Errors on the WCST Correlate with Language Proficiency Scores in Spanish-English Bilingual Children

Clemente Vega¹, Mercedes Fernandez²,*

¹Department of Neurology, Children’s Hospital Boston/Harvard University, Boston, MA, USA
²Farquhar College of Arts and Sciences, Nova Southeastern University, Ft Lauderdale, FL, USA

*Corresponding author at: Farquhar College of Arts and Sciences, Nova Southeastern University, 3301 College Avenue Ft Lauderdale, FL 33314, USA.
Tel.: +1-954-262-7804; fax: +1-954-262-3760.
E-mail address: mf934@nova.edu (M. Fernandez).

Abstract

This study examined the relationship between the degree of balance in bilinguals (i.e., degree to which an individual shows equivalence in their level of mastery of two languages) and performance on executive function tests. Twenty-four “more balanced” and 16 “less balanced” Spanish-English bilingual, third- and fourth-grade children completed the Wechsler Abbreviated Scale of Intelligence, the English and Spanish Bilingual Verbal Ability Test (BVAT) (an objective measure of English and Spanish language proficiency), and executive function tests (Wisconsin Card Sorting Test [WCST] and Stroop Test). Significant correlations were found between the degree of balance and error scores on the WCST, such that participants with more similar scores on both English and Spanish BVAT (i.e., the more-balanced group) scored lower on perseveration measures of the WCST than the less-balanced group. Between-group comparisons showed significantly fewer perseverative errors and perseverative responses on the WCST in the more-balanced group compared with the less-balanced group. Stroop Test scores did not reveal group differences. These findings support previous results documenting a “bilingual advantage” on executive function tests and provide support for the hypothesis that the degree of balance moderates the relationship between bilingualism and executive functions.

Keywords: Executive functions; Language; Bilingualism; WCST; Stroop

Introduction

Research documents relatively better working memory, selective attention, cognitive flexibility, and reasoning ability in bilinguals compared with monolinguals (see Bialystok, 2009, for a review). These enhanced abilities, collectively known as executive functions, are attributed to superior inhibitory control in bilinguals (Green, 1998). According to Bialystok & Martin (2004, p. 338), the inhibitory mechanism proposed by Green (1998) to explain how bilinguals manage two languages is mediated by the same cortical centers used to solve nonlanguage tasks that contain salient but misleading information. Because bilinguals exercise this inhibitory mechanism continuously, they are more efficient at performing cognitive tasks with distracting information.

However, these enhanced abilities appear to be mediated by second language proficiency. “Balanced bilinguals” (i.e., those who are equally fluent in two languages) exhibit a performance advantage, whereas “unbalanced bilinguals” (i.e., those who are not equally fluent in both language) do not exhibit the same level of inhibitory control or similar performance advantages on tests of executive function (Carlson & Meltzoff, 2008; Costa & Santesteban, 2004; Kharkhurin, 2008; Rosselli et al., 2002; Zied et al., 2004). Thus, the purpose of this study was to quantify bilinguals’ language proficiency in their two languages and to correlate the degree of mastery in both languages with measures of executive function.

One study evaluating the bilingual advantage in younger (30–54 years of age) and older (60–88 years of age) adults by Bialystok, Craik, Klein and Viswanathan (2004) found that bilinguals outperformed age-matched monolinguals on the...
Simon task, a timed task that requires subjects to ignore one dimension of a stimulus and to respond with a button-press to the other dimension. The monolingual group included English speakers living in Canada, and the bilingual group included Tamil-English speakers living in India. The bilinguals learned Tamil (a language from Southern India) as their first language and were educated in Tamil and English starting at 6 years of age. A language questionnaire revealed that on average, participants spoke English and Tamil on a daily basis (56% and 44% of the time, respectively). These researchers proposed that enhanced executive control develops as a function of bilinguals’ experience inhibiting one language when speaking in their other language.

In another study, Bialystok and Shapero (2005) found that 6-year-old bilingual children performed better than their monolingual counterparts on a card-sorting task and on a task requiring reversal in visual interpretation of ambiguous figures (i.e., tasks requiring re-assigning the meaning of an image). The monolingual children were English speakers, and the bilinguals spoke a language other than English (e.g., French, Kurdish, Chinese, Russian, etc.) as their first language and learned English as their second language. The bilingual children were raised speaking both languages and used both languages daily. Again, between-group differences were attributed to superior inhibitory control of attention.

Studies comparing balanced and unbalanced bilinguals suggest that the enhanced nonlanguage executive function abilities in bilinguals are moderated by second language proficiency, with balanced bilinguals performing better than unbalanced bilinguals and monolinguals (Carlson & Meltzoff, 2008; Costa & Santesteban, 2004; Kharkhurin, 2008; Rosselli et al., 2002; Zied et al., 2004).

For instance, Carlson & Meltzoff (2008) found a bilingual advantage in native bilingual children, but not in English speakers enrolled in a second language immersion program, on a demanding task that required attention and inhibition. The native bilingual group included children who were exposed to English and Spanish since birth and who frequently spoke both languages in their environment. “In this group, either both parents were native Spanish speakers or one parent spoke Spanish and the other spoke English” (p. 285). Based on parents’ report on a language questionnaire, it was determined that children in this group had approximately equal exposure to both languages. The immersion group consisted of English monolingual children enrolled in a language immersion program in a public school. These children received academic instruction in English half the day and Spanish or Japanese instruction the remainder of the day. These children had been enrolled in the language immersion program for 6 months at the time of participation in this study, and they came from a home where only English was spoken. The bilingual advantage was once again observed in those who spoke both languages daily and from birth and were therefore more likely to be balanced bilinguals.

In a study where proficiency in two languages was objectively assessed, Zied and colleagues (2004) found that balanced bilinguals performed similarly on two different versions (French and Arabic) of the Stroop test, a test employed by neuropsychologists as a measure of inhibition, whereas unbalanced bilinguals performed significantly better when responding in their dominant relative to their nondominant language. Language proficiency was measured by administering the Boston Naming Test (BNT, a confrontation naming test of common objects) in both French and Arabic. Those who scored within 1 SD on both versions of the BNT were classified as balanced bilinguals. Those who had a difference score >1 SD were classified as unbalanced bilinguals. Together, these studies suggest that compared with monolinguals, bilinguals exhibit enhanced executive control and that this advantage is moderated by language proficiency with balanced bilinguals outperforming unbalanced bilinguals.

Importantly, although the above-cited studies document the moderating effects of language proficiency on executive functioning, not all studies investigating the benefits of bilingualism employ objective measures to assess language proficiency; instead, studies rely on self-report questionnaires and frequency of language usage (e.g., Bialystok, Craik, & Ryan, 2006; Lee & Chan, 2000). The purpose of this study was to objectively quantify language proficiency in two languages (Spanish and English) and to correlate the degree of balance in both languages with performance on two tests of executive function.

Materials and Methods

Participants

Forty, Spanish-English bilingual, third- and fourth-grade children participated in the study. Healthy children without a history of psychiatric disorders, medical conditions, or learning disabilities/attention deficit hyperactivity disorder were recruited from a private school located in an upper-middle class community of South Florida.

Group Inclusion

Participants were classified into one of two groups, More-Balanced Bilinguals (MBB) and Less-Balanced Bilinguals (LBB), where “balanced” refers to the extent to which a participant was equally proficient in both languages. For each participant, the
score on the Spanish Oral Vocabulary subtest was subtracted from the English Oral Vocabulary subtest score of the Bilingual Verbal Ability Test (BVAT) (see description below). Participants with a difference score of $< 15$ points were classified as MBB ($n = 24$), whereas those with a difference score of $> 15$ points were classified as LBB ($n = 16$).

Materials

The Bilingual Verbal Ability Tests (BVAT) (Muñoz-Sandoval, Cummings, Alvarado, & Ruef, 1998). This test is composed of three subtests, Picture Vocabulary, Oral Vocabulary, and Verbal Analogies and generates a score of English Language Proficiency and of Bilingual Verbal Ability. All subtests are administered in English first (this score generates the English Language Proficiency Index). Those items that are missed in English are then administered in the person’s native language, which generates a gain score. The Bilingual Verbal Ability score is the sum of the English Language Proficiency score plus the gain score. This score reflects a bilingual’s overall verbal ability.

In our study, to determine group membership, the English and the Spanish versions of the BVAT were administered in their entirety. This represented a deviation from the standardized administration procedure, which dictates that only items missed in English are to be administered in the person’s other language. Additionally, because we did not follow the administration procedure, we could not use the BVAT scoring program, which generates transformed scores; we therefore report raw scores in this manuscript. We chose to administer the BVAT in this manner to enable a more “accurate” comparison of scores in English and Spanish. From among the three BVAT subtest scores, the Oral Vocabulary subtest was selected as the score to be used for determining group assignment. According to the manual, the Oral Vocabulary subtest provides the best representation of language proficiency based on its strong correlation with the Bilingual Verbal Ability score, a measure which “represents the combined verbal cognitive ability that is distributed across different domains in the bilingual child’s native and second language” (p. 11). In addition, when examining performance on the Spanish subtests, all participants in our cohort scored high on Picture Vocabulary, the easiest subtest, and scored poorly on Verbal Analogies; thus, the test that best distinguished between levels of Spanish proficiency was Oral Vocabulary. We used the 15-point cutoff score described in the manual (p. 81) to distinguish between groups.

The Wisconsin Card Sorting Test (WCST) (Heaton, 1993). This test measures concept formation and set shifting abilities (flexibility) by providing the examinee stimulus cards that differ by color, form, and number, and the examinee must figure out the correct dimension to sort the cards based on feedback (correct or incorrect match) provided by the examiner. The sorting rules change several times and the participant must figure out the new sorting rule every time. To successfully complete this task, a person must be able to shift cognitive sets in order to abandon the previously correct sorting rule for the new rule. Errors (perseverative errors and perseverative responses) are used as measures of cognitive flexibility and were used as measures of executive function in this study.

The Stroop Color-Word Test (Golden, 1978). This test provides a measure of ability to inhibit a prepotent or dominant response in favor of a nondominant response. Participants read color words printed in incongruent ink colors and report the color of the ink while ignoring the color word. This test was used to measure prepotent response inhibition, a component of executive function.

The Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999). This test provides a measure of intellectual ability. This battery includes four subtests: Vocabulary, Similarities, Block Design, and Matrix Reasoning, and generates Full-Scale IQ (FSIQ), Verbal IQ (VIQ), and Performance IQ (PIQ). This test was administered to compare the two groups in overall intellectual abilities.

A brief language questionnaire was completed by the parents to determine whether English or Spanish (or both) were spoken at home.

Design

Third and fourth graders were recruited by mailing letters to parents to inform them of the study. Consent forms were mailed to the parents of every child in the third and the fourth grade in the school. A language questionnaire was also sent along with the consent forms. Only children whose parents signed the consent forms and indicated some degree of Spanish language exposure at home were invited to participate. Participants were tested at school during their first class period (homeroom) and testing was completed within a 75 min session. All participants were administered the WASI, Stroop test, and WCST before the BVAT, so as to keep the experimenter blind to group assignment until after all other tests were administered. The Spanish and English versions of the BVAT were administered to all participants in a counterbalanced order, such that 50% of the participants were administered the English version first, whereas the others were administered the Spanish version first.
Results

Demographic Information

The two groups did not differ by age, sex, or by the number of students in the third or fourth grades. Table 1 shows demographic information for the two groups and results of statistical analyses.

To determine whether the two groups were different on measures of intellectual abilities, participants were compared on their FSIQ, PIQ, and VIQ scores. Group means, standard deviations, and statistical results are presented in Table 2. As can be seen, the two groups did not differ on any of these variables.

Language Questionnaire

Based on parent report, all children had some degree of exposure to Spanish and English at home. Additionally, during the testing session, all children reported taking a Spanish course as part of their regular school curriculum.

Primary Analyses

To test the hypothesis that the degree of balance of the two languages was correlated with performance on executive function tests, Pearson’s bivariate correlations were computed. Language proficiency was quantified by computing difference scores between English and Spanish Oral Vocabulary subtest for each participant, which generated a “Bilingualism Score (BS).” The BS was correlated with perseverative errors and with perseverative responses on the WCST, as well as with the

### Table 1. Demographic information by group

<table>
<thead>
<tr>
<th>Sex</th>
<th>More-Balanced Bilingual group (n = 24)</th>
<th>Less-Balanced Bilingual group (n = 16)</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>9</td>
<td>7</td>
<td>$\chi^2 = 0.16, p = .693$</td>
</tr>
<tr>
<td>Girls</td>
<td>15</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>3</td>
<td>8</td>
<td>$\chi^2 = 1.11, p = .292$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Age (in years)</td>
<td>9.02 (0.61)</td>
<td>9.25 (0.52)</td>
<td>$t(38) = 1.24, p = .223$</td>
</tr>
</tbody>
</table>

### Table 2. Means, standard deviations, and statistical analyses for Wechsler Abbreviated Scale of Intelligence, Stroop Test, Wisconsin Card Sorting Test, and the Bilingual Verbal Ability Test by group

<table>
<thead>
<tr>
<th></th>
<th>More-Balanced Bilingual group (n = 24)</th>
<th>Less-Balanced Bilingual group (n = 16)</th>
<th>Statistics</th>
<th>Effect size Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSIQ</td>
<td>109.92 (10.43)</td>
<td>111.25 (9.42)</td>
<td>$t(38) = 0.411, p = .683$</td>
<td></td>
</tr>
<tr>
<td>VIQ</td>
<td>112.54 (13.17)</td>
<td>113.31 (9.12)</td>
<td>$t(38) = 0.203, p = .840$</td>
<td></td>
</tr>
<tr>
<td>PIQ</td>
<td>106.04 (11.66)</td>
<td>106.81 (12.71)</td>
<td>$t(38) = 0.198, p = .844$</td>
<td></td>
</tr>
<tr>
<td>Stroop Test (raw score)</td>
<td>29.38 (6.91)</td>
<td>28.88 (7.37)</td>
<td>$t(38) = 0.218, p = .828$</td>
<td>0.07 (negligible)</td>
</tr>
<tr>
<td>Color-Word trial</td>
<td>12.25 (4.52)</td>
<td>14.06 (5.54)</td>
<td>$t(38) = 2.24, p = .031$</td>
<td>0.74 (medium)</td>
</tr>
<tr>
<td>Perseverative responses</td>
<td>8.92 (4.68)</td>
<td>9.83 (6.11)</td>
<td>$t(38) = 2.22, p = .032$</td>
<td>0.74 (medium)</td>
</tr>
<tr>
<td>Wisconsin Card Sorting Test (raw scores)</td>
<td>18.50 (3.49)</td>
<td>21.19 (2.34)</td>
<td>$t(38) = 2.70, p = .010$</td>
<td></td>
</tr>
<tr>
<td>Oral Vocabulary English</td>
<td>7.79 (4.25)</td>
<td>3.88 (2.25)</td>
<td>$t(38) = 3.37, p = .002$</td>
<td></td>
</tr>
</tbody>
</table>
color-word score of the Stroop test. Results revealed a significant and positive correlation between BS and perseverative errors ($r = .320, p = .044$) and perseverative responses ($r = .328, p = .039$). That is, as the difference score between Spanish and English Oral Vocabulary got smaller, so did the number of perseverative responses and perseverative errors. Similar results were not obtained when BS were correlated with Stroop Color-Word scores ($r = .060, p = .712$).

All participants regardless of group inclusion scored significantly higher on the English ($M = 19.58, SD = 3.33$) than on the Spanish ($M = 6.23, SD = 4.05$) version of the Oral Vocabulary subtest of the BVAT, $t(39) = 20.23$, $p < .001$. This was expected as all participants received most of their academic instruction in English, with the exception of the Spanish class.

### Secondary Analyses

To test the hypothesis that the two groups were different on perseverative errors and perseverative responses generated by the WCST and the color-word score generated by the Stroop test, independent sample $t$-tests were computed. Table 2 shows means, standard deviations, and results of statistical analyses for each test. As can be seen in this table, relative to the LBB group, the MBB group generated significantly fewer perseverative errors and perseverative responses on the WCST. These comparisons yielded a medium effect size (Cohen’s $d$; Thalheimer & Cook, 2002). However, the two groups did not differ on performance on the Stroop test, and the effect size was negligible.

### Discussion

This study supports the hypothesis that the degree of balance in one’s level of mastery of two languages moderates the relationship between bilingualism and cognitive flexibility. This study also highlights the importance of quantifying mastery in each language when assessing the construct of bilingualism. We objectively quantified the degree of balance in two languages to generate a BS and found that cognitive flexibility measures were positively related to this score, with those who were more balanced exhibiting the best performance on the WCST. In other words, this study documents a nonlinguistic advantage to speaking two languages for individuals who have a relatively balanced level of mastery in both languages.

These results could not be attributed to differences in overall intellectual abilities or tests of verbal abilities as the two groups scored similarly on these tests. Moreover, all participants in this study attended the same school and belong to a similar SES group; therefore, it is unlikely that demographic variables account for these results.

Although both the Stroop and WCST are considered executive function tests, this study did not reveal a relationship between Stroop Color-Word score and the BS. This may be attributable to differences in the task demands of these two tests. The WCST requires primarily conceptual reasoning and cognitive flexibility and has no time limit, whereas the Stroop is timed and measures reading speed and response inhibition. It is also possible that Stroop differences are detected when comparing monolinguals with bilinguals, but not when comparing bilinguals in their dominant language who only differ by the degree of balance in their two languages.

Another explanation may be that brain maturation processes necessary for optimal Stroop performance are not present at such an early age (our participants were between the ages of 8 and 10), and thus, the degree of bilingualism did not affect performance in our sample. It has been shown that overall brain activation during the Stroop task increases significantly in a number of areas including the anterior cingulate, middle frontal gyrus, parietal, and parieto-occipital regions after the age of 11 (Adleman et al., 2002). Thus, the benefits of bilingualism on Stroop performance may only be evident in older age groups.

Studies evaluating the development of automatic and control processes in bilinguals suggest that as bilinguals develop proficiency in a language, the Stroop interference effect diminishes (Tzelgov, Henik, & Leiser, 1990). With respect to Stroop interference, reading is the automatic process which interferes with a person’s ability to suppress reading the word in favor of reporting the ink color. As a person develops reading skills, or as they develop proficiency in another language, as in the case of bilinguals, they are better able to control the reading process and the interference effect decreases. In our study, VIQ scores of bilinguals and monolinguals were similar; thus, one may predict that both groups had similar control and therefore would perform similarly. Finally, it may be that the benefit of bilingualism on Stroop test performance is small, and given our small sample size, this effect was not detected.

### Limitations of this Study

These results add to the existing body of knowledge on the benefits of speaking two languages, but these findings must be interpreted with caution. First, the generalizability of our findings may be limited. Our sample was small and consisted of
children from relatively high SES who were enrolled in Spanish class as part of their regular school curriculum. It may be that children from lower SES or who are not enrolled in a second language class as part of their regular school curriculum do not show similar advantages associated with their bilingualism status. Despite this possibility, however, we note that previous research by Carlson & Meltzoff (2008) found no group differences on executive function tasks between English monolinguals and Spanish-English bilinguals from economic disadvantaged backgrounds prior to controlling for SES. They proposed that children from lower SES gain an advantage in cognition as a result of learning two languages that ascends their academic performance to the level of nondisadvantaged peers. They identified a surprising degree of lower parent education and lesser home-based reading, which would otherwise be predictive of poorer performance on neuropsychological tests.

Additionally, we did not measure acculturation in this study; therefore, we do not know the extent to which acculturation contributed to our findings. Yet, studies have found that acculturation affects performance on executive function tests (Ranani, Burciaga, Madore, & Wong, 2007) and specifically on the WCST (Coffey, Marmol, Schock, & Adams, 2005).

Future Directions

We propose that when studying the concept of bilingualism, categorizing participants into dichotomous groups, such as balanced and unbalanced, may not be as useful as quantifying the degree of balance between bilinguals’ two languages. Therefore, future studies should employ objective measures of the degree of balance between bilinguals’ two languages. Additionally, there is a need for studies to replicate these findings in larger groups and in older groups, to investigate whether this positive relationship holds across the lifespan.

To investigate whether our findings generalize to other aspects of executive function, for example, working memory, studies should incorporate neuropsychological tests that measure these different aspects of executive functions. Lastly, future studies should incorporate measures of acculturation and evaluate its impact on executive function measures.

Conflict of Interest

None declared.

References

