Published in final edited form as:
Clin Neuropsychol. 2021 February ; 35(2): 339–355. doi:10.1080/13854046.2019.1703042.

Demographically-Adjusted Normative Data for the Wisconsin Card Sorting Test-64 Item: Results from the Neuropsychological Norms for the U.S.-Mexico Border Region in Spanish (NP-NUMBRS) Project

M.J. Marquine[a], **D. Yassai-Gonzalez**[a], **A. Perez-Tejada**[a], **A. Umlauf**[a], **L. Kamalyan**[a], **A. Morlett**[a], **P. Suarez**[b], **M. Rivera Mindt**[c], **D. Franklin**[a], **L. Artiola I Fortuny**[d], **M. Cherner**[a], **Robert. K. Heaton**[a]

[a]Department of Psychiatry, University of California, San Diego, USA

[b]Department of Psychiatry and Behavioral Sciences, University of California Los Angeles, USA

[c]Department of Psychology and Latin American and Latina/o Studies Institute, Fordham University & Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, USA

[d]Private Practice, Barcelona, Spain

Abstract

Objective: The Wisconsin Card Sorting Test (WCST) is among the most commonly used tests of executive functioning. We aimed to generate normative data on the 64-item version of this test (WCST-64) for Spanish-speakers living in the U.S.-Mexico Border region.

Methods: Participants included 189 native Spanish-speakers (Age: 19-60; Education: 0-20; 59.3% female) from the Neuropsychological Norms for the U.S.-Mexico Border Region in Spanish (NP-NUMBRS) project who completed the WCST-64. Univariable and interactive associations between demographic variables and raw scores were examined via Spearman correlations, Wilcoxon Rank-sum tests and linear regressions. T-scores for various WCST-64 measures (Total Errors, Perseverative Responses, Perseverative Errors, Conceptual Level Responses and Number of Categories) were obtained using fractional polynomial equations with weights for age, education, and gender. Percentile scores were reported for Failures to Maintain Set. Rates of impairment (T-score<40) were calculated by applying the newly developed norms and published norms for non-Hispanic English-speaking Whites and Blacks.

Results: Older age was associated with worse performance and education was linked to better performance on most WCST-64 raw scores, with stronger education effects among females than males. The norms developed here resulted in expected rates of impairment (14-16% across measures). Applying published norms for non-Hispanic Blacks resulted in generally comparable

Corresponding author: María J. Marquine, PhD, 9500 Gilman Dr #0664, La Jolla, CA 92093, Phone: (858) 534-6748, Fax: (858) 534-5475.

Disclosures
In accordance with Taylor & Francis policy and ethical obligation as a researcher, Dr. Heaton is reporting that he receives royalties from the publisher of the Wisconsin Card Sorting Test-64 item, which is the focus of the present paper.

impairment rates. In contrast, applying previously published norms for non-Hispanic Whites overestimated impairment (38-52% across measures).

Conclusions: These data will enhance interpretation performance on the WCST-64 for Spanish-speakers living in the U.S.-Mexico Border region. Future work will need to examine the generalizability of these norms to other Hispanic/Latino groups.

Keywords

executive function; Spanish-speaking; cross-cultural assessment; ethnic differences

Introduction

The Wisconsin Card Sorting test (WCST; Berg, 1948; Grant & Berg, 1948; Heaton, 1981; Heaton, Chelune, Talley, Kay, & Curtis, 1993) is widely used in neuropsychological evaluations for the assessment of abstract reasoning and cognitive flexibility in response to changing environmental contingencies. These cognitive abilities are considered to be part of the broader cognitive domain of executive functioning. Lesions and dysfunctions to the frontal lobe systems of the brain often result in executive dysfunction, including poor performance on the WCST. Furthermore, neuroimaging studies have linked the WCST with activity in the lateral prefrontal cortex, anterior cingulate cortex, and inferior parietal lobule (Buchsbaum, Greer, Chang, & Berman, 2005). However, it is worth keeping in mind that the frontal lobes are highly complex and underlie a much wider set of cognitive functions than those assessed by the WCST. Thus, “normal” performance in the WCST does not necessarily indicate intact frontal or executive function, generally. Conversely, impaired performances on the WCST can result from dysfunction to other brain regions and systems subserving executive functioning and other cognitive abilities necessary to perform well in the WCST. Therefore, clinical interpretation of individual performance on the WCST should be conducted in the context of a comprehensive neuropsychological evaluation and considering other neuropsychological tests performances with medical, psychological, and other pertinent background information (Heaton et al., 1993).

The full WCST consists of four stimulus cards and two sets of 64 response cards that show figures that vary in form, color, and number of figures. The examinee is asked to match each of the response cards to the four stimulus cards— whichever one he or she thinks it matches. The examinee is given feedback for each sort regarding whether his or her response is right or wrong, but is not told the correct sorting principle. The actual correct principle changes, without warning, after the examinee achieves a certain number of correct sorts in a row (a completed “category”). The test proceeds this way until the examinee successfully completes six categories or sorts all 128 cards— whichever comes first. In order to be successful, the examinee must learn and consistently use a correct sorting category, but then flexibly shift to another sorting category when (s)he discovers that the previous sorting rule is no longer “correct.” Continued uses of previously correct sorting rules after they change are scored as “perseverations” which delay or prevent set-shifting. In sum, the WCST requires conceptualizing sorting principles, working memory (keeping track of feedback on previous sorts), consistent use of a sorting principle that is working, but also the ability to flexibly shift to a new sorting principle when the previous one is no longer working.

The WCST-64 item version (WCST-64) is a shortened version of the WCST. It retains all the features of the standard WCST, except it uses only one 64-card sorting deck instead of two (Greve, 2001; Haaland, Vranes, Goodwin, & Garry, 1987; Kongs, Thompson, Iverson, & Heaton, 2000). The WCST-64 is psychometrically comparable to the standard WCST (Axelrod, Henry, & Woodard, 1992; R. K. Heaton & Thompson, 1992; Smith-Seemiller, Franzen, & Bowers, 1997), and sensitive to neurocognitive changes related to an array of clinical conditions, including traumatic brain injury, Parkinson's Disease, Alzheimer's Disease and schizophrenia (Love, Greve, Sherwin, & Mathias, 2003; Paolo, Axelrod, Troster, Blackwell, & Koller, 1996; Purdon & Waldie, 2001).

Demographic influences on WCST-64 performance have been largely similar to those found in the standard (longer) version of the WCST. Age has been significantly associated with worse performance on the WCST-64, with typically medium effect sizes (Axelrod, Jiron, & Henry, 1993; Norman et al., 2011), and somewhat larger effects in non-Hispanic Whites than Blacks (Norman et al., 2011). Increased education has been associated with better performance on the WCST-64, with small to medium effects (Norman et al., 2011), and in general, there have been no significant differences in performance based on sex/gender (Norman et al., 2011). The limited data available examining racial/ethnic differences on WCST-64 indicate worse performance in non-Hispanic Blacks compared to non-Hispanic Whites (Norman et al., 2011). A few other studies have investigated the impact of culture/ethnicity on the WCST-128. Some of these studies have found no significant differences by racial/ethnic group, when comparing groups of English-speaking college students who self-identified as African American, European American, and Latino/a American (Proctor & Zhang, 2008). Yet, others have found significantly different performance by country (Artiola i Fortuny, Heaton, & Hermosillo, 1998), and degree of acculturation (Coffey, Marmol, Schock, & Adams, 2005) within Spanish-speakers. Specifically, Artiola i Fortuny and colleagues (1998) found Spanish speaking adults living in the U.S.-Mexico border region performed worse than their counterparts living in Spain, and this difference diminished with increasing levels of education (Artiola i Fortuny et al., 1998). They also found that increased years living in the United States (U.S.) was linked to better performance on the WCST-128 (Artiola i Fortuny et al., 1998). Relatedly, Coffey and colleagues found that higher acculturation was associated with better performance on the WCST-128 among Spanish-speaking Mexican Americans (Coffey et al., 2005).

In the U.S., over 37 million people speak Spanish at home (Ryan, 2013). The current number of Spanish-speakers in the U.S. comes second to only one other country in the world (Mexico) and surpasses the number of Spanish-speakers in Spain. The development of neuropsychological tests with normative data for Spanish speakers in the U.S. is important for accurate identification of neurological dysfunction in this large segment of the U.S. population (for a review on available norms for Spanish-speakers in the U.S. see Morlett Paredes, Gooding, et al., This Issue). Among adult Spanish-speakers, normative data have been developed for the 128-item version of the WCST for individuals living in the U.S.-Mexico border region (Artiola i Fortuny, Hermosillo Romo, Heaton, & Pardee, 1999) and for Spanish-speakers in other countries, such as Spain and some Latin American countries (R. K. Heaton, Chelune, Talley, Kay, & Curtiss, 2012). Norms have also been developed for a modified version of the WCST (i.e., 48 response cards with no ambiguous stimuli; Nelson,

1976) for various countries in Latin America (Arango-Lasprilla et al., 2015) and for a population from Spain (Del Pino, Peña, Ibarretxe-Bilbao, Schretlen, & Ojeda, 2016).

The main aim of the current study was to develop normative data for the WCST-64 among Spanish-speaking persons living in the U.S.-Mexico border region. The present manuscript is part of a series of papers in this special issue on the Neuropsychological Norms for the U.S.-Mexico Border Region in Spanish (NP-NUMBRS) project. This larger project presents demographically-corrected normative data on a comprehensive neuropsychological test battery in Spanish assessing multiple domains, i.e., verbal fluency (Marquine et al., This Issue), speed of information processing (Rivera Mindt et al., This Issue; Suarez, Diaz-Santos, Marquine, Rivera Mindt, et al., This Issue), attention/working memory (Gooding et al., This Issue; Scott et al., This Issue), executive functioning (Morlett Paredes, Carrasco, et al., This Issue; Suarez, Diaz-Santos, Marquine, Rivera Mindt, et al., This Issue, in addiiton to the current paper), learning and memory (Diaz-Santos et al., This Issue), visuospatial skills (Scott et al., This Issue), and fine motor skills (A. Heaton et al., This Issue).

In order to determine whether existing normative data on the WCST-64 for other cultural/linguistic groups might apply to the current sample of Spanish-speakers, after developing norms for this test in the current sample, we computed rates of impairment based on the newly developed norms and existing published norms for non-Hispanic English-speakers.

Methods

Participants

One hundred and eighty-nine native Spanish-speaking adults, who were participants in the larger NP-NUMBRS Project (*N*=254), completed the WCST-64. Participants with data on the WCST-64 were recruited from the U.S.-Mexico border regions of Arizona (*n*=59) and California (*n*=130) in two study waves (Cohort 1 [n=120]: 1998-2000 and Cohort 2 [n=69]: 2006-2009). Exclusion criteria included having a significant history of neurologic, metabolic, psychiatric, developmental, or substance abuse problems. See Cherner, Marquine and colleagues (Cherner et al., This Issue) for further details on the sample.

Procedures

The test was administered via computer by trained bilingual (Spanish-English) study psychometrists following standard procedures outlined in Kongs and colleagues (2000; see Artiola i Fortuny et al, 1999 for instructions in Spanish).

The following scores are reported:

1. Total Errors: Number of responses that do not match the correct sorting principle in effect at the time the response is made.
2. Perseverative Responses: Responses that match the “perseverated-to” principle, i.e., a persistent responding to a stimulus characteristic that is incorrect.
3. Perseverative Errors: Perseverative responses that are also errors.

4. **4.** Conceptual Level Responses: Consecutive correct responses occurring in runs of three or more.
5. **5.** Number of Categories Completed: The total number of categories (i.e., a sequence of 10 consecutive correct matches to the correct sorting principle) that the examinee successfully completes during the test.
6. **6.** Failure to Maintain Set: Five or more consecutive correct matches followed by an error without successfully completing the category.

T-scores are reported for *Total Errors*, *Perseverative Responses*, *Perseverative Errors,* and *Number of Categories Completed*. Given the skewed distribution of scores and limited range, only percentile scores are reported for *Failures to Maintain Set*.

A subset of participants completed self-report questionnaires aimed at collecting information on educational, sociocultural and language use background. Current performance-based language fluency was assessed via administration of the Controlled Oral Word Association Test with letters F-A-S in English and P-M-R in Spanish (Artiola i Fortuny et al., 1999; Strauss et al., 2006). Based on the ratio of words produced in the FAS task to total words in both tasks (language fluency ratio = FAS/[FAS+PMR]; Miranda, Renteria, … & Rivera Mindt, 2016), we classified participants as Spanish dominant (scores ≤ 0.33) English-dominant (scores ≥ 0.67) and bilingual (scores:0.34 to 0.66). Please see Cherner, Marquine and colleagues (This Issue) for further details on the collection of educational, social, and language factors.

Statistical Analyses

We computed descriptive characteristics of raw scores for Total Errors, Perseverative Responses, Perseverative Errors, Number of Categories Completed, and Failures to Maintain Set. We examined the distribution of these raw scores via Shapiro-Wilk tests. We then examined the linear and non-linear association of age and education with WCST-64 raw scores via a series of univariable linear regression analyses and the association between sex and WCST-64 via independent sample t-tests (or Wilcoxon Rank Sum tests for variables with skewed distributions). To investigate interaction effects of demographics on raw scores, we ran a series of linear regression analyses with two-way interaction terms of demographics as predictors (i.e., age X education, age X sex, education X sex) on Total Errors, Perseverative Responses, Perseverative Errors, and Number of Categories Completed.

Raw scores were converted to normalized Scaled Scores, with a mean of 10 and SD of 3. T-scores for measures with adequate distribution were obtained by applying fractional polynomial equations (Royston & Altman, 1994) controlling for age, education, and gender. This method allows for the consideration of both linear and non-linear effects for numeric predictors (i.e., age and education) and selects the best curve (p<0.05) from several options: linear, quadratic, logarithmic, and other combinations of fractional polynomials of first (e.g. x^m) and second degree (e.g. $x^{m1} + x^{m2}$) with powers (m_i) ranging from −2 to +3. Further details can be found in Cherner, Marquine and colleagues (This Issue).

We then examined the descriptive characteristics of the resulting T-scores and their distributions via Shapiro-Wilk tests, and investigated the association of age and education with the newly developed T-scores via Pearson product moment correlation coefficients, and the association of sex with T-scores via independent samples t-tests. We also compared T-scores based on the newly developed norms by testing site (Arizona and California) and study wave (Cohorts 1 and 2) via a series of independent sample t-tests.

We calculated T-scores for the raw WCST-64 scores based on published norms for English-speaking non-Hispanic Whites and non-Hispanic Blacks/African Americans in the United States (Norman et al., 2011). Rates of neurocognitive impairment (NCI; T-scores <40) obtained in the normative sample using the existing norms (Norman et al., 2011) were compared to expected scores based on the new norms via McNemar's test.

Results

Demographic Characteristics of the Norming Sample

Demographic characteristics of the norming sample for this test stratified by years of education are presented in Table 1. Similar to the overall NP-NUMBRS cohort, age ranged from 19 to 60 years (*M*=38.23, *SD*=10.27), education ranged from 0 to 20 years (*M*=10.85, *SD*=4.49), and a little over half of the sample was female (59.26%).

Table 2 lists educational, social, and language background characteristics of a subset of participants in the study sample with available data on these variables. Most participants completed more years of education in their country of origin than in the U.S., and almost a third of the sample had to stop attending school to work. Years of education completed by both parents was on average 6-7. Participants lived a majority of their lives in their country of origin. Most participants described their childhood socioeconomic status as middle class, with nearly a third reporting having been poor. Half of participants reported working for money during childhood, and nearly half of those stated they did so to help their families financially. A little over two thirds of participants were gainfully employed at the time of their participation in the present study. All but one of the participants reported that Spanish was the first language they learned. Almost two-thirds of the sample was monolingual Spanish-speaking or strongly Spanish dominant, with the remaining third being bilingual. Average ratings of language used in various everyday activities indicated that Spanish was the predominant language used in daily life.

Raw Scores to Scaled Scores Conversions

Table 3 shows descriptive characteristics of raw scores on each of the WCST-64 scores. Results from Shapiro-Wilk tests indicated that none of the variables were normally distributed in raw scores. There were no significant non-linear associations of age and education to WCST-64 raw scores. Table 4 shows the association of raw scores with demographic variables, based on Spearman ρ (for age and education) and Wilcoxon rank-sum tests (for gender). There were small effects of age and moderate to large effects of education on all scores except for Failures to Maintain Set. There were no significant main effects of gender on any of the raw scores. Separate linear regression models on WCST-64 raw scores, entering terms for two-way interactions among demographics (age X education; age X gender; education X gender), showed significant education X gender interactions on Total Errors (p=.04, Figure 1a), Perseverative Responses (p=.002; Figure 1b), and Perseverative Errors (p=.005; Figure 1c) raw scores, with no other significant interactions. These results indicated that while higher levels of education were significantly associated with better WCST performance (as reflected by raw scores in these WCST-64 measures) in both genders, this association was stronger among females than males.

Table 5 shows the raw-to-scale score conversions for *Total Errors*, *Perseverative Responses, Perseverative Errors, Conceptual Level Responses,* and *Number of Categories Completed*. Given the skewness of the distribution and the limited range of scores of *Failures to Maintain Set*, percentile ranges are provided for this measure (see Table 6).

T-Scores Equations

Table 7 shows the T-score equations used to compute individual T-scores (please see digital supplemental material with these formulas to ease computation of T-scores). While the methodology used allowed for the consideration of non-linear effects of age and education, the model selected linear effects of these variables as the best fit.

As expected, the resulting T-scores had a mean of 50 and a SD of 10, and were normally distributed as assessed by Shapiro-Wilk tests. T-scores ranged from 23 to 82 for *Total Errors,* 25 to 76 for *Perseverative Responses*, 24 to 75 for *Perseverative Errors*, 23 to 82 for *Conceptual Level Responses*, and 23 to 78 for *Number of Categories Completed*. Pearson product moment correlation coefficients showed no significant effect of age or education on any of the T-scores and there were no significant gender differences. There were also no significant demographic interactions on any of the T-scores, except for an education X gender interaction on Total Errors (p=.02). However, stratified analyses by gender showed education was not significantly associated with T-scores on these WCST scores in either males or females (p = .10 in both groups), so no further analyses to account for an interaction were conducted. There were no significant differences by site or cohort on any of the T-scores (Table 8).

Applications of Existing Norms

Figure 2 shows the proportion of participants who are classified as impaired using the norms developed in the present study and existing norms for English-speaking non-Hispanic Whites and Blacks (Norman et al., 2011). As expected, applying the newly developed norms resulted in 14-16% impairment across measures. Applying norms for Blacks resulted in impairment rates close to what would be expected (10-19%), with no significant differences compared to rates of impairment using current norms, except for Perseverative Errors (p=.01). In contrast, applying norms for Whites overestimated impairment (38-52% across measures) compared to using currently developed norms (all ps<.0001). In follow-up analyses, we ran similar analyses by education group contrasting rates of impairment utilizing non-Hispanic White and current norms. We found that in applying non-Hispanic White norms, misclassification was most prominent in persons with 6 years of education or less (53-72%), but still notable in other education groups (7-10 years of education: 31-44%; 11-12 years of education: 38-47%; 13+ of education years: 29-48; all *ps*< .01).

Discussion

Neuropsychological test norms for Spanish-speakers in the U.S. are scarce. In the present study, we developed norms for a cohort of native Spanish speakers living in the U.S.-Mexico border region on the WCST-64 as part of a larger normative effort in this population. Current findings showed older age was associated with worse performance and education was linked to better performance on most WCST-64 raw scores, with stronger education effects among females than males. We developed demographically adjusted T-scores, which resulted in expected rates of impairment (defined as T-scores<40). While applying published norms for non-Hispanic Blacks to the current sample of Spanish-speakers resulted in generally comparable impairment rates, applying previously published norms for non-Hispanic Whites overestimated impairment.

Consistent with prior studies (Axelrod et al., 1993; Norman et al., 2011), we found that more years of education was significantly associated with better raw scores (performances) on the WCST-64, and increased age was associated with worse performance, with no significant differences by gender. We also found that the link between higher education and better WCST-64 performance was stronger in females than males. While the reasons for this interaction are largely unknown, they might mean that education is representing slightly different underlying constructs across genders and/or that other unmeasured factors play a more important role among males. Future studies investigating the impact on test performance of adherence to traditional gender roles within the Latino culture, might help explain these differential associations by gender.

Relatively unexpected findings with this Spanish-speaking cohort residing in the U.S.-Mexico borderland region were that the distribution of the raw "Categories Completed" score was not significantly skewed, and this score showed comparably high associations with age and (especially) education as the other WCST variables (Table 4). This is somewhat at odds with prior findings with English-speaking non-Hispanic Whites and African Americans in the U.S., whose Categories Completed scores were so highly skewed in the negative direction (toward higher Categories Completed) as to preclude normalization transformations and creation of normative T-scores. This difference may have been due to much higher representation of Spanish-speakers with very low levels of formal education (Table 1), which were rare among the English-speaking U.S. cohorts (Norman et al., 2011). In any event, the final T-score distribution among our Spanish-speakers was normal (not skewed) and provided substantial demographic correction (no education or age effects).

Applying existing norms for non-Hispanic Whites in the present population of Spanish-speakers overestimated impairment, particularly among participants with less or equal to 6 years of education. Of note, rates of impairment were still notably elevated among participants with more than a high school education. These results indicate that while years of formal education are an important factor explaining the difference in rates of impairment when utilizing different norms, other unmeasured factors are also important. Applying norms developed for non-Hispanic Blacks resulted in rates of impairment generally similar to what would be expected, indicating that the differences found when applying non-Hispanic White norms cannot be fully attributed to testing language. These findings underscore the importance of considering the cultural/linguistic background of examinees when selecting neuropsychological test norms in the process of identifying underlying brain impairment via neuropsychological testing. As is the case with most studies proving normative data, it is crucial to consider that our findings tell us very little about the factors underlying ethnic/racial differences in neuropsychological test performance. While region specific neuropsychological norms provide an important tool for clinicians and researchers in the identification of underlying neurological dysfunction, identifying the factors driving differences in performance is a crucial step in moving the field forward.

Limitations of our study warrant further discussion to better guide future research. Hispanics/Latinos living in the U.S. are a highly heterogeneous group with significant differences in dialect, culture, and national origin. However, norms from our study were developed from Spanish-speakers residing in the U.S.-Mexico border region, which is largely comprised of individuals of Mexican origin. Thus, caution should be used when applying the present norms to other Spanish-speaking individuals living in the U.S., as this may impact interpretation of results. The age range for this sample was restricted from 19 to 60 years. Application of these norms to individuals outside of this age may result in improper interpretation and should be used with caution. The present data were collected a number of years ago, and thus there could be cohort effects when applying the norms developed here that have not been accounted for. Of note, we found no significant cohort effects during the two waves of testing in the present study, providing some evidence that, if present, cohort effects might not be notable. Furthermore, clinicians and researchers might consider the educational, social and language background characteristics of participants in the present study in deciding whether the norms presented here might apply to their patients or population to be studied.

Although substantial sociodemographic details were collected from a subset of participants, we did not investigate the impact of these factors on WCST-64 performance in the present paper. Considering the impact of bilingualism on cognition (Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystock, 2017), a paper in this special issue investigates the association between bilingualism and performance on the entire NP-NUMBRS test battery (Suarez, Diaz-Santos, Marquine, Gollan, et al., This Issue). We are also currently investigating the link between the educational, social and language background characteristics presented in Table 2 and performance on the NP-NUMBRS test battery (including the WCST-64), which will be the focus of another report. Future studies might measure additional factors not assessed in the current project that might also influence test performance, such as acculturation, generational membership, country of origin/age of immigration for first-generation immigrants, and location/quality of education, among others. While efforts were made to have the sample include Spanish-speaking persons with varied years of education and equally distributed in terms of gender, the present study is not population-based, and as such, might have been impacted by selection bias. Future studies examining variables impacting neurocognitive test performance in population-based samples of Spanish-speakers in the US are important to assure representation of all segments of the

Spanish-speaking US population. Another paper in this special issue examines the utility of the entire battery (including the WCST-64) in detecting HIV-associated neurocognitive impairment (Kamalyan et al., This Issue). Further research examining the utility of the norms developed in the NP-NUMBRS project in detecting neurocognitive impairment in other neurological conditions would lend further support to the validity of the tests in the battery and the accompanying normative data among Spanish-speakers in the U.S.

Overall, the present study has significant implications for clinicians and researchers. The normative data developed provides clinicians a tool to more accurately detect cognitive impairment among Spanish-speakers living in the U.S., specifically those residing near the U.S.-Mexico border region, aged 19-60. The data presented adds to the growing literature on culturally-focused neuropsychology, which is crucial for improvement of future diagnostic validity and treatment in diverse groups.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by grants from the National Institutes of Health (P30MH62512, R01MH57266, K23MH105297, P30AG059299) and the UCSD Hispanic Center of Excellence.

References

Adesope OO, Lavin T, Thompson T, & Ungerleider C (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. Review of Educational Research, 80(2), 207–245.

Arango-Lasprilla JC, Rivera D, Longoni M, Saracho CP, Garza MT, Aliaga A, . . . Perrin PB (2015). Modified Wisconsin Card Sorting Test (M-WCST): Normative data for the Latin American Spanish speaking adult population. NeuroRehabilitation, 37(4), 563–590. doi:10.3233/NRE-151280 [PubMed: 26639931]

Artiola i Fortuny L, Heaton RK, & Hermosillo D (1998). Neuropsychological comparisons of Spanish-speaking participants from the U.S.-Mexico border region versus Spain. J Int Neuropsychol Soc, 4(4), 363–379. [PubMed: 9656610]

Artiola i Fortuny L, Hermosillo Romo D, Heaton RK, & Pardee RE (1999). Manual de Normas y Procedimientos para la Bateria Neuropsicologica en Espanol. Tucson, AZ: m Press.

Axelrod BN, Henry RR, & Woodard JL (1992). Analysis of an abbreviated form of the Wisconsin Card Sorting Test. The Clinical Neuropsychologist, 6, 27–31.

Axelrod BN, Jiron CC, & Henry RR (1993). Performance of Adults Ages 20 to 90 on the Abbreviated Wisconsin Card Sorting Test. Clinical Neuropsychologist, 7(2), 205–209. doi:Doi 10.1080/13854049308401523

Berg EA (1948). A simple objective test for measuring flexibility in thinking. Journal of General Psychology, 39, 15–22.

Bialystock E (2017). The bilingual adaptation: How minds accommodate experience. Psychological bulletin, 143(3), 233. [PubMed: 28230411]

Buchsbaum BR, Greer S, Chang WL, & Berman KF (2005). Meta-analysis of neuroimaging studies of the Wisconsin card-sorting task and component processes. Hum Brain Mapp, 25(1), 35–45. doi:10.1002/hbm.20128 [PubMed: 15846821]

Cherner M, Marquine MJ, Umlauf A, Morlett Paredes A, Rivera Mindt M, Suarez P, . . . Heaton R (This Issue). Neuropsychological Norms for the U.S.-Mexico Border Region in Spanish (NP-NUMBRS) Project: Methodology and Sample Characteristics. The Clinical Neuropsychologist.

Coffey DM, Marmol L, Schock L, & Adams W (2005). The influence of acculturation on the Wisconsin Card Sorting Test by Mexican Americans. Arch Clin Neuropsychol, 20(6), 795–803. doi:10.1016/j.acn.2005.04.009 [PubMed: 15916877]

Del Pino R, Peña J, Ibarretxe-Bilbao N, Schretlen D, & Ojeda N (2016). Test modificado de clasificación de tarjetas de Wisconsin: normalización y estandarización de la prueba en población española. Revista de Neurologia, 62, 193–202. [PubMed: 26916322]

Diaz-Santos M, Suarez P, Umlauf A, Marquine MJ, Rivera Mindt M, Artiola i Fortuny L, . . . Group, H. (This Issue). Updated Demographically Adjusted Norms for the Brief Visuospatial Memory Test-Revised and Hopkins Verbal Learning Test-Revised in Native Spanish Speakers from the U.S.-Mexico Border Region: NP-NUMBRS Project. The Clinical Neuropsychologist.

Gooding A, Seider T, Marquine MJ, Suarez P, Umlauf A, Rivera Mindt M, . . . Cherner M (This Issue). Demographically-adjusted norms for the Paced Auditory Serial Addition Test and Letter Number Sequencing Test in Spanish-Speakers from the US-Mexico Border Region. The Clinical Neuropsychologist.

Grant DA, & Berg EA (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. J Exp Psychol, 38(4), 404–411. [PubMed: 18874598]

Greve KW (2001). The WCST-64: A standardized short-form of the Wisconsin Card Sorting Test. Clinical Neuropsychologist, 15(2), 228–234. doi:Doi 10.1076/Clin.15.2.228.1901

Haaland KY, Vranes LF, Goodwin JS, & Garry PJ (1987). Wisconsin Card Sort Test performance in a healthy elderly population. J Gerontol, 42(3), 345–346. [PubMed: 3571874]

Heaton A, Gooding A, Cherner M, Umlauf A, Franklin D Jr., Rivera Mindt M, . . . Marquine MJ (This Issue). Demographically-Adjusted Norms for the Grooved Pegboard and Finger Tapping Tests in Spanish-Speaking adults: Results from the Neuropsychological Norms for the U.S.-Mexico Border Region in Spanish (NP-NUMBRS) Project. The Clinical Neuropsychologist.

Heaton RK (1981). A Manual for the Wisconsin Card Sorting Test: Psychological Assessment Services. Odessa, FL.

Heaton RK, Chelune GH, Talley JL, Kay GG, & Curtiss G (2012). Test de Clasificacion de Tarjetas de Wisconsin (Cruz A. b. M. V. d. l., Trans.): TEA Ediciones.

Heaton RK, Chelune GJ, Talley JL, Kay GG, & Curtis G (1993). Wisconsin Card Sorting Test (WCST) Manual Revised and Expanded. Odessa, FL: Psychological Assessment Resources.

Heaton RK, & Thompson LL (1992). Wisconsin Card Sorting Test: Is one deck as good as two? Journal of Clinical and Experimental Neuropsychology, 14, 63.

Kamalyan L, Hussain MA, Diaz MM, Umlauf A, Franklin D Jr., Cherner M, . . . Marquine MJ (This Issue). Neurocognitive Impairment in Spanish-speaking Latinos Living with HIV in the US: Application of NP-NUMBRS: Neuropsychological Norms for the US-Mexico Border Region in Spanish. The Clinical Neuropsychologist.

Kongs SK, Thompson LL, Iverson GL, & Heaton RK (2000). Wisconsin Card Sorting Test-64 Card Version: Professional Manual. Odessa, FL: Psycho-logical Assessment Resources.

Love JM, Greve KW, Sherwin E, & Mathias C (2003). Comparability of the standard WCST and WCST-64 in traumatic brain injury. Applied Neuropsychology, 10(4), 246–251. doi:Doi 10.1207/S15324826an1004_7 [PubMed: 14690806]

Marquine MJ, Morlett Paredes A, Madriaga C, Blumstein Y, Umlauf A, Kamalyan L, . . . Cherner M (This Issue). Demographically-Adjusted Norms for Selected Tests of Verbal Fluency in a Spanish-Speaking Adult Population: Results from the Neuropsychological Norms for the US-Mexico Border Region in Spanish (NP-NUMBRS) Project. The Clinical Neuropsychologist.

Miranda C, Arce Rentería M, Fuentes A, Coulehan K, Arentoft A, Byrd D, Rosario A, Monzones J, Morgello S, & Rivera Mindt M (2016). The relative utility of three English language dominance measures in predicting the neuropsychological performance of HIV+ bilingual Latino/a adults. The Clinical Neuropsychologist, 30, 165–184. PMID: 26934820 [PubMed: 26923937]

Morlett Paredes A, Carrasco J, Kamalyan L, Cherner M, Umlauf A, Rivera Mindt M, . . . Marquine MJ (This Issue). Demographically Adjusted Normative Data for the Halstead Category Test in a Spanish- Speaking Adult Population: Results from the Neuropsychological Norms for the US-Mexico Border Region in Spanish (NP-NUMBRS). The Clinical Neuropsychologist.

Morlett Paredes A, Gooding A, Artiola i Fortuny L, Rivera Mindt M, Heaton RK, Cherner M, & Marquine MJ (This Issue). The State of Neuropsychological Test Norms for Spanish-Speaking Adults in the United States. The Clinical Neuropsychologist.

Nelson HE (1976). A modified card sorting test sensitive to frontal lobe defects. Cortex, 12(4), 313–324. doi:10.1016/s0010-9452(76)80035-4 [PubMed: 1009768]

Norman MA, Moore DJ, Taylor M, Franklin D Jr., Cysique L, Ake C, . . . Group, H. (2011). Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. J Clin Exp Neuropsychol, 33(7), 793–804. doi:10.1080/13803395.2011.559157 [PubMed: 21547817]

Paolo AM, Axelrod BN, Troster AI, Blackwell KT, & Koller WC (1996). Utility of a Wisconsin card sorting test short form in persons with Alzheimer's and Parkinson's disease. Journal of Clinical and Experimental Neuropsychology, 18(6), 892–897. doi:Doi 10.1080/01688639608408310 [PubMed: 9157112]

Proctor A, & Zhang J (2008). Performance of three racial/ethnic groups on two tests of executive function: clinical implications for traumatic brain injury (TBI). NeuroRehabilitation, 23(6), 529–536. [PubMed: 19127006]

Purdon SE, & Waldie B (2001). A short form of the Wisconsin Card Sorting Test. Journal of Psychiatry & Neuroscience, 26(3), 253–256.

Rivera Mindt M, Marquine MJ, Aghvinian M, Scott T, Cherner M, Morlett Paredes A, . . . Heaton RK (This Issue). Demographically-Adjusted Norms for the Processing Speed Subtests of the WAIS-III in a Spanish-Speaking Adult Population Living in the U.S.-Mexico Border Region. The Clinical Neuropsychologist.

Royston P, & Altman DG (1994). Regression using fractional polynomials for continuous covariates: Parsimonious parametric modeling. Journal of the Royal Statistical Society, Series C 43(3), 429–467.

Ryan C (2013). Language Use in the United States: 2011. American Community Survey Reports. Retrieved 1 23, 2015 http://www.census.gov/prod/2013pubs/acs-22.pdf

Scott T, Morlett Paredes A, Taylor MJ, Umlauf A, Artiola i Fortuny L, Heaton RK, . . . Rivera Mindt M (This Issue). Demographically-Adjusted Norms for the WAIS-R Block Design and Arithmetic Subtests: Results from the Neuropsychological Norms for the US-Mexico Border Region in Spanish (NP-NUMBRS) Project. The Clinical Neuropsychologist.

Smith-Seemiller L, Franzen MD, & Bowers D (1997). Use of Wisconsin card sorting test short forms in clinical samples. Clinical Neuropsychologist, 11(4), 421–427. doi:Doi 10.1080/13854049708400472

Strauss E, Sherman EM, & Spreen O (2006). A compendium of neuropsychological tests: Administration, norms, and commentary. American Chemical Society.

Suarez P, Diaz-Santos M, Marquine MJ, Gollan T, Artiola i Fortuny L, Heaton RK, . . . Group, H. (This Issue). Role of English Fluency on Verbal and Non-Verbal Neuropsychological Tests in Native Spanish Speakers from the U.S.-Mexico Border Region using Demographically Corrected Norms. The Clinical Neuropsychologist.

Suarez P, Diaz-Santos M, Marquine MJ, Rivera Mindt M, Umlauf A, Heaton RK, . . . Cherner M (This Issue). Demographically Adjusted Norms for the Trail Making Test in Native Spanish- speakers from the U.S.-Mexico Border Region. The Clinical Neuropsychologist.

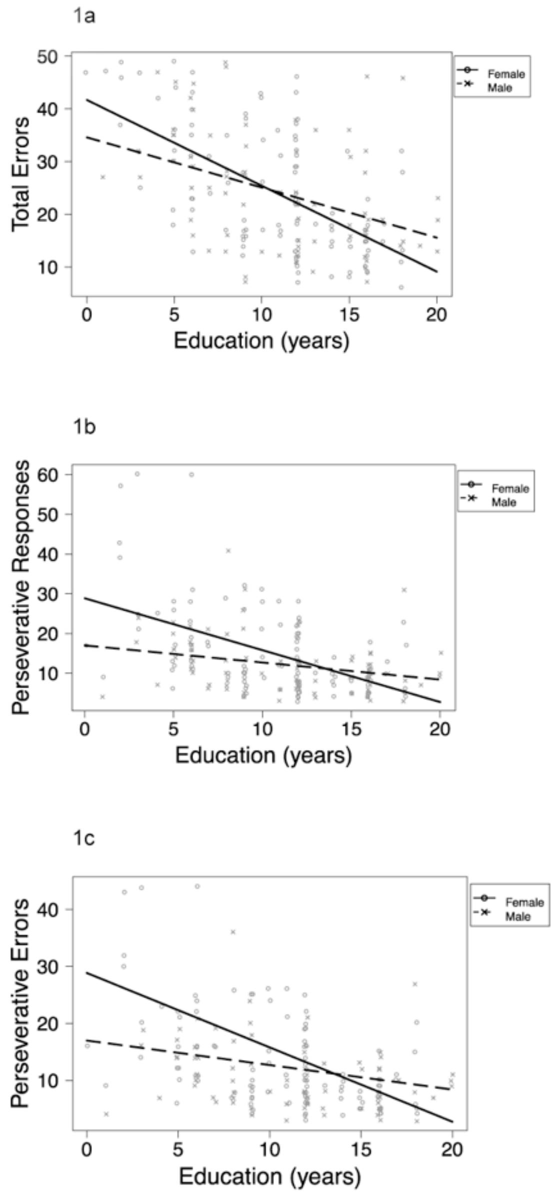

Figure 1.

Linear regression models on WCST-64 total errors (Fig. 1a), perseverative responses (Fig. 1b), and perseverative errors (Fig. 1c) with terms for education, gender and their interaction.

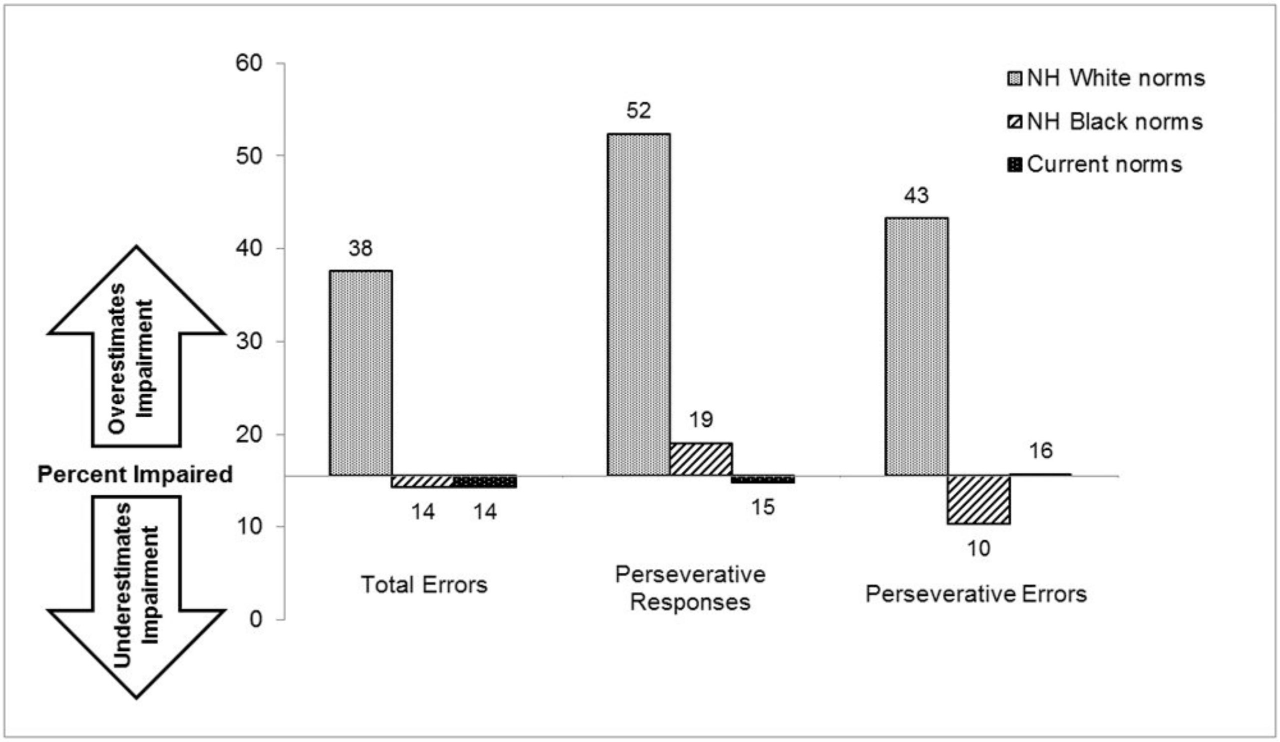

Figure 2.
Percent impaired (T-score<40) on WCST-64 total errors, perseverative responses, and perseverative errors using previously published norms in non-Hispanic (NH) Whites and Blacks (Norman et al., 2011) and newly developed norms.

Table 1.

Demographic characteristics of the normative sample stratified by years of education

	Years of Education			
	≤ 6 (n=43)	**7-10 (n=39)**	**11-12 (n=47)**	**≥13 (n=60)**
Age (years), M (SD)	40.44 (9.99)	38.44 (9.19)	35.51 (10.64)	38.65 (10.60)
Education (years), M (SD)	4.63 (1.68)	8.64 (0.90)	11.83 (0.38)	15.97 (1.70)
% Female	60.47	56.41	68.09	53.33

Table 2.

Educational, Social, and Language Background Characteristics of NP-NUMBRS Participants with Data on the WCST-64

Characteristics	Descriptives *M*(*SD*), %	*n*
Educational Background		
Years of education in country of origin	8.65 (4.80)	173
Years of education in the U.S.	2.50 (4.86)	173
Proportion of education by country	--	173
More years of education in country of origin	86.13%	149
More years of education in the U.S.	12.72%	22
Equal number of years of education in both countries	1.16%	2
Type of school attended[a]	--	179
Large	55.87%	100
Regular	39.11%	70
Small	5.03%	9
Number of students in the class	--	183
Less than 21	16.94%	31
21 to 30	37.16%	68
31 to 40	24.04%	44
40+	21.86%	40
Had to stop attending school to work	--	170
Yes	29.41%	50
Social Background		
Mother’s years of education	5.96 (3.86)	117
Father’s years of education	7.13 (5.22)	108
Years lived in country of origin	27.02 (12.42)	181
Years living in the U.S.	10.60 (10.97)	181
Childhood SES[b]	--	187
Very poor	5.35%	10
Poor	29.41%	55
Middle class	54.01%	101
Upper class	11.23%	21
Worked as a child	--	184
Yes	50%	92
Reason to work	--	92
Help family financially	40.22%	37
Own benefit	59.78%	55
Age started working as a child	12.52 (3.29)	89
Currently Gainfully Employed	--	164
Yes	67.07%	110
Language		

Characteristics	Descriptives *M*(*SD*), %	*n*
First Language	--	186
Spanish	98.39%	183
English	0.54%	1
Both	1.08%	2
Current Language Use Rating[c]		
Radio or TV	2.36 (1.03)	187
Reading	2.21 (1.16)	187
Math	1.50 (1.01)	185
Praying	1.29 (0.78)	179
With family	1.57 (0.91)	182
Performance-based language fluency		150
Spanish dominant	63.33%	95
English dominant	0.00%	0
Bilingual	36.67%	55

Note. *M*: mean; *SD*: standard deviation; SES: socioeconomic status

[a]Type of school attended: 'large' refers to large school that had many classrooms and room to play); 'regular' refers to a school of regular size that had at least one classroom per grade and room to play; and small school refers to a small school with less than one classroom per grade.

[b]Childhood SES was assessed by the following question and response options: "As a child, your family was: (1) Very Poor; (2) Poor; (3) Middle Class); (4) Upper Class".

[c]Ratings for each activity ranged from 1 "Always in Spanish" to 5 "Always in English", with 3 being "similarly in English and Spanish").

Table 3.

Mean, standard deviation, and range of the WCST Raw scores

	***Mean* (*SD*)**	***Range***
Total Errors	24.20 (11.56)	6-49
Perseverative Responses	13.82 (9.82)	3-60
Perseverative Errors	12.21 (7.67)	3-44
Conceptual Level Responses	33.15 (15.63)	0-58
Categories Completed	2.38 (1.48)	0-5
Failure to Maintain Set	0.52 (0.73)	0-4

Table 4.

Association between raw scores and demographic characteristics

	Age[a]	Education[a]	Gender		
			Male (*n*=77)	**Female (*n*=112)**	***p***[b]
Total Errors	0.21**	−0.51***	24.01 (11.41)	24.33 (11.70)	.85
Perseverative Responses	0.17*	−0.41***	12.21 (7.31)	14.93 (11.19)	.20
Perseverative Errors	0.19*	−0.42***	10.97 (6.08)	13.08 (8.53)	.19
Conceptual level Responses	−0.21**	0.51**	33.27 (15.80)	33.07 (15.58)	.86
Categories Completed	−0.19**	0.53**	2.45 (1.51)	2.33 (1.46)	.57
Failure to Maintain Set	0.03	0.02	0.45 (0.64)	0.56 (0.79)	.48

Note. Based on results from Spearman ρ[a] and Wilcoxon rank-sum tests[b].

* $p<.05$

** $p<.01$

*** $p<.001$

Table 5.

Raw-to-scale score conversions

Scaled	Total Errors	Perseverative Responses	Perseverative Errors	Conceptual Level Responses	Categories Completed
19	0	0	0	60	6
18	1-6	1	1	58-59	--
17	7	2-3	2-3	57	--
16	8	--	--	56	5
15	9	4	4	53-55	--
14	10-11	5	5	51-52	--
13	12-13	6	6	48-50	4
12	14-15	7-8	7	45-47	--
11	16-18	9	8-9	41-44	3
10	19-24	10-12	10-11	33-40	--
9	25-30	13-15	12-14	25-32	2
8	31-35	16-19	15-16	19-24	--
7	36-39	20-23	17-20	13-18	1
6	40-45	24-28	21-24	6-12	--
5	46	29-31	25-27	3-5	0
4	47-48	32-51	28-40	2	--
3	49-53	52-60	41-45	0-1	--
2	53-63	61-63	46-62	--	--
1	64	64	63-64	--	--

Table 6.

Raw scores to percentiles conversions for Failures to Maintain Set

Percentile	Failures to Maintain Set
60^{th}	0
91^{st}	1
98^{th}	2
99^{th}	3 or more

Table 7.

T-score equations

Measure	Equation
Total Errors	$10 \times \left(\frac{\text{SS Total Errors} - (8.23568 - 5.53473 * \frac{age}{100} + 3.29995 * \frac{(\text{edu} + 1)}{10} - 0.17484 * \text{gender})}{2.53754}\right) + 50$
Perseverative Responses	$10 \times \left(\frac{\text{SS Pers. Resp.} - (8.13820 - 4.14280 * \frac{age}{100} + 2.76561 * \frac{(\text{edu} + 1)}{10} + 0.44179 * \text{gender})}{2.72544}\right) + 50$
Perseverative Errors	$10 \times \left(\frac{\text{SS Pers. Errors} - (8.1984 - 3.99399 * \frac{age}{100} + 2.69882 * \frac{(\text{edu} + 1)}{10} + 0.54986 * \text{gender})}{2.7539}\right) + 50$
Conceptual level Responses	$10 \times \left(\frac{\text{SS Concep Resp} - (8.29546 - 5.33912 * \frac{age}{100} + 3.22828 * \frac{(\text{edu} + 1)}{10} - 0.17058 * \text{gender})}{2.50474}\right) + 50$
Categories Completed	$10 \times \left(\frac{\text{SS Categories} - (7.47484 - 4.69471 * \frac{age}{100} + 3.48816 * \frac{(\text{edu} + 1)}{10} + 0.01942 * \text{gender})}{2.57271}\right) + 50$

Note. These formulas should be applied to education level ranges from 0-20 and age 19-60. Using values outside these ranges might result in extrapolation errors. Gender: Male=1; Female=0

Edu=years of education; Age= years of age; Concep Resp = Conceptual Level Responses; Pers. Errors = Perseverative Errors; Per. Resp. = Perseverative Responses

Table 8.

Comparisons by study site and cohort on WCST-64 T-scores based on newly developed norms

	Study Site			**Study Cohort**		
	Arizona (n=59)	**California (n=130)**	**p**	**Cohort 1 (n=120)**	**Cohort 2 (n=69)**	**p**
Total Errors	50.37 (10.90)	49.87 (9.58)	.76	50.62 (10.23)	49.0 (9.53)	.28
Perseverative Responses	50.24 (10.39)	49.93 (9.83)	.85	50.60 (9.92)	49.03 (10.08)	.30
Perseverative Errors	50.72 (10.52)	49.66 (9.76)	.52	50.68 (9.94)	48.82 (10.04)	.23
Conceptual Level Responses	50.45 (10.67)	49.71 (9.66)	.65	50.40 (10.31)	49.15 (9.37)	.39
Categories	49.98 (11.09)	49.99 (9.53)	.99	49.96 (9.80)	50.04 (10.44)	.96