de Balamand (Ed.), La Psycholinguistique au Liban: Etat des lieux (pp. 59–75). Liban: Publications de l'Université de Balamand.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H., 2008, Language assessment and development in toddlers with autism spectrum disorders. Journal of Autism and Developmental Disorders, 38, 1426–1438.

Lyytinen, H., Aro, M., Eklund, K., Erskine, J., Guttorm, T. K,, Laakso, M.-L., Lep pänen, P. H. T., Lyytinen, P., Poikkeus, A.-M., Richardson, U., & Torppa, M. (2004). The development of children at familial risk for dyslexia: birth to school age. Annals of Dyslexia, 54, 184—220

Martin, N. A., & Brownell, R. (2011). *Expressive One-Word Picture Vocabulary Test 4*. Novato, CA: Academic Therapy Publications.

Northrup, D. (2005) Globalization and the Great Convergence: Rethinking World History in the Long Term. *Journal of World History* 16(3): 249–267.

Oliver, B., Dale, P. S., & Plomin, R. (2004). Verbal and nonverbal predictors of early language problems: An analysis of twins in early childhood back to infancy. *Journal of Child Language*, *31*(3), 609-631.

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*(4), 1255-1264.

PNUD report, Lebanon toward a citizen state, 2009, p 132.

Rescorla, L., & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. Journal of Speech, Language, and Hearing Research, 44, 434-445

Sayigh, R. (1965). The bilingualism controversy in Lebanon. *The World Today*, *21*(3), 120-130.

Shaaban, K. A. (1997). Bilingual education in Lebanon. *Bilingual Education* (pp. 251-259). Springer Netherlands.

Simhairi, V. (2010). *An Investigation into the Validity of Adopting Western Psycho-educational Assessments in a Society They are Not Designed for: The Case of Lebanon* (Doctoral dissertation, King's College London).

Spinelli, M., Rocha, A. C. D. O., Giacheti, C. M., &Richieri-Costa, A. (1995). Word-finding difficulties, verbal paraphasias, and verbal dyspraxia in ten individuals with fragile x syndrome. *American Journal of Medical Genetics Part A*, *60*(1), 39-43.

Tervo, R. C. (2007). Language proficiency, development, and behavioral difficulties in toddlers. *Clinical Pediatrics*, *46*(6), 530-539.

van Viersen, S., de Bree, E. H., Verdam, M., Krikhaar, E., Maassen, B., van der Leij, A., & de Jong, P. F. (2017). Delayed early vocabulary development in children at family risk of dyslexia. Journal of Speech, Language, and Hearing Research, 60, 937–949. doi:10.1044/2016_ JSLHR-L-16-0031

Vogt, P., Douglas, J. M., & Aussems, S. (2015). Early vocabulary development in rural and urban Mozambique. *Child Development Research*, *2015*, 15. doi:10.1155/2015/189195

Vogt, P., Douglas, J. M., & Aussems, S. (2015). Early vocabulary development in rural and urban Mozambique. *Child Development Research*, *2015*, 15. doi:10.1155/2015/189195

Yaghan, M. A. (2008). “Arabizi”: A contemporary style of Arabic slang. *Design Issues*, *24*(2), 39-52.

Zebib, R., Henry, G., Khomsi, A., Messarra, C., & Hreich, E. (2017). Evaluation du langage oral chez l’enfant libanais: ELO-L, Liban Tests Edition (ECPA Pearson).

Zec, R. F., Burkett, N. R., Markwell, S. J., & Larsen, D. L. (2007). A cross-sectional study of the effects of age, education, and gender on the Boston Naming Test. *The Clinical Neuropsychologist*, *21*(4), 587-616.

Table 1. *Distribution of the Sample across Age and Gender*

Age (years)	Females (N=44)	Males (N=31)	Total (N=74)	
			N	%
3	6	6	12	16.2
4	4	1	5	6.8
5	1	3	4	5.4
6	8	3	11	14.9
7	5	4	9	12.2
8	10	6	16	21.6
9	10	7	17	23.0
Total %	55.4	44.6	74	100.0

Table 2. *Members of the Committee of Experts*

	Current Profession	Highest Level of Education	Years of Experience	Language Proficiency
1	Assistant Director of a Preschool in Beirut	Graduate Studies	>30 years	AR-FR-ENG
2	Doctor in Educational Psychology	Doctorate Studies	>30 years	AR-FR-ENG
3	Speech and Language Therapist	Graduate Studies	>15 years	AR-FR-ENG
4	Special Education Coordinator at a Private School in Beirut	Teaching Diploma in Special Education	7 years	AR-ENG
5	Neuropsychologist	Graduate Studies	>8 years	AR-FR-ENG
6	Consultant on Childhood Education and Literacy Coach	Graduate Studies	>30 years	AR-ENG
7	Child Behavioral Therapist	Graduate Studies	7 years	AR-FR-ENG
8	English Language Homeroom Teacher	Graduate Studies	>15 years	AR-ENG

Table 3. *Proportions of Semantic Categories*

Superordinate Category	% in the 157 pictures	% in the 219 pictures
Natural Kind	30.3	29.7
Animals and insects	13.2	12.8
Food	11.2	9.6
Natural element/plant	5.9	7.3
Artifact	58.7	57.1
Clothing (or part of) and accessories	9.9	8.7
Furniture	8.6	7.3
Tools	7.9	8.7
Container/receptacle	6.6	4.6
Kitchen utensils and appliances	5.3	4.6
Vehicle (or part of)	5.3	4.6
Desk/writing material	3.9	4.1
Media and communication tools	3.3	3.2
Toy/game	3.3	5.9
Shape	2.6	1.8
Musical instrument	2.0	2.7
Fiction	0.0	0.9
Activity	11.2	13.3
Human (or body part) and interaction	7.2	7.3
Outdoor places or parts	2.0	2.3
Profession	2.0	3.7

Table 4. *Percent Language Responses of 10 First Items*

Item	% Arabic Responses	% French Responses	% English Responses	% No responses
Dog	20	22	58	0
Hand	27	18	54	1
Apple	20	22	58	0
Balloon	26	24	47	3
Car	24	23	53	0
Bird	30	19	46	5
Fish	26	22	53	0
Banana	19	20	61	0
Sun	27	22	50	1
Spoon	36	15	34	15

Table 5. *Mean Performance Across Age Groups*

Age	N	Mean	SD	SE	Min	Max
3 to 5	21	80.52	23.32	5.08	46.00	119.00
6 to 7	20	110.15	18.99	4.24	70.00	142.00
8 to 9	33	114.03	15.44	2.68	68.00	138.00
Total	74	103.47	23.69	2.75	46.00	142.00

Table 6. *Regression Parameters*

Model		B	SE B	β
1	(Constant)	68.95	9.06	
	Gender	5.73	4.10	.12
	Age	7.80	1.01	.70***
	Schooling	-18.76	4.43	-.38***

***. Correlation is significant at the 0.001 level.

Table 7. *Internal Reliability Measures*

Age Groups	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
3 to 5	.958	.955	149
6 to 7	.935	.934	143
8 to 9	.909	.910	141
Total (N=74)	.958	.957	155

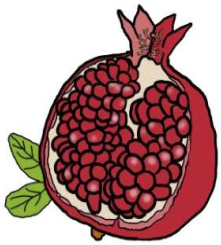
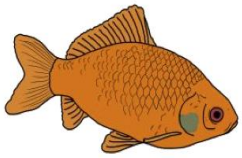
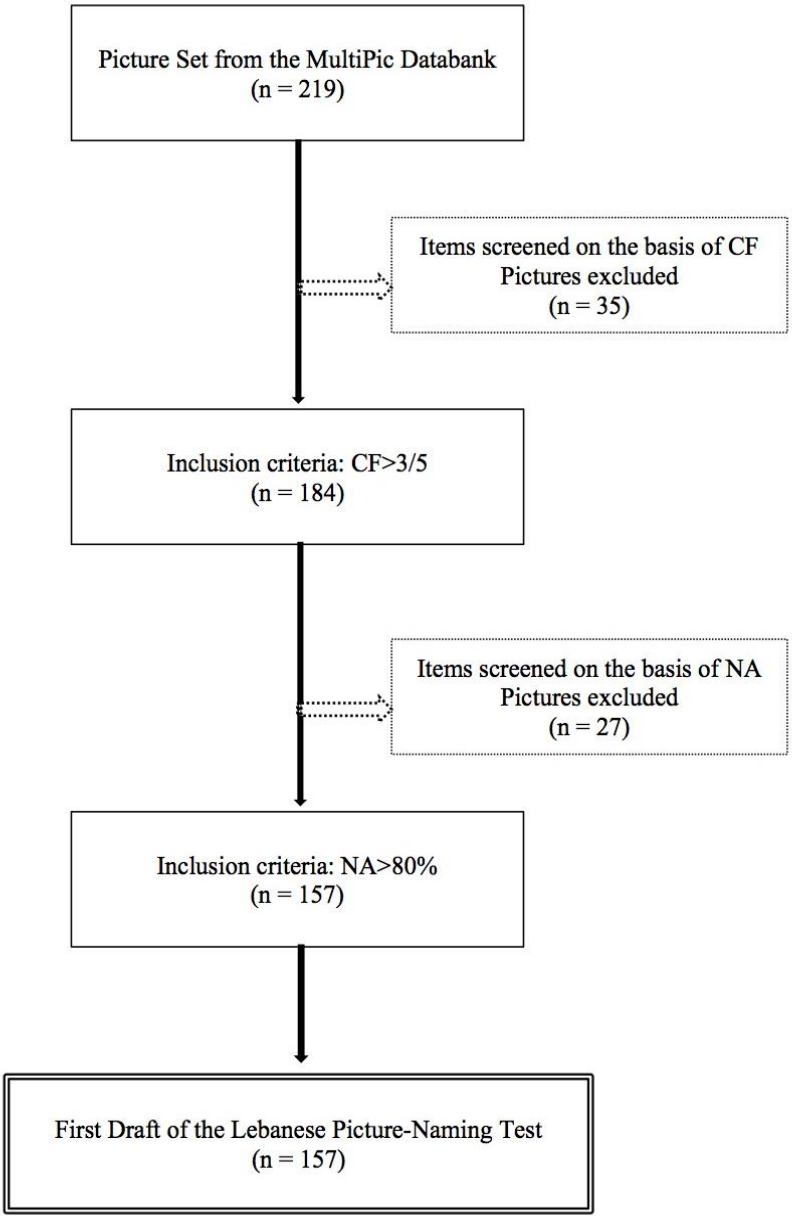
		
Pencil (Desk/Writing Material)	Leaf (Natural Element/ Plant)	Pomegranate (Food)
		
Fish (Animals)	Thumb (Human (Body Part) and Interaction)	Tambourine (Musical Instrument)

Figure 1. Examples of pictures from the MultiPic Databank from various semantic categories

Figure 2. Flowchart of the Systematic Process for Picture Selection



Notes. CF = Cultural Familiarity. NA = Name Agreement.

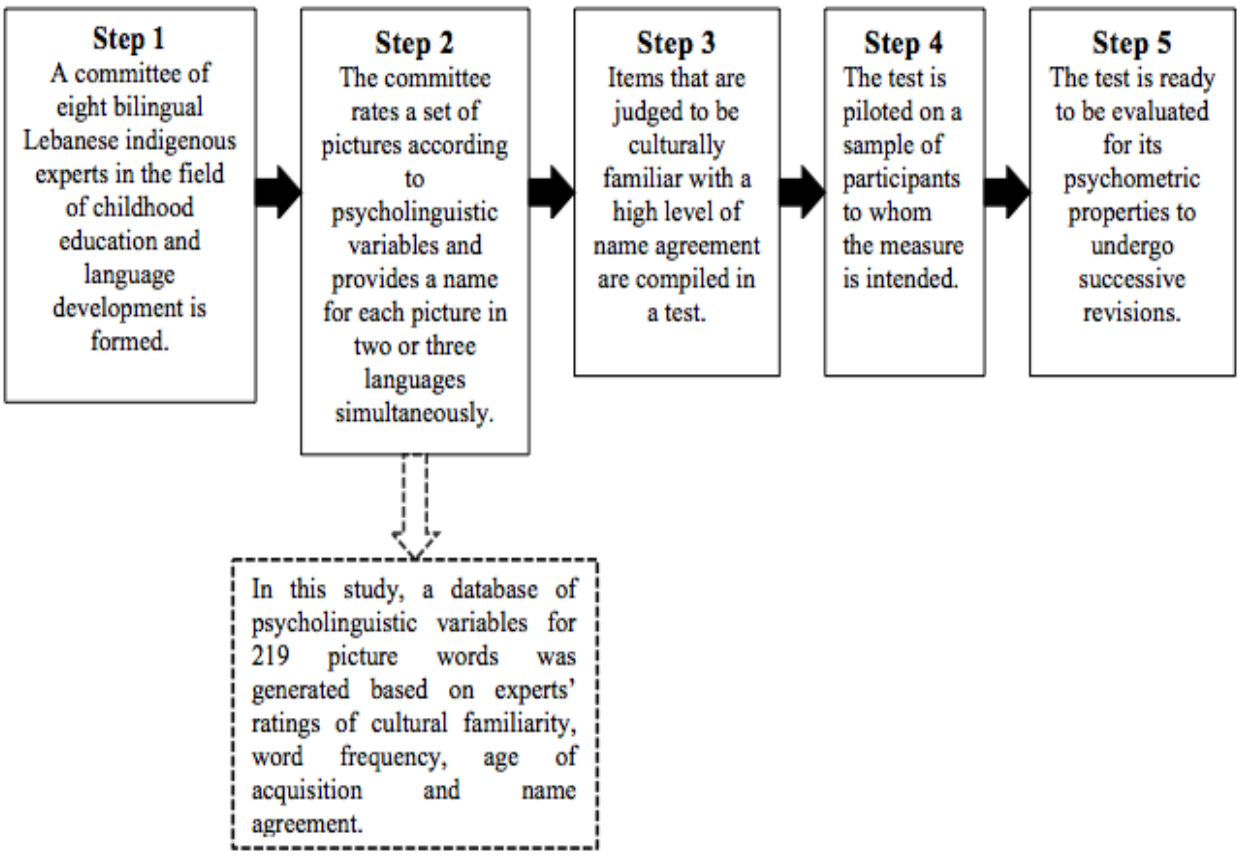


Figure 3. Implementation of the Dual-Focus Approach (Erkut et al., 1999) in the development of the first Lebanese picture-naming test.