Methodology for the development of normative data for ten Spanish-language neuropsychological tests in eleven Latin American countries

Joan Guàrdia-Olmos, Maribel Peró-Cebollero, Diego Rivera and Juan Carlos Arango-Lasprilla

Abstract

BACKGROUND: Within the field of neuropsychology, there is a significant lack of normative data for individuals in Latin America.

OBJECTIVE: To describe the methodology utilized to obtain the data and create norms for 10 Spanish-language neuropsychological tests administered in 11 Latin-American countries in a sample of 3,977 healthy individuals between the ages 18 and 90.

METHOD: The same data manipulation process was applied to the data collected (regardless of the scale or country) using a regression-based procedure that takes into account sex, age, and educational influences on neuropsychological test scores.

CONCLUSIONS: Following this procedure, we were able to generate age, education, and sex (if relevant) based norms for each test in each of the 11 countries studied. These norms are presented in the 10 articles that comprise this special issue.

Keywords: Psychometric, norms, Latin America, neuropsychological tests

1. Introduction

Most psychological and neuropsychological tests are built using the Classical Test Theory (CTT), from this perspective norms are created using simple procedures like transformation in percentile scales, or T and D scores (Abad, Olea, Ponsoda, & García, 2011; Crawford, 2003). Within the field of neuropsychology, there is a significant lack of norms for Spanish-speakers in Latin America. This article presents the methodology employed to standardize various neuropsychological tests administered in 11 Latin American countries.

There is extensive literature describing which demographic variables (e.g., sex, age, socio-economic status, and years of education) must be taken into account to generate appropriate norms for neuropsychological tests. It is widely accepted that the systematic error generated by both chronological age and education needs must be controlled for (Salthouse, 2000, 2001, 2009; Salthouse, Atkinson, & Berish, 2003). Therefore, there are several ways to control for age and education effects:
1) A method that involves the adjustment of control systems for error correction, based on the generation of normative data for each reference group. In this procedure, it would be necessary to use the stratum system to identify normative groups based on classic standardization. Though this option is easy, it generates an enormous quantity of normative tables that would turn the final product into a measureless and useless contribution from the point of view of applied clinical research (Brooks, Strauss, Sherman, Iverson, & Slick, 2009; Kim et al., 2014; La Cour & Andersen, 2006; La Paglia et al., 2014; Mungas, Reed, Crane, Haan, & González, 2004; Peintinger & Klimenmann, 2013; Salvadori, Poggesi, Pracucci, Inzitari, & Pantoni, 2015),

2) The score correction obtained from the use of partial regression coefficients from a linear model (Blesa et al., 2001; Villaseñor, Guárdia, Jiménez, Rizo, & Peri, 2010), and 3) The use of regression coefficients combined with the standard deviation of the residual to obtain normative data based on a percentile scale (Van der Elst et al., 2006, 2011, 2012). Though this procedure does not radically solve the issues of asymmetric distributions, it does make it possible to take them into account, as the predictions of scale setting are based on the establishment of confidence intervals close to the average values (Russell, Russell, & Hill, 2005).

2. Method

2.1. Participants

The present study was conducted with a sample of 5,402 healthy individuals from 20 cities in 12 Spanish-speaking countries in Latin America (Argentina, Bolivia, Chile, Colombia, Cuba, El Salvador, Guatemala, Honduras, Mexico, Paraguay, Peru, and Puerto Rico). The results obtained from 1,425 Colombians have been published elsewhere (Arango & Rivera, 2015). Therefore, this article will present the procedures used to generate normative data from 3,977 healthy individuals from 14 cities for the remaining 11 countries.

The sample was designed taking into account the literacy level, the percentage of people with primary, secondary and tertiary studies, and the age distribution for each country. Thus, an empirical quota sampling was used, in order to have a statistical precision between 0.063 to 0.049 with a 95% confidence level estimated for each country. Thus, an empirical quota sampling was used, in order to have a statistical precision between 0.063 to 0.049 with a 95% confidence level estimated for each country. Therefore, this article will present the procedures used to generate normative data from 11 Spanish-speaking countries from Latin America.
Table 1
Summary of the sampling designed and final sample obtained

<table>
<thead>
<tr>
<th>Country</th>
<th>Age in years</th>
<th>Designed sample</th>
<th>Final sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;6</td>
<td>6 to 12</td>
<td>&gt;12</td>
</tr>
<tr>
<td>Argentina</td>
<td>56</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Bolivia</td>
<td>56</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Chile</td>
<td>56</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Cuba</td>
<td>56</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>El Salvador</td>
<td>56</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Guatemala</td>
<td>56</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Mexico</td>
<td>56</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Peru</td>
<td>56</td>
<td>50</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 2
Sample distribution by age, education and gender

<table>
<thead>
<tr>
<th>Country</th>
<th>Age in years</th>
<th>Education</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>1 to 12</td>
<td>&gt;12</td>
</tr>
<tr>
<td></td>
<td>n (%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>Argentina</td>
<td>320</td>
<td>45.3 (19.5)</td>
<td>148 (46.3%)</td>
</tr>
<tr>
<td>Bolivia</td>
<td>274</td>
<td>55.8 (22.0)</td>
<td>226 (82.9%)</td>
</tr>
<tr>
<td>Chile</td>
<td>320</td>
<td>55.1 (19.6)</td>
<td>241 (75.3%)</td>
</tr>
<tr>
<td>Cuba</td>
<td>363</td>
<td>56.1 (19.7)</td>
<td>234 (76.5%)</td>
</tr>
<tr>
<td>El Salvador</td>
<td>257</td>
<td>56.0 (20.7)</td>
<td>203 (79.0%)</td>
</tr>
<tr>
<td>Guatemala</td>
<td>214</td>
<td>53.2 (17.4)</td>
<td>133 (62.1%)</td>
</tr>
<tr>
<td>Honduras</td>
<td>184</td>
<td>48.6 (18.8)</td>
<td>140 (76.1%)</td>
</tr>
<tr>
<td>Mexico</td>
<td>1300</td>
<td>52.5 (20.5)</td>
<td>1005 (77.3%)</td>
</tr>
<tr>
<td>Paraguay</td>
<td>263</td>
<td>53.0 (14.8)</td>
<td>216 (82.1%)</td>
</tr>
<tr>
<td>Peru</td>
<td>245</td>
<td>43.4 (20.6)</td>
<td>87 (35.5%)</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>294</td>
<td>50.9 (14.8)</td>
<td>140 (45.4%)</td>
</tr>
</tbody>
</table>

Additional information:
- native language, d) had completed at least one year of formal education, e) be able to read and write at the moment of evaluation. f) scored ≥23 on the Mini-Mental State Examination (MMSE, Folstein, Folstein, & McHugh, 1975; Ostrosky-Soló, López-Arango, & Ardila, 2000; Villaseñor-Cabrera, Guardia-Olmos, Jiménez-Maldonado, Rizo-Curriel, & Peric-Cebollero, 2010), g) scored ≤4 on the Patient
Health Questionnaire–9 (PHQ-9, Kroenke, Spitzer, & Williams, 2001), and h) scored ≥90 on the Barthel Index (Mahoney & Barthel, 1965).

The exclusion criteria were: a) having a personal history of central nervous disease (stroke, epilepsy, multiple sclerosis, brain tumour, severe head trauma, etc.); b) having history of alcohol abuse or other psychotropic substances; c) having an active systemic disease or uncontrolled disease associated with cognitive impairment (diabetes mellitus, hypothyroidism, B12 vitamin deficit); d) having a history of psychiatric illness (major depression, bipolar disease, psychosis, etc.); e) having severe sensory deficits (vision and/or auditory loss) that could affect the administration of the test or the participants’ performance in these test; f) Use of psychiatric or other drugs that could affect one’s cognitive performance; and g) taking medications for chronic pain (e.g. Monoamine Oxidase Inhibitors – MAOI).

2.2. Instruments

A sociodemographic questionnaire was created for this project in order to collect data on age, years of education, sex, laterality (right, left handed, ambidextrous), residence zone (rural, urban), race (white, black, mestizo, indigenous), employment status (employed, unemployed, student, retired, housekeeping), and marital status (single, married, cohabited, separated, divorced, widower). Then, all of the individuals were administered a comprehensive neuropsychological evaluation that was composed of the 10 neuropsychological tests: 1. Rey-Osterrieth Complex Figure test (ROCF; Rey, 2009). 2. Stroop colour and words test (Golden, 2010). 3. Modified Wisconsin card sorting test (M-WCST; Nelson, 1976; Schretlen, 2010). 4. Trial making test (TMT A-B; Reitan & Wolfson, 1985). 5. Brief test of attention (BTA; Schretlen, 1997). 6. Phonological and semantical verbal fluency test. 7. Boston naming test (Goodglas, Kaplan, & Barresi, 2005). 8. Symbol digit modalities test (SDMT; Smith, 2002). 9. Hopkins verbal learning test - Revised (HVLT-R; Benedict, Schretlen, Groninger, & Brandt, 1998). 10. Test of memory malingering (TOMM; Tombaugh, 2011).

2.3. Procedure

2.3.1. Training and administration procedure of the battery

The present study began November 1, 2012 with the participation of 15 institutions. Subsequently, the proposal that included the methodology and ethics was drafted and later delivered to the ethics committee of the University of Deusto (Bilbao, Spain). After approval, the publishing manuals, answer sheets and materials (booklets and encouragement cards) for each of the neuropsychological tests were bought. Spanish-language versions of instructions booklets were available for all instruments except for M-WCST and BTA. Spanish administration and instruction manuals for these instruments were created in collaboration with the publishers.

Likewise, tools and visual aids were created to achieve standard management process of the battery. These tools consisted of: a) a randomized list to determine the order of administration of the test for each participant in order to avoid order bias and cognitive conditioning. To do so, the function $f(x) = \text{RAND}(x)$ was used in Microsoft Excel and was set to take into account the interaction of the language test and verbal memory test; b) a framework for decision-making in the evaluation process; c) a template in Microsoft Excel for entering information to limit bias input information. The template was designed using the configuration options: data validation = Customized (numeric variables), dropdown lists (categorical variables) and setting formats; d) examples showing the most frequent errors in the administration and scoring of each test, and e) a virtual folder with a security key for each city, administered by the study coordinator to track data entry.

Moreover, each coordinator selected a group of 6–12 undergraduates and/or graduate students, with which the coordinator reviewed the instructions and application and qualification of the tests. Once the group was ready, an online virtual training was performed through telemedicine platform VSee. During the two-hour long training, the administration and scoring of the tests were reviewed and any doubts were resolved.

A pilot test with 40 protocols (these were excluded from the analysis of normative data) was conducted by analyzing data from the first two cases collected in each of the 20 centers in order to determine the adequate functioning for collection of information with the proposed design.

Data collection began March 2013 and ended August 2014. The protocol was administered in a single day and lasted about 70 minutes. Before starting the battery...
administration, participants had to sign the informed consent.

Once the database was consolidated, we reviewed the data distribution of frequencies, comparing the values of various statistics and graphs, in order to check for correct processing and characteristics of distributions for each of the variables analyzed. This certified the database was generated correctly and the properties of the observed variables were known.

2.3.2. Normative procedure

An independent t-test by country was run for each neuropsychological measure to first determine whether there were any significant differences in test scores based on participant sex within each country. Since the sample size is very large, significant differences for gender in neuropsychological scores were needed with a medium or high effect size ($r > 0.3$, Cohen, 1988) in order to consider that differences in the scores between men and women are relevant. The effect size was obtained using the following expression:

$$r = \sqrt{\frac{t^2}{t^2 + v}}$$

when relevant, country sample sizes were stratified by sex when significant differences were found with medium or high effect sizes.

The linear regression model was built using the raw (direct/unconverted) score in the scale being normed as the criterion variable. Participant age was included as a predictor variable, and years of education (dichotomized in two groups: 1 to 12 years coded as 0 and >12 years coded as 1) was also considered as a predictor variable for those cases where the differences in the raw scores between men and women met the above-mentioned criteria. All the regressions were run differentiating by country. If any predictive variable was not statistically significant, the linear regression analysis was repeated eliminating that predictive variable. For all linear regression models, the following were assessed: collinearity between the predictive variables, value of variance inflation factor (VIF) near to 1, the residuals normality using the Q-Q plot and the residuals histogram, the homoscedasticity as in the scatter plot of the standardized residuals and the predicted values, and finally, the existence of influential values by Cook’s distance.

In order to obtain the exact percentile for an exact score, the procedure used by Van der Elst et al. (2006, 2011, and 2012) was followed. The steps to obtain the exact percentile were:

1. Obtain the predicted value using the regression equation obtained in the country of origin of the person where the original $y_i$ was the score of each neuropsychological test:

$$\hat{y}_i = b_0 + b_1 \cdot \text{Age}_i + b_2 \cdot \text{Educational level}_i + b_3 \cdot \text{Sex}$$

2. Obtain the residual value:

$$e_i = y_i - \hat{y}_i$$

3. Standardize the residual obtained (transform to z score). To do this, it is necessary to divide the residual value by the standard deviation for the residual obtained in the fitted regression model.

$$z = \frac{e_i}{SD_{\varepsilon}}$$

4. Search the exact probability associated with the z value using accumulated probability of the standardized normal distribution.

It is possible that some clinicians consider this procedure slow and expensive in time. For this reason, in the papers that the different normative data were presented for each scale, we present also tables with the approximated percentile. To do this, following the procedure used by Van der Elst et al. (2006, 2011, 2012), we used class mark age value starting with 20 years ± 2 years until a class mark of 80 years old. For each class mark age and educational level (1: more than 12 years or 0:1 to 12 years) we apply the procedure explained in order to obtain the percentile. In this case we do not have the percentile for each age in years, we have the approximation for intervals of 5 years instead.

3. Limitations and future directions

Among the limitations encountered in the present study, there was an observed low symmetry in sample sizes between males and females for certain countries (e.g., Argentina). Additionally, discrepancy between designed and final sample was observed in certain countries more so than in others due to issues beyond researchers’ control. For instance, in the case of Cuba, where education is obligatory and accessible to the population, it was extremely difficult to find participants with fewer than six years of education. Furthermore, the study used healthy participants. This resulted in floor and ceiling effects on some of the neuropsychological measures used. Additional limitations include lack of clinical sample(s) in the normative data, as well as lack of adequate representation of indigenous populations.
from Latin America. Among inclusion criteria for this study was the requirement that participants must be able to read. Therefore illiterate individuals were not part of the final sample of this study. Directions for future work in this line of research may include creating normative data for illiterate individuals, as well as individuals with varying neurological disorders (e.g., dementias, TBI, multiple sclerosis, etc.) from Latin America.

4. Conclusions
The procedure used by Van der Elst et al. (2006, 2011, 2012) allows us to propose a convenient procedure to obtain sex, age, and education norms for 10 neuropsychological tests without the necessity to work with large samples. This procedure is, in fact, a simple procedure useful in clinical practice. Clinicians have two possibilities to obtain the percentile. The first one is to obtain the exact percentile using the procedure described and the information of the regression coefficients and the residual standard error provided in the different articles of each test. They also have tables corresponding to the standardized normal distribution or software that provide the cumulative probability to a z score in order to complete the process and obtain the exact percentile. The second one, for those clinicians who do not need to obtain the exact percentile, is to use the tables obtained for each test with the approximate percentile. These tables are very simple to use, and are similar to those tables that psychologists use in their regular practice. They only need to know the sex, age and educational level of the person assessed and, finally, the raw score. In this line of research may include creating normative data for illiterate individuals, as well as individuals with varying neurological disorders (e.g., dementias, TBI, multiple sclerosis, etc.) from Latin America.

Acknowledgments
The Grup de Recerca en Técniques Estadístiques Avançades Aplicades a la Psicologia (GTEAAP) members of the Generalitat de Catalunya’s 2014 SGR 326 Consolidated Research Group (GRC) provided methodological and statistical support for this study. They are funded by the PSEI2013-41400-P project of Ministerio de Economía y Competitividad of the Spanish Government.

Conflict of interest
None.

References
patients. "Persuasive Computing Paradigms for Mental Health" (pp. 23-32). Springer International Publishing.


