Using Monolingual Neuropsychological Test Norms with Bilingual Hispanic Americans: Application of an Individual Comparison Standard

Philip Gerard Gasquoine*, Cassandra Dayanira Gonzalez

Department of Psychology and Anthropology, University of Texas-Pan American, Edinburg, TX, USA

*Corresponding author at: Department of Psychology and Anthropology, University of Texas-Pan American, Edinburg, TX 78541, USA.
Tel.: +1-361-888-7999; fax: +1-956-665-3333.
E-mail address: drgdrg13@yahoo.com (P.G. Gasquoine).

Accepted 9 January 2012

Abstract

Conventional neuropsychological norms developed for monolinguals likely overestimate normal performance in bilinguals on language but not visual-perceptual format tests. This was studied by comparing neuropsychological false-positive rates using the 50th percentile of conventional norms and individual comparison standards (Picture Vocabulary or Matrix Reasoning scores) as estimates of preexisting neuropsychological skill level against the number expected from the normal distribution for a consecutive sample of 56 neurologically intact, bilingual, Hispanic Americans. Participants were tested in separate sessions in Spanish and English in the counterbalanced order on La Bateria Neuropsicológica and the original English language tests on which this battery was based. For language format measures, repeated-measures multivariate analysis of variance showed that individual estimates of preexisting skill level in English generated the mean number of false positives most approximate to that expected from the normal distribution, whereas the 50th percentile of conventional English language norms did the same for visual-perceptual format measures. When using conventional Spanish or English monolingual norms for language format neuropsychological measures with bilingual Hispanic Americans, individual estimates of preexisting skill level are recommended over the 50th percentile.

Keywords: Bilingual advantage; Neuropsychological norms; False positive; Deficit measurement; Preexisting neuropsychological skill estimate

Introduction

Norms developed for monolingual speakers in either Spanish or English may not give an accurate depiction of normal performance for bilingual Hispanic Americans (Gasquoine, Croyle, Cavazos-Gonzalez, & Sandoval, 2007; Roberts, Garcia, Desrochers, & Hernandez, 2002; Rosselli, Ardila, Navarrete, & Matute, 2010). This is because bilingualism has a general effect on cognition, the nature of which is much debated. One popular perception is that bilingualism has a positive effect on cognitive skills, known as “the bilingual advantage.” As an illustration, the American Speech-Language-Hearing Association (2010) listed the advantages of bilingualism as: being able to learn new words easily; breaking down words by sounds; being able to use information in new ways; putting words into categories; coming up with solutions to problems; and good listening skills.

Research supporting such advantage is equivocal. As an illustration, while Kaushanskaya and Marian (2009) found bilinguals outperformed monolinguals when learning non-words with imaginary translations, Bialystok, Majumder, and Martin (2003) found that bilingual children do not develop phonological awareness more easily than monolingual children. The main experimental paradigm that has demonstrated an advantage for bilinguals over monolinguals has involved tasks that require suppression of a previously correct response (“response inhibition”), such as: the Dimensional Change Card Sort Task, an analogue of the Wisconsin Card Sorting Test for young children (Bialystok, 1999); perseverative errors on the Wisconsin Card Sorting Task (Vega & Fernandez, 2011); theory of mind false-belief tasks for young children (Bialystok & Senman, 2004; Goetz, 2003); the Simon task, where participants must respond to the dimension of a stimulus, such as...
color, despite misleading cue information (Bialystok, Martin, & Viswanathan, 2005); the Posner Attentional Network Task, where participants must indicate the direction of an arrow accompanied by either congruent or incongruent cue information (Costa, Hernandez, & Sebastian-Galles, 2008); and the Stroop Color and Word Test interference condition (Bialystok, Craik, & Luk, 2008). The bilingual advantage has been found mainly with young children and may diminish over the lifespan (Bialystok et al., 2005).

In contrast, most studies that have used tasks that emphasize skills related to language processing have found bilinguals at a disadvantage compared with monolinguals all across the lifespan. These tasks include: receptive vocabulary (Bialystok, 2007; Portocarrero, Burright, & Donovich, 2007); tip-of-the-tongue retrieval failures (Gollan & Acenas, 2004); naming deficits in both speed and accuracy (Bialystok et al., 2008; Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Roberts et al., 2002); speed deficits in reading color names (Rosselli et al., 2002); and reduced scores on category fluency (Bialystok et al., 2008; Gollan, Montoya, & Werner, 2002; Portocarrero et al., 2007). In a questionable practice, some studies using language format tasks have claimed a bilingual advantage when bilinguals and monolinguals scored at equivalent levels. As an example, Moreno, Bialystok, Wodniecka, and Alain (2010) found equal performance for bilinguals and monolinguals on a linguistic judgment task (identifying sentences with conflicting semantic information that contained grammatical errors) evidence of a “bilingual advantage” (p. 576), as the bilinguals had previously scored below the monolinguals on picture vocabulary and in the judgment of grammatical and semantic errors.

Although not tested directly, the bilingual language processing disadvantage may impact Hispanic American response patterns on intelligence tests. The US census race/ethnic classification first incorporated Hispanic Americans as a separate grouping in 1970 and subsequent national normative intelligence test samples showed group mean Hispanic American scores were below those of White non-Hispanics on language format measures, but equivalent on visual-perceptual format measures. As an illustration, Hispanic American Wechsler Intelligence Scale for Children, 3rd edition (Wechsler, 1991) mean scores were about 0.5 SD below those of White non-Hispanics on Verbal IQ, whereas those on Performance IQ were comparable (Neisser et al., 1996; Puente & Salazar, 1998). Pinpointing the reason for this disparity has been complicated as test developers classify Hispanic Americans by ethnic self-report (mirroring the census), thereby ignoring potentially important individual differences in acculturational, bilingualism, and English language proficiency that are known to impact performance on cognitive tests (Gasquoine, 1999). A bilingual advantage in response inhibition would not positively impact intelligence test scores, as item content does not emphasize this aspect of cognition.

For bilingual Hispanic Americans, low scores on language format tasks occur with both Spanish and English languages of administration (Gasquoine et al., 2007; Gasquoine, Cavazos, Cantu, & Weimer, 2010; Weimer, Meza, & Gasquoine, In press), indicating that it is some aspect of bilingualism and not reduced Spanish or English language proficiency that is responsible. Demographic explanations (i.e., low socio-economic status [SES] or poor-quality education) for the low scores also appear unlikely. Mean scores from the Woodcock–Munoz Language Survey-Revised (WMLS-R; Woodcock, Munoz-Sandoval, Ruef, & Alvarado, 2005) Picture Vocabulary subtest for 103 bilingual Hispanic American 3–7-year olds from higher SES homes (mean yearly household income of $58,000 compared with the regional median of $30,000) in the Rio Grande Region of Texas were 80.3 in English and 89.7 in Spanish compared with the national \( M = 100; SD = 15 \) (Weimer et al., In press). These scores were similar to WMLS-R cluster scores obtained from a sample of bilingual Hispanic American adults (\( M = 14.4; SD = 2 \) years of education) from the same region who had means of 85.8 in English and 87.9 in Spanish (Gasquoine et al., 2007). Visual-perceptual processing skills, as assessed by the Wechsler Adult Intelligence Scale-III (WAIS-III: Wechsler, 1997a) Matrix Reasoning subtest (national \( M = 10; SD = 3 \)), were in the average range (11.3 in English and 10.7 in Spanish) for the latter group. Exactly what aspect of bilingualism is responsible for the low language processing scores remains unclear, although it has been hypothesized that the parallel activation of both languages (as occurs when bilinguals utilize language brain zones) causes inter-language interference thereby slowing processing time and increasing the possibility of errors, especially on difficult and/or time-dependent tasks (e.g., Bialystok et al., 2008).

Systematic negative or positive effects of individual difference variables like bilingualism on neuropsychological test scores impact the primary purpose of neuropsychological assessment, namely the identification of cognitive impairment via “deficit measurement” (Lezak, Howieson, & Loring, 2004). In this process, cognitive deficit is inferred whenever the difference between an obtained post-onset neuropsychological test score and a preexisting estimate is greater than a certain amount, such as 1 SD. A common approach to estimating the preexisting neuropsychological skill level is to use the 50th percentile of conventional norms compiled from neurologically intact individuals (Heaton, Miller, Taylor, & Grant, 2004). Conventional norms are typically stratified by age, educational level, and, to a lesser extent, gender to account for non-brain injury-related variance from these factors. In a similar vein, race norms for African Americans have recently been generated (e.g., Lucas et al., 2005), as group mean scores fall below those of White Americans on many neuropsychological tests. This endeavor has not been without its critics (Brandt, 2007; Gasquoine, 2009). One
argument against the use of race as a stratification variable is the difficulty with its operational definition, a charge that also applies to bilingual norms that are now being advocated (Rivera Mindt, Byrd, Saez, & Manly, 2010) and produced (Rosselli et al., 2010) for Hispanic Americans.

Bilingualism is a multidimensional continuous construct with many possible discrete classification schemes (Grosjean, 1998; Rhodes, Ochoa, & Ortiz, 2005). Key assessment metrics include both proficiency (an ability rating in each language) and dominance (a difference score between proficiency measures in two languages). Proficiency and dominance can vary across expression, comprehension, reading, and writing tasks, being influenced by language use characteristics, such as age of second language acquisition or amount of second language exposure (e.g., Gutierrez-Clellen & Kreiter, 2003; Portocarrero et al., 2007). One such characteristic that is likely especially pertinent for Hispanic Americans is the elective versus circumstantial distinction of Valdes and Figueroa (1994). Elective bilinguals choose to learn a second language, whereas circumstantial bilinguals are forced to learn a second language due to life events. It can be surmised that group mean proficiency scores for elective bilinguals (often selected as experimental subjects) would be higher than that for an immigrant population like Hispanic Americans, as the latter contain a large number of circumstantial bilinguals.

Heterogeneity within the bilingual Hispanic American population may restrict the use of generic bilingual norms as the “utility of any set of normative data is largely dependent upon the degree of similarity between the individual test-taker and the characteristic features of the normative sample” (Lucas et al., 2005, p. 176). As an illustration, language dominance in Hispanic American bilinguals ranges from English-dominant through balanced bilingualism to Spanish-dominant. The two language-dominant groups differ significantly in scores on certain neuropsychological measures depending on the language of test administration (Gasquoine et al., 2007).

An alternative method of estimating the preexisting neuropsychological skill level to the 50th percentile of conventional norms is the “individual comparison standard,” whereby the estimate is generated from individual data (Lezak et al., 2004). The nature of individual data used has included the following, either alone or in various combinations: demographics (e.g., Barona, Reynolds, & Chastain, 1984); school achievement test records (Baade & Schoenberg, 2004); obtained post-onset scores from so-called “hold” (i.e., measures of crystallized intelligence that are considered relatively resistant to brain injury) Wechsler subtests of Vocabulary, Information, Matrix Reasoning, and Picture Completion (e.g., Schoenberg, Duff, Scott, & Adams, 2002); obtained post-onset hold word pronunciation test scores (e.g., Schretlen et al., 2009); and the best obtained post-onset score on any neuropsychological test or demographic data point (Lezak et al., 2004). Regardless of the type of data used or how it is combined, all individual comparison standards give an estimate that is conceived of as preexisting FSIQ (i.e., “g”) or other composite IQ score that can be compared with obtained post-onset neuropsychological test scores, in a similar fashion to the 50th percentile estimate. Importantly, this comparison takes place on conventional normative tables derived from a sample of neurologically intact individuals that mirror the general characteristics of the population, such as those of the meta-norm type (Mitrushina, Boone, Razani, & D’Elia, 2005). These conventional norms do not need to be stratified by race/ethnicity, bilingualism, or other individual difference variables, as the non-brain injury-related effect of these variables is taken into account in the preexisting skill estimate.

There is a lack of consensus in determining the best method for calculating the individual preexisting skill estimate, as intermethod comparison ideally requires a standard that is typically unavailable, namely the “true” preexisting skill level. In the absence of this gold standard, comparisons have been made in neurologically intact groupings by assuming that the best estimate for any given study sample is the one that produces scores most approximate to the normal distribution (Ball, Hart, Stutts, Turf, & Barth, 2007; Langeluddecke & Lucas, 2004; Powell, Brossart, & Reynolds, 2003; Schoenberg, Scott, Ruwe, Patton, & Adams, 2004). Results using this method of comparison have been decidedly mixed, with no one method proving superior. All methods of estimating the preexisting neuropsychological skill level from individual data suffer regression to the mean (Baade & Schoenberg, 2004; Ball et al., 2007; Griffin et al., 2002; Langeluddecke & Lucas, 2004; Powell et al., 2003; Spinks et al., 2009; Yates, 1956).

Individual estimates of the preexisting skill level will be more accurate than the 50th percentile of conventional norms in cases where the “true” preexisting skill level falls toward either end of the normal distribution, with the advantage increasing as a function of the difference between the “true” preexisting skill level and the 50th percentile. For Hispanic Americans, it is hypothesized that an individual estimate of the preexisting skill level will be more accurate than the 50th percentile of conventional norms when using language format measures upon which Hispanic Americans historically obtain low scores. In contrast, the two methods should show little difference for visual-perceptual format measures where Hispanic American group performance approximates the conventional mean. This study tested these two hypotheses using a sample of bilingual, neurologically intact, Hispanic Americans assessed in both Spanish and English. Relative accuracy of the two methods was assessed by comparing the number of false positives (i.e., neurologically intact individuals misdiagnosed with cognitive impairment) with that expected in a normal distribution of scores.
Method

Participants

Sixty Spanish/English bilingual adults were recruited by word of mouth from the Rio Grande Valley region of South Texas. This included all \( N = 36 \) participants whose data were analyzed in Gasquoine and colleagues (2007), a study that looked at score differences from Spanish and English languages of test administration among bilinguals. Consecutive volunteers were included if they were within the age range of 18–65 years, subjectively able to carry on a conversation in both languages, and had no self-reported history of brain damage, psychiatric disorder, or drug/alcohol abuse. Four participants did not complete the project (failing to return for the second session) and were excluded from analysis. The final sample \( N = 56 \) consisted of 24 male and 32 female Mexican Americans with a mean age of 32.4 years \( (SD = 11.6) \). Education ranged from 9 to 19 \( (M = 13.9; SD = 2.2) \) years. Thirty-two participants were born in the USA and 24 in Mexico. Those born in Mexico had lived in the USA from 3 to 37 \( (M = 15.6; SD = 7.8) \) years. Most participants \( (40) \) were educated solely in the USA, 7 solely in Mexico, and 9 in both countries. Spanish was the first language for 48 participants, and 39 were residing in Spanish language-dominant homes.

Measures

Language proficiency and dominance. Administration of four subtests (Picture Vocabulary, Verbal Analogies, Letter-Word Identification, and Dictation) from the WMLS-R Spanish and English Form A versions allowed calculation of summary Broad Spanish and English Ability scores, with age-corrected norms \( (M = 100; SD = 15) \). These were subtracted from each other to produce a measure of language dominance with positive difference scores indicating greater Spanish proficiency.

Visual-perceptual processing skills. The Matrix Reasoning subtest of the WAIS-III has items that require visual-perceptual matches of sameness and symmetry and the solving of analogy problems. It was administered in both Spanish and English with the only difference being language of administration. Spanish instructions were generated from the standard back-translation techniques. Scale scores \( (M = 10; SD = 3) \) were generated for each language of administration from the single set of national, English language, age-corrected norms.

Neuropsychological test battery. La Bateria Neuropsicologica en Espanol (Artiola i Fortuny, Hermosillo, Heatton, & Pardee, 1999) was administered along with versions of the original English language tests from which this battery was adapted. The Spanish battery includes eight tests: (a) “Story Memory”: adapted from Heaton, Grant, and Matthews (1991), the paragraph content differs from the English version but both are scored out of 29. Three measures were analyzed: Trial 1 recall; learning; and delayed recall. (b) “Verbal Learning”: the format parallels the California Verbal Learning Test-II (Delis, Kramer, Kaplan, & Ober, 2000) with different words. Seven measures were analyzed: Trial 1 recall; total words; list B recall; short-delay free recall; short-delay cued recall; long-delay free recall; and long-delay cued recall. (c) “Digit Span”: number sequences were translated from the WAIS-R (Wechsler, 1981). The measure analyzed was digit span total. (d) “Letter Fluency”: the format parallels the Word Fluency subtest of the Neurosensory Center Comprehensive Examination for Aphasia (Spreen & Benton, 1969), but uses the letters P, M, and R. The measure analyzed was total words generated. (e) “Figure Memory”: adapted from Heaton and colleagues (1991), this test uses the three geometric figures from the original Wechsler Memory Scale (WMS; Wechsler, 1945). Three measures were analyzed: Trial 1 recall; learning; and delayed recall. (f) “Wisconsin Card Sorting Test”: the format is as in Heaton, Chelune, Talley, Kay, and Curtiss (1993). The measure analyzed was total errors. (g) “Spatial Span”: format parallels the WMS-III (Wechsler, 1997b) subtest. The two measures analyzed were number correct forward and backward. (h) “Stroop Color and Word Test”: format as in Golden and Freshwater (1998). The measure analyzed was the number correct in the interference condition.

Overall, there were 19 separate measures, 13 using primarily a language format (measures from Story Memory, California Verbal Learning Test, Digit Span, Letter Fluency, Stroop Color and Word) and 6 using primarily a visual-perceptual format (measures from Figure Memory, Wisconsin Card Sorting Test, Spatial Span). Spanish language norms used were age and education corrected from the USA–Mexico border region as published in the manual. An exception was Digit Span Total where WMS-III norms were used in both Spanish and English, as the Bateria Neuropsicologica norms were separated for forward and backward administrations and this has no English equivalent in either the WAIS-III or the WMS-III. For the English tests, Letter Fluency used the letters F, A, and S, and digit and tapping sequences for Digit and Spatial Span, respectively, were from the WMS-III. Conventional norms from the various manuals were used for the California Verbal Learning Test-II, Digit Span, Wisconsin Card Sorting Test, and Spatial Span. Age-corrected meta-analytical norms (Mitrushina et al., 2005)
were used for Letter Fluency and the Stroop Color and Word Test. Gender-, age-, and education-corrected conventional Caucasian norms were used for Story Memory and Figure Memory (Heaton et al., 2004), as there were no separate tables for English-speaking Hispanics.

Procedure

Participants were tested over two sessions 14–61 (M = 27.6; SD = 8.5) days apart. At each session, both examiner and participant spoke either English or Spanish only. Participants were randomly assigned to a language condition whereby half of the participants spoke English during the first session and Spanish during the second and the other half was tested first in Spanish then English. The WMLS-R, Matrix Reasoning, and neuropsychological test battery were administered in both sessions in the assigned language. The length of each session was approximately 2 h.

WMLS-R Picture Vocabulary subtest scores in each language were used to make the individual estimate against which to compare obtained scores on neuropsychological measures that primarily used a language format. Use of vocabulary scores is the oldest method of preexisting skill estimation, dating back to Babcock (1930) who used the score discrepancy between measures of vocabulary and memory/timed tasks as a measure of “mental deterioration” in psychotic patients at Manhattan State Hospital. She chose Vocabulary as it had been shown to hold in chronic schizophrenia and was highly correlated with “g.” Wechsler Matrix Reasoning subtest scores in each language were used to make the individual estimate against which to compare obtained scores on neuropsychological measures that primarily used a visual-perceptual format. Matrix Reasoning is a hold subtest originally recommended for use in estimating the preexisting visual-perceptual skill level in cases where obtained post-onset language skills are potentially compromised by brain injury, such as that resulting from lesions to the language zones of the dominant hemisphere (e.g., Schoenberg et al., 2002).

False positives using the 50th percentile of conventional norms to estimate the preexisting skill level were scores on any measure ≤16th percentile. False positives using individual estimates of preexisting skill level were language format measure scores ≤1 SD below the obtained WMLS-R Picture Vocabulary subtest score in the corresponding language and visual-perceptual format measure scores ≤1 SD below the obtained WAIS-III Matrix Reasoning subtest score in the corresponding language. Note that in clinical neuropsychology, the probability of obtaining false positives increases as a function of the number of measures administered (Crawford, Garthwaite, & Gault, 2007). Often called the “base rate” (of obtaining x false-positive scores from y inter-measure comparisons), this probability was the same for the two methods of preexisting skill estimation, as the number and type of measures analyzed was identical.

Results

Broad WMLS-R Ability scores (language proficiency) ranged from 68 to 122 (M = 94; SD = 12) in Spanish and from 71 to 119 (M = 92; SD = 9.6) in English. Spanish minus English difference scores (language dominance) ranged from −28 to +48 (M = 1.0; SD = 15). Means (SDs) for the two individual comparison standards, WMLS-R Picture Vocabulary and WAIS-III Matrix Reasoning, are shown in Table 1 and indicate a visual-perceptual over language processing advantage.

Table 2 shows the number of false positives on the 13 language format measures in Spanish and English using the 50th percentile of conventional norms and the WMLS-R Picture Vocabulary subtest score as the estimate of the preexisting skill level. Based upon the normal distribution, the number of false positives expected for each measure was 9 (i.e., .56 × 16%). The number of false positives actually observed was strikingly different across methods of preexisting skill estimation ranging from a mean low of 2.5 per measure (WMLS-R Picture Vocabulary in Spanish) to a mean high of 24.1 (50th percentile in English). The method of preexisting skill estimation that generated a false-positive mean most approximate to the normal distribution was obtained using the WMLS-R Picture Vocabulary subtest score in English (M = 10). Repeated-measures multivariate analysis of variance (MANOVA) with the method of false-positive calculation as the repeated measure was significant, F(4, 48) = 88.4, p < .0001, η² = 0.97. Post hoc analysis of planned comparisons (four methods of preexisting skill estimation

Table 1. Means (SDs) in Spanish and English for individual comparison scores

<table>
<thead>
<tr>
<th>Subtest</th>
<th>National</th>
<th>Spanish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMLS-R Picture Vocabulary</td>
<td>100 (15)</td>
<td>84.4 (8.8)</td>
<td>86.0 (9.6)</td>
</tr>
<tr>
<td>WAIS-III Matrix Reasoning</td>
<td>10 (3)</td>
<td>10.9 (2.6)</td>
<td>11.3 (2.5)</td>
</tr>
</tbody>
</table>

vs. expected) using a p-level of .05/2 = .025 (two separate MANOVAs were completed in this study) showed that all methods of preexisting skill estimation generated a mean number of false positives significantly different from the expected mean except WMLS-R Picture Vocabulary in English.

Table 3 shows the number of false positives on the six visual-perceptual format measures in Spanish and English using the 50th percentile of conventional norms (50%) and the WAIS-III MR subtest score as the estimate of preexisting skill level (N = 56).

Table 3. False positives on six visual-perceptual format measures in Spanish and English using the 50th percentile of conventional norms (50%) and the WAIS-III MR subtest score as the estimate of preexisting skill level (N = 56)

<table>
<thead>
<tr>
<th>Neuropsychological measure</th>
<th># False positive</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spanish</td>
<td>English</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50% PV</td>
<td>50% PV</td>
<td></td>
</tr>
<tr>
<td>Figure Memory Trial 1 recall</td>
<td>4 3</td>
<td>7 9</td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>3 3</td>
<td>6 7</td>
<td></td>
</tr>
<tr>
<td>Delayed recall</td>
<td>2 1</td>
<td>2 2</td>
<td></td>
</tr>
<tr>
<td>Wisconsin Card Sort total errors</td>
<td>5 7</td>
<td>16 22</td>
<td></td>
</tr>
<tr>
<td>Spatial span</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward</td>
<td>9 18</td>
<td>14 24</td>
<td></td>
</tr>
<tr>
<td>Backward</td>
<td>7 10</td>
<td>14 21</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30 42</td>
<td>59 85</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>5.0 (2.6)*</td>
<td>7.0 (6.3)*</td>
<td>9.8 (5.6)*</td>
</tr>
</tbody>
</table>

*Significantly different from the expected mean of 9 at p < .025.

vs. expected) using a p-level of .05/2 = .025 (two separate MANOVAs were completed in this study) showed that all methods of preexisting skill estimation generated a mean number of false positives significantly different from the expected mean except WMLS-R Picture Vocabulary in English.

Table 3 shows the number of false positives on the six visual-perceptual format measures in Spanish and English using the 50th percentile of conventional norms and the WAIS-III MR subtest score as the estimate of preexisting skill level. Again the expected number of false positives was nine per measure. The mean number of false positives actually generated was strikingly different across methods of preexisting skill estimation ranging from a low of five per measure (50th percentile in Spanish) to a high of 14.1 (Matrix Reasoning in English). The method of preexisting skill estimation that generated a false-positive mean most approximate to the normal distribution was the 50th percentile of conventional norms in English (M = 9.8). Repeated-measures MANOVA with the method of false-positive estimation as the repeated measure was significant, F(4, 20) = 255.3, p < .01, η² = 0.99. Post hoc analysis of planned comparisons showed that the only method of preexisting skill estimation that generated a false-positive mean that was significantly different from expected was the 50th percentile in Spanish (M = 5).
Discussion

As hypothesized from previous research on Hispanic American language versus visual-perceptual processing test score differences, the optimal method of preexisting skill estimation differed for language format compared with visual-perceptual format measures. For language format measures, the individual estimate (picture vocabulary scores) generated the mean number of false positives most approximate to that expected from the normal distribution, while the 50th percentile of conventional norms did the same for visual-perceptual format measures.

In both instances, the mean number of false positives most approximate to the normal distribution occurred with English as the language of test administration. Although English language norms were found to be more suitable than Spanish language norms during group analysis, this language of administration would not necessarily be preferred for all individuals within the group. Language-dominant bilinguals should be tested in their dominant language. Spanish language norms from La Bateria Neuropsicologica are regression-based and were collected in the USA–Mexico border region on a sample that was predominantly women (75%), resident of Mexico (60%), and monolingual (76%). The total N of 185 was divided among 5 age and 7 education levels, equating to less than six individuals per age × education cell. They represent a distribution below that of the corresponding English language norms and were found unsuitable for this sample of Hispanic Americans.

Consistent with previous findings of low scores for bilinguals on language format neuropsychological tests (Bialystok, 2007; Bialystok et al., 2008; Gollan & Acenas, 2004; Gollan et al., 2002, 2005; Portocarrero et al., 2007; Roberts et al., 2002; Rosselli et al., 2002), mean WMLS-R language proficiency scores for the sample studied here were about 0.6 SD below average in both languages, despite a mean 14 years of education. Also consistent with previous findings on Hispanic Americans (degree of bilingualism unknown: Neisser et al., 1996; Puente & Salazar, 1998), scores on visual-perceptual format tests (e.g., Matrix Reasoning) were superior to those on language format tests (e.g., WMLS-R Picture Vocabulary), although with this sample the difference was more pronounced being >1 SD. Language dominance, measured as the difference in WMLS-R language proficiency scores between Spanish and English, showed balance across the sample as indexed by a mean difference score of only 1, ever so slightly in favor of Spanish.

The unsuitability of conventional monolingual norms in either Spanish or English for bilingual Hispanic Americans has been noted elsewhere (Roberts et al., 2002; Rosselli et al., 2010), although current findings demonstrate that this is restricted to language format measures. In practice, the use of the 50th percentile of monolingual language format measure norms as an estimate of preexisting skill level for bilinguals should be avoided. One possible alternative is the production of separate bilingual Hispanic American norms (Rosselli et al., 2010). Problems with this solution include: the intensive labor involved; the heterogeneity of the bilingual population; the existence of different ethnic groupings within the Hispanic American community (e.g., Mexican vs. Cuban Americans) that may require separate norms; and the difficulty of determining if a bilingual Hispanic American qualifies for bilingual or monolingual norms. Separate bilingual/monolingual norms require parceling a continuous variable (bilingualism) into an artificial dichotomy (balanced vs. language-dominant) and placement of borderline individuals in adjacent normative categories can result in wild swings in percentile scores (Zachary & Gorsuch, 1985). The other alternative is the use of an individual estimate of preexisting skill level with conventional monolingual normative tables and the efficacy of this approach has been demonstrated here.

The localized nature of the study participants and the heterogeneity of the bilingual population within the USA may limit the generalization of the findings to other Hispanic American and/or bilingual groupings. This study was conducted in an area of Texas near the Mexican border where Spanish and English language use in both the home and community is common practice for many Hispanic Americans. Participants were consecutive volunteers who were subjectively bilingual in the sense that they were able to carry on a conversation in both languages. No attempt was made to exclude participants on the basis of subjective judgments about language abilities, so as to approximate the natural setting faced by clinicians. Instead, the level of bilingualism of each participant was formally assessed in terms of language proficiency and dominance. WMLS-R Broad Ability scores (encompassing measures of expression, comprehension, reading, and writing) were used to measure language proficiency in Spanish and English. These scores ranged from about 2 SDs below average to about 1.5 SD above average in both languages.

There is no agreement in clinical neuropsychology as to the optimal size of the cutoff used to create an artificial dichotomy of cognitive impairment versus unimpairedness from a continuous distribution of test scores, with cutoffs of 1 SD through 2 SD being reported in the literature (Heaton et al., 2004; Gasquoine, 2011). Differing cutoff values reflect differing relative statistical probabilities of making true- and false-positive identifications of cognitive impairment and particular cutoffs may be preferred in specific contexts. The 1 SD cutoff was chosen here primarily to maximize the number of false-positive identifications in the study. In all studies of neurologically intact participants, true positive rates (i.e., correctly identified neurologically impaired individuals) are not measured, although this metric is equally as important as the false-positive rate in deficit measurement.
Conflict of Interest

None declared.

Acknowledgements

Part of the data collected in this study was used in a thesis by the second author in partial fulfillment of the requirements for the degree of Master of Science in Experimental Psychology at the University of Texas-Pan American.

References


