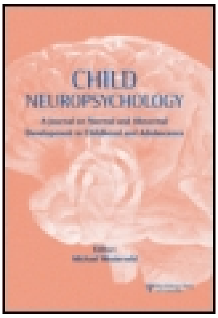

Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ncny20

Reliability and validity of the Arabic version of the computerized Battery for Neuropsychological Evaluation of Children (BENCI)

Ahmed F. Fasfous[ab], Maria Isabel Peralta-Ramirez[ab], María Nieves Pérez-Marfil[a], Francisco Cruz-Quintana[ab], Andrés Catena-Martinez[ab] & Miguel Pérez-García[ab]

[a] Facultad de Psicología, Universidad de Granada, Granada, Spain

[b] Centro de Investigación Mente, Cerebro y Comportamiento (CIMCYC), Universidad de Granada, Granada, Spain

Published online: 03 Apr 2014.

Child Neuropsychology, 2014
http://dx.doi.org/10.1080/09297049.2014.896330

Reliability and validity of the Arabic version of the computerized Battery for Neuropsychological Evaluation of Children (BENCI)

Ahmed F. Fasfous[1,2], Maria Isabel Peralta-Ramirez[1,2], María Nieves Pérez-Marfil[1], Francisco Cruz-Quintana[1,2], Andrés Catena-Martinez[1,2], and Miguel Pérez-García[1,2]

[1]Facultad de Psicología, Universidad de Granada, Granada, Spain
[2]Centro de Investigación Mente, Cerebro y Comportamiento (CIMCYC), Universidad de Granada, Granada, Spain

Batería de Evaluación Neuropsicológica Infantil (BENCI) is a computerized battery for the neuropsychological evaluation of children. This battery has been used in different studies to evaluate neuropsychological functions and neurodevelopment in children. The objective of this study is to test the validity and reliability of the first Arabic version of the BENCI on an Arabic population where neuropsychological tests are very scarce. We administrate the BENCI to 198 school-age children (98 boys and 100 girls) from Morocco. To examine the test retest reliability of the BENCI battery, we administered the battery 2 times to 43 children (23 boys and 20 girls) with 15 days in between the pre- and posttest. The results revealed good validity and reliability of the battery in Arabic children. Also, the BENCI battery has demonstrated the capacity to differentiate between children by their age group. This battery can be of great use to both the research and clinical areas of Arabic countries and/or in assistance to Arabic immigrants that live outside of their native country.

Keywords: Arab children; Neuropsychological battery; Morocco; test-retest; neurodevelopment.

Neuropsychological tests are scarce in developing countries (Nell, 2000) and are even less available for children. Recently, our team has developed a computerized battery for the neuropsychological evaluation of children (*BENCI: Batería de Evaluación Neuropsicológica Infantil*; Cruz-Quintana, Pérez-García, Roldan-Vílchez, Fernández López, & Pérez-Marfil, 2013) by means of international cooperative projects. This battery has been used in different studies to evaluate neuropsychological domains and neurodevelopment in children (Cruz-Quintana et al., 2013) and has demonstrated discriminate validity in a study comparing the neuropsychological performance of preterm and normal 7-year-old children (García Bermúdez et al., 2012). Also, BENCI subtests have shown

This research is a part of project (Project number: PI18X) that has been carried out with the financial of the *Centro de Iniciativas de Cooperación al Desarrollo (CICODE)—Granada University-Spain.*

Address correspondence to Ahmed Fasfous, Department of Clinical Psychology, Faculty of Psychology, Universidad de Granada, Granada, Spain. E-mail: Fasfous@ugr.es

good convergent validity with other similar and validated paper-and-pencil tests and have demonstrated good test-retest reliability in a sample of Hispanic children (Cruz-Quintana et al., 2013). The BENCI includes tests that evaluate different neuropsychological areas such as processing speed, visuomotor coordination, attention, memory, language, and executive functions. In addition to the fact that the BENCI battery has good psychometric properties, it is easy and enjoyable for children.

In the Arab world, over 350 million inhabitants live divided between 22 countries (Mirkin, 2010). Of these inhabitants, over 35% are children (Mirkin, 2010). These children have been exposed to distinct factors (Shonkoff, 2010; Shonkoff, Richter, van der Gaag, & Bhutta, 2012) that can affect their neurodevelopment. Among these factors is the exposure to war violence (e.g., Palestine and Iraq) or poverty and malnutrition, (e.g., Mauritania and Somalia). Nevertheless, there is a very scarce number of studies that have evaluated the impact of these factors on the neuropsychological development of these children (Abu Zaydeh, Zalina, Wan, & Aljeesh, 2012; El Hioui, Azzaoui, Touhami Ahami, Rusinek, & Aboussaleh, 2012). Therefore, it is important to evaluate the neurodevelopment of Arabic children in order to understand how these variables are influential and to develop prevention programs and interventions.

Nevertheless, despite the millions of people in the Arabic population, there are very few neuropsychological tests adapted to this culture. In a literature review about tests mostly used in Arabic culture, Fasfous, Puente, Peralata-Ramírez, and Pérez-García (2014) have found that more than 46% of the neuropsychological tests had not been translated, adapted, or its psychometric properties had not been examined in this culture. Only 13 tests had normative data in some Arab countries (Fasfous et al., 2014). Literature reveals a cultural effect on neuropsychological tests (e.g., Ardila, 1995, 2005; Greenfield, 1997; Puente, Perez-Garcia, Vilar-Lopez, Hidalgo-Ruzzante, & Fasfous, 2013). Recently, of the few neuropsychological studies in the Arab world, different works demonstrate that there is cultural difference in neuropsychological performance between Arabs and individuals from other cultures (e.g., Fasfous, Hidalgo-Ruzzante, Vilar-Lopez, Catena-Martinez, & Perez-Garcia, 2013; Sobeh & Spijkers, 2013). Sobeh and Spijkers (2013) have found a differential attentional pattern when comparing Syrian and German children. Fasfous and colleagues (2013) found that Spanish adults outscored Moroccan Arab adults in verbal and nonverbal neuropsychological tests where each cultural group utilized different neuropsychological abilities to perform the same intelligence test. Different studies have confirmed the need to create and validate specified neuropsychological tests for the Arabic population (Alansari & Baroun, 2004; Fasfous et al., 2013).

In summary, despite the large number of inhabitants in the Arabic world, there are few valid neuropsychological tests, and the comprehensive batteries for the evaluation of children are nonexistent. To our knowledge, BENCI is the first computerized battery to evaluate the neurodevelopment of Arabic children. Therefore, the objective of this study is to test the validity and reliability of the Arabic version of the computerized battery for the neuropsychological evaluation of children (BENCI) in a sample of Moroccan children. This battery uses the classic Arabic language. Classic Arabic is the common and official language for the 22 Arab countries and it is the common writing system in those countries. School children learn classic Arabic in their schools.

METHOD

Participants

In this study, 198 children participated (98 boys and 100 girls) from the city of Chefchaouen, Morocco. This city is the capital of the Chefchaouen province and it has more than 35,000 inhabitants in the northwest of Morocco. The main economic activities are trade and tourism. The unemployment rate is currently 7.18% and the literacy rate was 26% in 2004 (Chefchaouen Townhall, 2013). The participants were selected from two public schools from the middle-class neighborhoods of this city. Both schools have a computer lab where the BENCI tests were administered. According to the recorded information, most of the parents were working as civil servants or private business owners. The sample was made up of three age groups: 7-year-olds (51 children, 25 boys and 26 girls from second grade), 9-year-olds (63 children, 30 boys and 33 girls from fourth grade), and 11-year-olds (84 children, 43 boys and 41 girls from sixth grade). To examine the test-retest reliability of the BENCI battery, we administered the battery two times to 43 children (23 boys and 20 girls) with 15 days in between the pre- and posttest.

Schools were selected by the Chefchaouen Education Office of the Moroccan Ministry of Education based on the representation criteria of the children of Chefchaouen. There were two classrooms for each grade and 52 children (26 boys and 26 girls) were selected randomly from each classroom. And from the fourth and sixth grades, more than 43 children were selected to conduct the study of test-retest reliability. All of the selected children had accepted to participate in this study, except for 1 child from the second grade who did not complete the battery.

According to professors and parents in the initial interviews, all of the children were free of medical problems. The questions were about antecedents of any relevant illness and medication intake, and this information was included as a part of a comprehensive interview about health status. Nevertheless, the children with developmental problems attend a special and separate program in the Moroccan educational system, and all of the children from our study were selected from the ordinary schools. In ordinary schools, children attend 4 hours (two in French per day) in the classroom, 6 days a week. During that time, they take classes in science, math, Arabic language, French, and social sciences. Arabic is the native and everyday speaking language.

Instruments

The original version of the BENCI battery (Cruz-Quintana et al., 2013) was developed among Ecuadorian children as a result of an International Cooperative Program. BENCI is composed of basic neuropsychological domains required to conduct a comprehensive neuropsychological assessment (Lezak, Howieson, & Loring, 2004): speed processing, visuomotor coordination, attention, memory, language, and executive functions. Neuropsychological tests were developed to cover these domains. All these tests were developed using valid neuropsychological procedures based on the literature of neuropsychological assessment. For example, verbal episodic memory tests usually use a list of words (Delis, Kramer, Kaplan, & Ober, 1987; Rey, 1964). The BENCI battery also includes a list of new words; the number of these words depends on the age of children. In order to evaluate the validity of the BENCI subtest, Cruz-Quintana et al. calculated the correlation between the BENCI subtests and other valid tests. For example, visuomotor tests were significantly correlated with the Children's Color Trails Test (CCTT-1&2:

Llorente, Williams, Satz, & D'Elia, 2003) and the verbal memory test with the Spanish adaptation of California Verbal Learning Test (Test de Aprendizaje Verbal España-Complutense Infantil (TAVECI): Pamos, Alejandre, & Benedet, 2007).

As a result, the BENCI consists of 14 neuropsychological tests. This version has norms for 216 (108 boys and 108 girls) Ecuadorian children between 6 and 11 years old and it has demonstrated good psychometric properties. The test-retest reliability was good (correlation ranged between $r = .927$ and $r = .351$). The convergent validity of the BENCI was calculated using a number of valid tests such as the Stroop Color Word test (Golden, 2001), Backward Digits (Woodcock & Sandoval, 1996), Raven´s Progressive Matrices (Raven, 1977), CCTT (Llorente et al., 2003), and the Spanish adaptation of the California Verbal Learning Test (TAVECI: Pamos et al., 2007). Results showed acceptable correlations between the BENCI subtest and these tests (Cruz-Quintana et al., 2013). Specifically, the correlation between the visuomotor subtest and the CCTT-1 was .539; the alternate visuomotor subtest and CCTT-2 was .567; the Spatial Stroop subtest and Stroop test was .383; the verbal memory subtest and TAVECI was .689; the Abstract Reasoning subtest and Raven´s test was .605; and the Working Memory subtest and Backward Digits was .469. Furthermore, BENCI has demonstrated discriminant validity in a study that compared the neuropsychological performance of 25 preterm children (14 boys and 11 girls; M = 7 years and SD = 0.645) and 25 normal children (10 boys and 15 girls; M = 6.7 years and SD = 0.614) (García Bermúdez et al., 2012). Neuropsychological differences between the two groups were found in all of the BENCI subtests, except for in the reaction time (RT) and Planning subtests. These findings coincide with previous literature (García Bermúdez et al., 2012). Cohen´s d was 0.3 and 0.7 in the two visuomotor tests and ranged between 0.83 and 1.55 on the remaining tests, which is considered a large effect size. In addition to the good psychometric properties of the BENCI battery, it is computerized but requires no previous experience to be administered. This computerized format allows for standardized administration, records hits, errors, or RT (when proceeded) in a very easy manner, and it is simple and enjoyable for children.

The Arabic version of the BENCI battery (Cruz-Quintana et al., 2013) was administered to 198 children by an Arabic neuropsychologist (A. Fasfous). Translation and backtranslation were completed. One bilingual neuropsychologist translated the test from Spanish to Arabic and another bilingual neuropsychologist translated the BENCI from Arabic to Spanish. Adaptations to Arabic culture were needed for some words of the memory list and for some of the pictures (e.g., we changed the pig's picture to a sheep). We did not need to adapt numbers since Arabic numbers are used in Morocco. This battery has been translated and adapted according to the international Test Commission Guidelines and the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; International Test Commission, 2010).

BENCI consists of 14 neuropsychological tests. It takes about 75 minutes to complete in one session with a break of 10 minutes. The order of administration was the same for all subjects and the order of administration was determined following the recommendations of Lezak et al. (2004). Domains and tests are described in the following sections.

Processing Speed.

Simple Reaction Time Test. This test requires that the child press any key as fast as possible every time a cross (+) appears on the screen. For this test, the dependent measure was the mean RT in ms.

Visuomotor Coordination.

Visuomotor. The child should connect numbers in an increasing order (or elements, according to the given sequence) that appear on the screen out of order. The dependent measure was the total completion time in ms.

Alternate Visuomotor. The child should connect in an alternate order and reach the increasing numbers of the two distinct series that appear on the screen out of order. The dependent measure was the total completion time in ms.

Sustained Attention.

Continues Performance Test (CPT). In this task, various blocks in a series of letters (essays) will appear on the screen so that the child should press the key each time the correct stimulus appears (i.e., an A after an X). The rest of the letters are distractor elements. The dependent measures were RT in ms and the total number of correct answers (CA).

Memory.

Verbal Memory. The child listens to a series of words and should memorize as many as possible. After each sequence, the user should repeat all of the words he or she remembers. The dependent measure was the total number of CA.

Verbal Memory (Delayed Recall). After 20 minutes, the child should repeat out loud all of the words he or she can remember that were presented in the list of the previous task (test of *Verbal Memory*). The dependent measure was the total number of CA.

Verbal Memory (Essay of Recognition). The child listens to a series of words, half of which are from the previous list. The child answers "yes" or "no" to whether or not each word was in the memory list. The dependent measure was the total number of CA.

Visual Memory. Pictures are presented with a series of common objects to which the child should memorize the greatest number possible. After each sequence, the child should repeat all of the objects that he or she remembers. The dependent measure was the total number of CA.

Visual Memory (Delayed Essay). After 20 minutes, the child should name out loud all of the images that he or she can remember from the sequence in the visual memory task. The dependent measure was the total number of CA.

Visual Memory (Essay of Recognition). Next, a series of images will be presented, and for each one, the child must indicate if it had appeared in the sequence of the previous test. The dependent measure was the total number of CA.

Language.

Verbal Comprehension (Images). A combination of images of a given category (e.g., animals) are shown and the user is given a series of instructions (auditory), indicating that he or she should select a given image that fulfils the indicated conditions (e.g., type of animal, position, type of activity that one can carry out, and/or color: e.g., “Touch the frog that is next to the dog”). The dependent measure was the total number of CA.

Verbal Comprehension (Figures). Similar to the previous test, the images on the screen are replaced by geometric shapes (circles, triangles, squares) of different colors. The child has to select the ones that fulfil the indicated conditions (figure, size, position, and color: e.g., “Touch the small blue circle”). The dependent measure was the total number of CA.

Phonetic Fluency. This test indicates a letter and the child should respond, as fast as he or she can, with all of the words he or she knows that start with the same letter. The dependent measure was the total number of CA given within a time of 60 seconds.

Executive Function.

Working Memory. The child listens to a series of number sequences and colors. After each sequence, the user should separately repeat the numbers and the colors that he or she has heard (first all of the numbers and next all the colors or in reverse order). The dependent measure was the total number of CA.

Abstract Reasoning. This exercise is a sequence of logical series. For each one, the child should select the element that appears that completes the series. The dependent measures were RT in ms and the total number of CA.

Semantic Fluency. The child listens to a semantic category and he or she states all of the elements that he or she knows from the given category (colors/animals). The dependent measure was the total number of CA given within a time of 60 seconds.

Inhibition: Go/No-Go. During this test, two alternating elements keep appearing on the screen. In the first phase of the test, the child should state the distinguishing element of the two and press any key when it appears. Afterwards, by listening to a sound that represents the phase change, the distinguished element will appear to be the other, to which the user should press a key when it appears. The dependent measures were RT in ms and the total number of CA.

Flexibility: Spatial Stroop. A sequence of arrows appears on the screen, either pointing to the left (←) or to the right (→). Each time that the arrow appears, the child should press the LEFT ARROW key if the arrow points to the left or the RIGHT ARROW key if the arrow points towards the right. In some occasions, the stimuli are presented on the side of the screen in coherence with the arrow and in other occasions in the reverse order. The dependent measures were RT in ms and the total number of CA.

Planning: Attraction park. An attraction park is presented and the child’s goal is to collect the highest number of attractions with the money he or she is allotted. The user is informed that there is a time limit for being in the festival and that each attraction expires part of this time. The user’s objective should be to acquire the highest number of different attractions with the available resources. The dependent measure was the number of activities.

Procedure

This study was a part of an International Cooperative Project about neurodevelopment between the University of Granada and the town hall of Chefchaouen. After the Ethical Committee at the University of Granada (Spain) approved the study, permission by the Department of Education in the province of Chefchaouen (Morocco) was obtained in order to begin the study in the selected schools. Next, parental informed consent was collected for each of the participants.

All of the participants were selected randomly utilizing a list of students from the two chosen schools for this study. The researchers received permission from the parents and acceptance from the children to participate voluntarily in the study. The directors of the two schools provided an empty classroom in each school to carry out the study. Afterwards, the BENCI battery was administered on an individual basis to participants during the school day.

Finally, in order to validate the Arabic version of the BENCI, we had a lot of difficulties in finding neuropsychological tests that covered different neuropsychological domains in the Arabic population. For this reason, we did not complete convergent validity and used construct validity.

Statistical Analysis

G3*Power was used to calculate the sample size required for the test-retest reliability and convergence validity analysis. In order to obtain a Pearson's correlation = .35 with an alpha of .05 and a power of .80, 46 children in each group were required.

The sample characteristics were analyzed with descriptive statistics and frequency distributions. The test-retest reliability was calculated using the intraclass correlation coefficient (ICC), the recommended statistic for this type of study (Armstrong, White, & Saracci, 1992; Scientific Advisory Committee of the Medical Outcomes Trust, 2002). To determine this, a two-dimensional analysis of variance (ANOVA) of random effects was conducted. The obtained ICC is essentially the same as the analyzed Kappa coefficient based on quadratic weight in disagreement, provided that the number of observations is sufficiently large, as it is in this case (Fleiss & Cohen, 1973). Given that the calculated ICC of this study is a measurement of absolute agreement, the same as the Kappa coefficient, the standards of Landis and Koch (1977) were followed in order to interpret the magnitude. The analyses fulfilled the scaled units of measurements (reliability) of the statistical program SPSS version 20.

To confirm the factorial structure of the BENCI, a confirmatory factorial analysis was carried out using AMOS 5.0. The method of confirmatory factorial analysis was based on the method of standard estimate in models of structural equations and on the criteria of the highest probability, in accordance with the assumption of normal multivariate distribution. To evaluate the adjustment of the confirmatory models, a variety of ratings were used, just as is recommended in the literature. The comparative fit index (CFI) and Tucker-Lewis index (TLI) were utilized for those that indicate 0.9 or more, which is considered the minimum to accept the model, root mean square error of approximation (RMSEA) with values less than 0.08 are considered acceptable and 0.06 considered optimal, in addition to the chi-squared test (Hu & Bentler, 1995).

Finally, an ANOVA was carried out to test the differentiated validity of the BENCI between the different ages.

To prevent Type I errors with multiple comparison, Bonferroni adjustment was obtained for the *p* value. According to this adjustment, the level of significance was established at .002.

RESULTS

Reliability: Test-Retest and Cronbach's Alpha

In Table 1, the intraclass correlation coefficients (ICC) for each of the battery tests and the Pearson correlation coefficients are presented. Forty-three children participated, and some outliers were eliminated in general (maximum of 2).

The intraclass correlation coefficients of the BENCI battery constituent tests fluctuate between .51 and .81. The coefficients are higher in the two components of the battery that measure Reasoning (ICC = .81) and Working Memory (ICC = .81), with a notable magnitude in both cases. The magnitude of the ICC values for Selective Attention (RT) (ICC = .51), Delayed Verbal Memory (ICC = .52), Immediate Verbal Memory (ICC = .58), and Visual Memory of Recognition (ICC = .53) are only moderate. In the case of the scales for Verbal Memory of Recognition (ICC = .03) and Planning (ICC =–.23), the values reveal that there is no correlation. On the other hand, the Pearson correlation coefficients show almost identical values to the ICC for each one of the battery components, which occurred because the measurements and the variances of the two series are the same (Fleiss & Cohen, 1973). This situation rules out the presence of biases between the two measurements, specifically the association between the effects of memory and learning.

We also calculated the Cronbach's alpha for applicable tests. Results indicate excellent internal consistency on five of the tests (Sustained Attention CPT [RT]: α = .96; Go/No-Go [RT]: α = .90; Comprehension of Figures [CA]: α = .92; Selective

Table 1 Reliability test-retest of the BENCI battery.

Test	ICC	CI 95%	Pearson correlation
Visuomotor Coordination (RT)	.75	0.61–0.84	.74
Alternate Visuomotor Coordination (RT)	.73	0.59–0.83	.73
Sustained Attention CPT (CA)	.69	0.53–0.81	.68
Sustained Attention CPT (RT)	.68	0.51–0.80	.67
Immediate Verbal Memory (CA)	.58	0.33–0.70	.57
Delayed Verbal Memory (CA)	.52	0.24–0.71	.50
Verbal Recognition Memory (CA)	.03	−0.28–0.32	.03
Immediate Visual Memory (CA)	.74	0.59–0.83	.73
Delayed Visual Memory (CA)	.78	0.65–0.86	.77
Visual Recognition Memory (CA)	.53	0.31–0.69	.52
Comprehension of Figures (CA)	.75	0.47–0.78	.75
Working Memory (CA)	.81	0.70–0.88	.81
Reasoning (CA)	.81	0.70–0.88	.81
Semantic Fluency (CA)	.77	0.64–0.86	.76
Go/No-Go (*d*-prime)	.79	0.62–0.88	.77
Selective Attention (CA)	.77	0.64–0.86	.78
Selective Attention (RT)	.51	0.29–0.68	.53
Planning (# Visited Attractions)	−.23	−0.46–0.02	−.23

Note. ICC= Intraclass Correlation Coefficient (two-dimensional ANOVA of the random effects, definition of absolute agreement); CI = confidence interval; RT = reaction time; CA = correct answer.

Attention [CA]: α = 0.96; Reasoning [CA]: α = .93) and good internal consistency on the other tests (Sustained Attention CPT [CA]: α = .80; Immediate Verbal Memory [CA]: α = .84; Immediate Visual Memory [CA]: α = 0.80; Selective Attention [RT]: α = .88; Go/No-Go [precision]: α = .80; Reasoning [RT]: α = .87; planning: α = .77; Reaction time: α = .82).

Construct Validity

With regard to the battery that measures different factors based on the measurement criteria of all the important power factors involved in neurodevelopment, a theoretical model does not exist to explain the structure of the entire battery. Nevertheless, for the decision on which tests of executive function were to be included, the Diamond model was followed. This was done to examine the theoretical model of five factors of executive function shown in the BENCI by Diamond (2013). An analysis was carried out utilizing an estimation of highest likelihood and a covariance matrix between the items such as input for the factor analysis. The adjustment of the model was evaluated with a combination of absolute and relative adjustment indexes that included, within the absolutes, a *p* value, associated with the chi-squared statistic, which tests the null model in front of the hypothesized model; the CFI, which indicated the relative amount of variance and covariance reproduced by the specific model compared with the saturated model. Its value should be the same or more than .90 for the adjustment model and RMSEA to be considered minimally acceptable, which minimizes the problem derived from the sample size and in the values of .06 or lower indicating an excellent adjustment (Hu & Bentler, 1999).

The confirmatory factorial analysis results indicate that five factors exist: Inhibition, Flexibility, Fluency, Reasoning, and Working Memory. This five-factor model demonstrates that the chi-square analysis was not statistically significant. The CMIN/DF ratio (chi square/degrees of freedom) of the model was 1.240 which indicates a good adjustment. With respect to the global adjustment indicators, the CFI = 0.939 and the TLI = 0.924 demonstrate that it is an acceptable model, and by the RMSEA = 0.042, also indicates an excellent adjustment. These results suggest that it is a strong adjustment model (see Figure 1).

Validity in Detecting Neurodevelopmental Changes

Because the BENCI is a battery used to evaluate the neurodevelopment of children, we have examined its detection of neuropsychological development based on age. The results demonstrate significant differences between children at the ages of 7, 9, and 11 years on the majority of the neuropsychological variables (see Table 2).

DISCUSSION

The objective of this work was to study the psychometric properties of the first Arabic version of the computerized battery for the neuropsychological evaluation of children (BENCI). The results demonstrated good validity and reliability of the battery in Arabic children living in Morocco. It also showed that the BENCI battery is capable of differentiating children by their age group.

In general, the reliability of the BENCI battery has been good and similar to that of other neuropsychological studies (Gualtieri & Johnson, 2006; Schatz & Ferris, 2013). Ten of the 12 tests of the battery have obtained an excellent or good Intraclass Correlation Coefficient. On

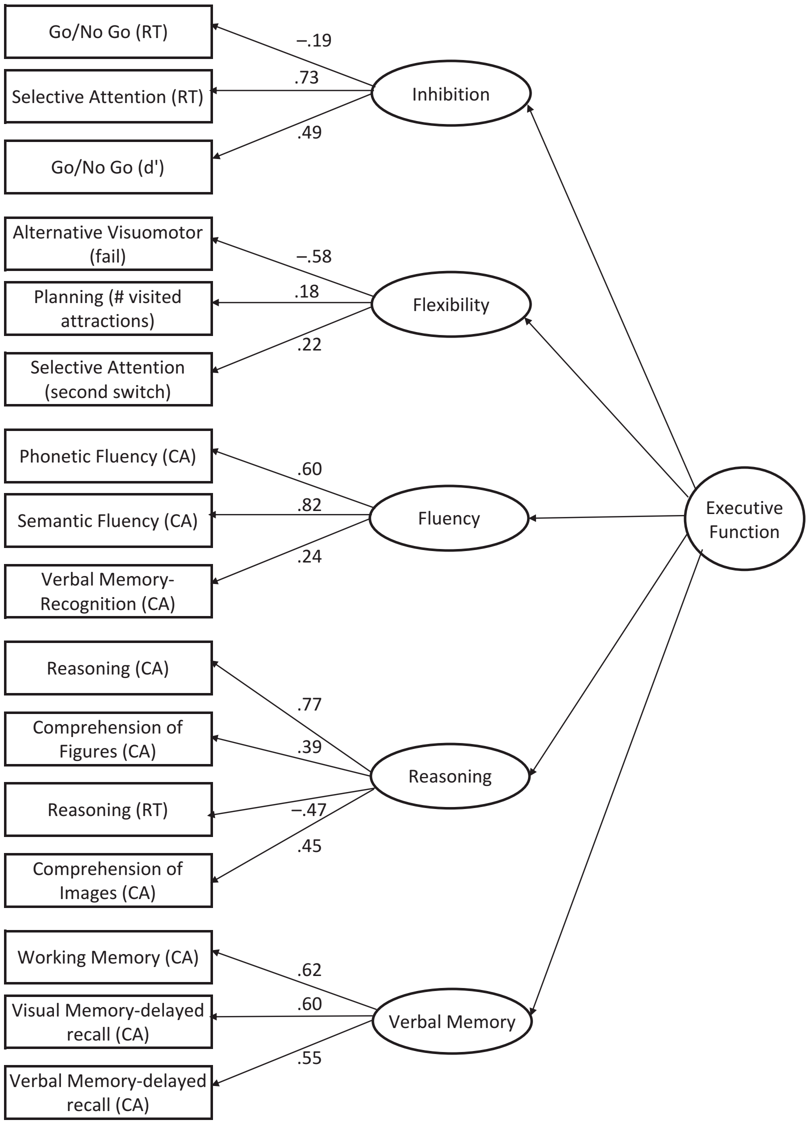

Figure 1 Structural diagram model. Model with five factors in the first order and three or four factors correlated in the second order.
Note. RT = Reaction time; CA = Correct answers.

the verbal and visual memory tests (except for the recognition essay), the ICC was acceptable. This result is in accordance with other studies, where the reliability of the test-retest was similar on these kinds of tests (Gualtieri & Johnson, 2006; Serra-Mayoral & Peña-Casanova, 2006). On the planning test, the ICC was very low. Different studies show a low test-retest reliability for planning tests in children (e.g., Bishop, Aamodt-Leeper, Creswell, McGurk, & Skuse, 2001) and, when these tests are computerized, the reliability is even lower (Luciana & Nelson, 1998). The novelty, difficulty, and computerized presentation of these tests can affect the results (Lowe & Rabbitt, 1998; Luciana & Nelson, 1998).

Table 2 Age differences on the BENCI tests.

Tests	7 years (n = 51) *M* (*SD*)	9 years (n = 63) *M* (*SD*)	11 years (n = 84) *M* (*SD*)	*F*	*p*	post hoc
Comprehension of Figures (CA)	8.59 (1.02)	8.95 (1.16)	9.48 (0.84)	13.33	**<.001**	11 > (7 = 9)
Comprehension of Images (CA)	8.39 (1.22)	8.83 (1)	9.15 (1)	6.48	**.002**	11> 7; 7 = 9; 11 = 9
CPT (RT_1)	444.8 (97.95)	387.10 (92.80)	377.89 (69.50)	10.56	**<.001**	(11 = 9) > 7
CPT (RT_2)	431.47 (107.99)	389 (89.29)	379.68 (64.18)	6.16	**.003**	(11 = 9) > 7
Immediate Verbal Memory (CA)	4.63 (0.81)	5.36 (0.87)	5.83 (0.95)	27.29	**<.001**	11 > 9 > 7
Delayed Verbal Memory (CA)	4.31 (1.31)	5.30 (0.95)	5.75 (1.40)	20.34	**<.001**	(11 = 9) > 7
Working Memory (CA)	4.47 (1.30)	6.25 (1.67)	6.6 (1.48)	32.72	**<.001**	(11 = 9) > 7
Phonetic Fluency (CA)	3.90 (1.98)	5.44 (2.41)	6.25 (2.96)	11.27	**<.001**	(11 = 9) > 7
Semantic Fluency (CA)	7.7 (2.75)	10.92 (3.14)	13.31 (3.33)	48.88	**<.001**	11 > 9 > 7
Immediate Visual Memory (CA)	5.94 (2.15)	7.41 (1.95)	9.33 (2.33)	38.13	**<.001**	11 > 9 > 7
Delayed Visual Memory (CA)	4.35 (1.96)	5.95 (2)	7.63 (2.16)	39.18	**<.001**	11 > 9 > 7
Go/No-Go (*d*-prime)	2.7 (1.25)	3.38 (0.85)	3.67 (0.62)	19.04	**<.001**	(11 = 9) > 7
Go/No-Go (RT)	605.51 (117.71)	574.83 (77.11)	544.96 (78.99)	7.31	**.001**	7 = 9 11 >7 11 = 9
Verbal Memory—Recognition (CA)	10.84 (1.97)	11.46 (1.04)	11.37 (1.73)	2.15	.12	
Planning (# visited Attractions)	18.33 (2.55)	20.08 (3.40)	20.31 (3.60)	6.23	**.002**	(11 = 9) > 7
Reasoning (RT)	22018.23 (9,874.74)	16702.67 (7,438.99)	13809.68 (5,029.58)	20.05	**<.001**	11 > 9 > 7
Reasoning Correct Answers	12.65 (4.42)	17.05 (3.90)	18.83 (3.09)	44.16	**<.001**	11 > 9 > 7
Selective Attention Resume	73.69 (9.40)	83.24 (9.33)	86.95 (8.81)	33.93	**<.001**	11 > 9 > 7
RT average	519.13 (138.37)	445.79 (103.46)	431.50 (119.33)	7.74	**.001**	(11 = 9) > 7
RT median	470.36 8147.53)	400.08 (103.58)	388.96 (109.96)	6.75	**.002**	(11 = 9) > 7
Visuomotor (completion time)	139,188.16 (56,937.30)	90,916.77 (34,758.77)	67,750.54 (17,890.48)	60.14	**<.001**	11 > 9 > 7
Alternate Visuomotor (completion time)	146,143.16 (60,636.66)	88,373.52 (34,972.37)	69,576.11 (25,749.86)	59.00	**<.001**	11 > 9 > 7

Note. CPT = Continuous Performance Test; RT = Reaction time (milliseconds); CA = Correct answer; *M* = Mean; *SD* = Standard deviation; boldface = $p < 0.01$.

With regard to the validity of the battery, our results showed a good validity construct. The confirmatory factorial analysis indicated a model of five factors: Inhibition, Flexibility, Fluency, Reasoning, and Working Memory. This model is acceptable and has a good adjustment model by Diamond (2013). Planning was the only component to not be adjusted to the model. This could be because planning is the last component of executive function to develop in adolescents around 15–18 years. In this case, the sample was made up of children under the age of 12. It is probable that this function was still not developed in the children of this sample.

Moreover, the BENCI battery has demonstrated the capacity to differentiate between children by their age group. The older children score better than the youngest ones on all of the battery tests. These results are similar to a study conducted by Waber et al. (2007), which finds that age has a clear effect on the neuropsychological tests performance of children between 6 and 10 years of age. These results support good validity of the BENCI for detecting changes in neurodevelopment.

When comparing our results to the Spanish norms, Moroccan children had higher scores than Ecuadorian children on the Verbal Memory and Spatial Stroop tests. On the other hand, Ecuadorian children outperformed the Moroccan children in Semantic Fluency and Visual Memory. These results could be explained by cultural differences between the two groups. In other words, differences in language, educational systems, attitudes towards time, and cultural values affect neuropsychological performance (Ardila, 2005; Nell, 2000; Ostrosky-Solís, Ramírez, Lozano, Picasso, & Vélez, 2004; Puente et al., 2013). However, there were no differences between Moroccan and Ecuadorian children on seven of the tests. This may demonstrate that these tests are valid for each cultural group.

According to our findings, the BENCI battery is an important contribution to the field of the neuropsychological evaluation of children in the Arabic population. This battery evaluates the neurodevelopment of children utilizing computerized tests that cover different neuropsychological domains. In societies where tests are scarce (e.g., the Arabic world), creating and validating psychological tests is ethically and professionally vital. To our knowledge, this study is the first to create and validate a computerized comprehensive battery to evaluate the neurodevelopment of Arabic children. The BENCI battery is a valid and reliable instrument for evaluating the neurodevelopment of Arabic children and responds to the need for creating neuropsychological tests specific to the Arabic world (Abdul Razzak, 2013; Fasfous et al., 2013; Sobeh & Spijkers, 2013).

Finally, our study has several limitations. The IQ of the children was not controlled for in this study. Nevertheless, because the objective of this study was to test the validity and reliability of an instrument, we believe that the IQ does not affect the statistical properties of the battery. In any case, in the Moroccan educational system, children with developmental problems attend special programs and all the children of this study were selected from the ordinary schools. On the other hand, the sample was chosen from only one Arabic country, Morocco, and only one city. Also, dialect differences could enhance differences among individuals from Arab countries in the neuropsychological performances, but no evidence exists about these distinctions. Nevertheless, this study is preliminary and future studies should test the validity and reliability of the BENCI in more Arabic countries, using larger and more representative samples, and utilizing the BENCI to compare healthy children to children with brain trauma.

In conclusion, the BENCI is a neuropsychological battery that has demonstrated good validity and reliability in Arabic children. This battery can be of great use to both the research and clinical areas of Arabic countries and/or in assistance to Arabic immigrants that live outside of their native country. To facilitate this, BENCI is ready to be used and can be obtained by contacting the first author.

Original manuscript received August 22, 2013
Revised manuscript accepted February 16, 2014
First published online March 28, 2014

REFERENCES

Abdul Razzak, R. (2013). A preliminary study on the trail-making test in Arabic–English bilingual young adults. *Applied Neuropsychology: Adult, 20*(1), 53–60. doi:10.1080/09084282.2012.670163

Abu Zaydeh, H., Zalina, I., Wan, A., & Aljeesh, Y. (2012). Visual spatial and executive functions disorders among Palestinian children living under chronic stress in Gaza. *Journal of Natural and Engineering Studies, 20*(2), 55–71. Retrieved from http://www.iugaza.edu.ps/ar/periodical/

Alansari, B., & Baroun, K. (2004). Gender and cultural performance differences on the Stroop Color and Word test: A comparative study. *Social Behavior and Personality, 32*(3), 235–245.

American Educational Research Association, American Psychological Association, & National council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17*(1), 143–150. doi:10.1080/13803399508406589

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review, 15*(4), 185–195. doi:10.1007/s11065-005-9180-y

Armstrong, S. K., White, E., & Saracci, R. (1992). *Principles of exposure measurement in epidemiology.* Oxford: Oxford University Press.

Bishop, D. V. M., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001). Individual differences in cognitive planning on the tower of Hanoi task: Neuropsychological maturity or measurement error?. *Journal of Child Psychology and Psychiatry, 42*(4), 551–556. doi:10.1111/1469-7610.00749

Chefchaouen Townhall “official web page” (2013). *Socio-demographic information about people living in the city of Chefchouen*. Retrieved from http://www.chaouen.ma

Cruz-Quintana, F., Pérez-García, M., Roldan-Vílchez, L. M., Fernández López, A., & Pérez-Marfil, M. N. (2013). *Manual de la Batería de Evaluación Neuropsicológica Infantil (BENCI)*. Granada: Ediciones CIDER S.C.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California verbal learning test: Adult version*. Manual. San Antonio, TX: Psychological Corporation.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. doi:10.1146/annurev-psych-113011-143750

El Hioui, M., Azzaoui, F. Z., Touhami Ahami, A. O., Rusinek, S., & Aboussaleh, Y. (2012). Iron deficiency and cognitive function among Moroccan school children. *Nutritional Therapy Metabolism, 30*(2), 84–89. doi:10.5301/NTM.2012.9675

Fasfous, A. F., Hidalgo-Ruzzante, N. A., Vilar-Lopez, R., Catena-Martinez, A., & Perez-Garcia, M. (2013). Cultural differences in neuropsychological abilities required to perform intelligence tasks. *Archives of Clinical Neuropsychology, 28*(8), 784–790. doi:10.1093/arclin/act074

Fasfous, A. F., Puente, A. E., Peralata-Ramírez, M., & Pérez-García, M.. (2014). *Neuropsychology in the Arab World during the last 20 years*. Manuscript in preparation.

Fasfous, A. F., Puente, A. E., Perez-Marfil, M. N., Cruz-Quintana, F., Peralta-Ramirez, I., & Perez-Garcia, M. (2013). Is the color trails culture free? *Archives of Clinical Neuropsychology, 28*(7), 743–749. doi:10.1093/arclin/act062

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*, 613–619. doi:10.1177/001316447303300309

García Bermúdez, O., Cruz-Quintana, F., Sosa, M. A., de la Cruz, J., Mañas, M., & Pérez-García, M. (2012). Alteraciones neuropsicológicas y emocionales en niños prematuros de muy bajo peso al nacer. *Revista Argentina de Ciencias del Comportamiento, 4*(2), 3–10.

Golden, C. J. (2001). *STROOP. Test de colores y palabras*. Madrid: TEA.

Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52*(10), 1115–1124. doi:10.1037/0003-066X.52.10.1115

Gualtieri, C. T., & Johnson, L. G. (2006). Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology, 21*(7), 623–643. doi:10.1016/j.acn.2006.05.007

Hu, L., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (pp. 76–99). London: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. doi:10.1080/10705519909540118

International Test Commission (2010). *International Test Commission Guidelines for translating and adapting tests*. Retrieved from http://www.intestcom.org

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. doi:10.2307/2529310

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed). New York, NY: Oxford University Press.

Llorente, A. M., Williams, J., Satz, P., & D'Elia, L. F. (2003). *Children's Color Trails Test (CCTT)*. Odessa, FL: PAR.

Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia, 36*(9*),* 915–923. Retrived from http://dx.doi.org/10.1016/S0028-3932(98)00036-0

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally guided working memory systems in four- to eight-year-old children. *Neuropsychologia, 36*(3*),* 273–293. Retrived from http://dx.doi.org/10.1016/S0028-3932(97)00109-7

Mirkin, B. (2010). *Population levels, trends and policies in the Arab region: Challenges and Opportunities*. United Nations Development Programme—Arab Human Development Report (Research Paper Series). Retrieved from http://www.arab-hdr.org

Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practice*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Ostrosky-Solís, F., Ramírez, M., Lozano, A., Picasso, H., & Vélez, A. (2004). Culture or education? Neuropsychological test performance of a Maya indigenous population. *International Journal of Psychology, 39*(1), 36–46. doi:10.1080/00207590344000277

Pamos, A., Alejandre, M. A., & Benedet, M. J. (2007). *TAVECI. Test de Aprendizaje Verbal España-Complutense Infantil*. Madrid: TEA.

Puente, A. E., Perez-Garcia, M., Vilar-Lopez, R., Hidalgo-Ruzzante, N., & Fasfous, A. F. (2013). Neuropsychological assessment of culturally and educationally dissimilar individuals. In F. Paniagua & A. M. Yamada, (Eds.), *Handbook of multicultural mental health: Assessment and treatment of diverse population* (2nd ed., pp. 225–242). New York, NY: Elsevier.

Raven, J. C. (1977). *Matrices Progresivas en Color*. Madrid: TEA.

Rey, A. (1964). *L 'examen clinique en psychologie* [Clinical tests in psychology]. Paris: Presses Universitaires de France.

Schatz, P., & Ferris, C. S. (2013). One-month test–retest reliability of the ImPACT test battery. *Archives of Clinical Neuropsychology, 28*(5*),* 499–504. doi:10.1093/arclin/act034

Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11*, 193–205. doi:10.1023/A:1015291021312

Serra-Mayoral, A., & Peña-Casanova, J. (2006). Fiabilidad test-retest e interevaluador del Test Barcelona. *Neuropsicología, 21*(6), 277–281.

Shonkoff, J. P. (2010). Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Development, 81*(1), 357–367. doi:10.1111/j.1467-8624.2009.01399.x

Shonkoff, J. P., Richter, L., van der Gaag, J., & Bhutta, Z. A. (2012). An integrated scientific framework for child survival and early childhood development. *Pediatrics, 129*(2), e460–e472. doi:10.1542/peds.2011-0366

Sobeh, J., & Spijkers, W. (2013). Development of neuropsychological functions of attention in two cultures: A cross-cultural study of attentional performances of Syrian and German children of pre-school and school age. *European Journal of Developmental Psychology, 10*(3), 318–336. doi:10.1080/17405629.2012.674761

Waber, D., De Moor, C., Forbes, P., Almli, C., Botteron, K., Leonard, G., … Group, BDC. (2007). The NIH MRI study of normal brain development: Performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. *Journal of the International Neuropsychological Society, 13*, 729–746. doi:10.1017/S1355617707070841

Woodcock, R. V., & Sandoval, A. F. (1996). *Batería Woodcock-Muñoz: Pruebas de habilidad cognitiva-revisada*. Itasca, IL: Riverside.