Published in final edited form as:

Clin Neuropsychol. 2013 August ; 27(6): 946–961. doi:10.1080/13854046.2013.796406.

Normative Performance on the Brief Smell Identification Test (BSIT) in a Multi-Ethnic Bilingual Cohort: A Project FRONTIER Study

Chloe Menon, M.A.¹, Holly James Westervelt, Ph.D.^{2,3}, Danielle R. Jahn, M.A.¹, Jeffrey A. Dressel, Ph.D.⁴, and Sid E. O'Bryant, Ph.D.^{5,6}

¹Department of Psychology, Texas Tech University, Lubbock, Texas ²Rhode Island Hospital, Providence, Rhode Island ³Alpert Medical School of Brown University, Providence, Rhode Island ⁴CSSI, Inc., Washington, DC ⁵Institute for Aging and Alzheimer's Disease Research, University of North Texas Health Science Center, Fort Worth, Texas ⁶University of North Texas Health Science Center, Department of Internal Medicine, Fort Worth, Texas

Abstract

The Brief Smell Identification Test (BSIT) is a commonly used measure of olfactory functioning in elderly populations. Few studies have provided normative data for this measure, and minimal data are available regarding the impact of sociodemographic factors on test scores. This study presents normative data for the BSIT in a sample of English- and Spanish-speaking Hispanic and non-Hispanic Whites. A Rasch analysis was also conducted to identify the items that best discriminated between varying levels of olfactory functioning, as measured by the BSIT. The total sample included 302 older adults seen as part of an ongoing study of rural cognitive aging, Project FRONTIER. Hierarchical regression analyses revealed that BSIT scores require adjustment by age and gender, but years of education, ethnicity, and language did not significantly influence BSIT performance. Four items best discriminated between varying levels of smell identification, accounting for 59.44% of total information provided by the measure. However, items did not represent a continuum of difficulty on the BSIT. The results of this study indicate that the BSIT appears to be well-suited for assessing odor identification deficits in older adults of diverse backgrounds, but that fine-tuning of this instrument may be recommended in light of its items' difficulty and discrimination parameters. Clinical and empirical implications are discussed.

Keywords

Olfactory functioning; norms; geriatrics; cognition

Introduction

Over the past several decades, increased empirical attention has been directed at evaluating the utility of olfactory measures in detecting and diagnosing neurodegenerative disorders

Address correspondence to: Sid E. O'Bryant, Ph.D., University of North Texas, Health Sciences Center, Department of Internal Medicine, Fort Worth, Texas, USA. Phone: (817)-735-2961. Sid.O'Bryant@unthsc.edu.

Author Contributions: All authors meet the criteria for authorship and contributed to the development of this work, including conception, data collection and analyses, as well as manuscript preparation. C. Menon was responsible for the design of this study, data extraction and analyses, as well as composition of the manuscript. S. O'Bryant, H. Westervelt, D. Jahn, and J. Dressel advised on the research design and analyses, while also contributing to the composition and review of the written product.

Conflict of Interest: No conflict of interest is present.

(Devanand et al., 2000; Kjellvik et al., 2007; Suzuki et al., 2004; Wilson, Arnold, Tang, & Bennett, 2006; Westervelt et al., 2003). Although these measures are often assumed to be “culture-free,” (Rosselli & Ardila, 2003), there has been little research investigating whether these measures can be appropriately applied to different ethnic, racial, and language populations.

The Brief Smell Identification Test (BSIT; Doty, Marcus & Lee, 1996) was developed as a quick tool to measure odor identification deficits. The BSIT is a brief, easily administered, and convenient instrument derived from the 40-item University of Pennsylvania Smell Identification Test (UPSIT; Doty, Shaman, & Dann, 1984). It is composed of twelve odorants embedded on scent strips that are released when scratched with a pencil tip. While the BSIT has been utilized among North-American (Wilson et al., 2007), Greek (Economou, 2003), Japanese-American (Graves et al., 1999), Norwegian (Kjellvik, Sando, Aasly, Engedal, & White, 2007), Mexican (Rodriguez-Violante, Lees, Cervantes-Arriaga, Corona, & Silveira-Moriyama, 2010), and Chinese samples (Wang et al., 2002), little work has been conducted to formally assess the impact of demographic factors on BSIT scores among culturally and linguistically diverse populations.

In addition to the dearth of studies evaluating the impact of sociodemographic factors on BSIT scores among diverse ethnic and language groups, there is also relatively little normative data available on this measure. Although normative data are provided with the test (Doty et al., 1996), they were extrapolated from the original 40-item University of Pennsylvania Smell Identification Test (Doty et al., 1984). This is a significant potential limitation to the extant normative data given that the distractor choices were altered in 9 of the 12 BSIT items. To our knowledge, the possible impact of this alteration has not been examined, though Hawkins et al. (2011) have provided some initial support for the use of the BSIT in a sample of African Americans. The authors demonstrate that BSIT performance is relatively resistant to sociodemographic influences including education in their cohort. However, the use of the BSIT has not yet been examined in a Hispanic sample, the largest ethnic minority group within the U.S. The importance of culturally-appropriate normative reference data has been increasingly recognized (Manly, Jacobs, Touradji, Small, & Stern, 2002), and has received increased empirical attention in the context of other smell tests including the UPSIT (e.g., Silveira-Moriyama, 2010) and the Sniffin’ Sticks test (Yuan, Lee, Lee, Lin, Shu, 2010). Efforts are needed to extend these findings to the BSIT in samples of varying sociodemographic backgrounds.

This study aimed to evaluate the use of the BSIT in a bilingual sample of Hispanics (primarily Mexican-Americans) and non-Hispanic Whites and to provide normative data for this population. The influence of various sociodemographic factors such as age, gender, education, language, and ethnicity on BSIT performance was examined. Lastly, item-level analyses were conducted using Rasch analysis. Rasch analyses have been applied to the UPSIT in previous studies, and have found that UPSIT items cover a range of difficulty (Minor et al., 2004; Lange, Donathan, & Hughes, 2002; Minor, Wright, & Park, 2004). However, no studies to date have analyzed the BSIT using a Rasch model. This study sought to replicate previous findings for the BSIT, identifying the items that best discriminate between varying levels of smell identification, as well as the items that provide the most information about the construct being measured (Thissen & Steinberg, 2009). This study also sought to identify a cutoff score to differentiate cognitively impaired from unimpaired participants using ROC curve analysis. It was hypothesized that BSIT performance would be resistant to sociodemographic influences including education, language, and ethnicity, and that most BSIT items would discriminate well at varying levels of difficulty on the smell identification ability continuum.

Materials and Methods

Participants

Data were analyzed from a sample of 302 Hispanic and non-Hispanic White older adults recruited as part of Project FRONTIER, an ongoing epidemiological study of rural health. Inclusion criteria are (1) age 40 and above; and (2) residing in one of the counties part of Project FRONTIER, including Cochran County and Parmer County, both located on the Texas-New Mexico border. These criteria are reflective of the nature of Project FRONTIER, which is an investigation of aging among rural-dwelling adults and older adults who may be at increased risk for cognitive and medical problems. Participant recruitment is conducted by community recruiters through brochures/flyers, presentations and events, and in-person or door-to-door solicitation. Project FRONTIER is conducted under an Institutional Review Board approved protocol that includes a standardized medical examination, clinical labs, interviews with participants and informants, and neuropsychological tests. The BSIT was administered as part of an extensive test battery used in Project FRONTIER. The medical exam and clinical labs were conducted by the local hospital and the interview and neuropsychological testing completed at the Project FRONTIER office at that same hospital by psychological technicians. Project FRONTIER research staff has received extensive training on interview and neuropsychological testing procedures with annual quality checks (to protect against experimenter drift) performed. The detailed interview collected information regarding demographic background, SES (current and past income as well as current and past occupational history), residential history, medical history, and included portions of the CDC Behavioral Risk Factor Surveillance System (BRFSS) questionnaire capturing information regarding hypertension, cholesterol, diabetes, cardiovascular disease, and cancer.

All participants signed written informed consent. Self-report of race and Hispanic origin (ethnicity) was collected using census methods. The full Project FRONTIER protocol has been described elsewhere (more detail provided in O'Bryant, Schrimsher, Johnson, & Zhang, 2011). Self-reported number of years of education was used as an indicator of educational attainment.

Approximately half of the Hispanic subgroup (48.5%) reported being of Mexican origin and the remainder indicated that they were born and raised in the United States (51.5%). A number of Hispanic participants described having greater mastery of Spanish than English (22.1%), and all other non-Spanish speakers of Hispanic origin endorsed English as their primary language (77.9%). The test instructions were provided in Spanish among participants who reported being more proficient in Spanish than English. All participants who reported adequate mastery of English were tested in this language.

For the acquisition of a well-defined "healthy" sample, all information obtained for each participant was systematically examined by consensus review, and participants were excluded if the data suggested the presence of cognitive impairment (e.g., current or preexisting diagnosis of dementia, mild cognitive impairment) and medical conditions susceptible of influencing neurocognitive functioning (e.g., histories of stroke, transient ischemic attack, head injury/concussion, seizures, Parkinson's disease, brain hemorrhage, and significant upper respiratory conditions). Participants with nasal congestion resulting from significant upper respiratory conditions (e.g., colds, allergies) were also excluded. The Project FRONTIER Consensus Diagnostic Committee is composed of an internist, a psychologist, and a neuropsychologist, and diagnoses are assigned according to published criteria (i.e., NINCDS-ADRDA possible/probable Alzheimer's disease, vascular dementia, "other" dementia [do not meet criteria for AD or VaD], mild cognitive impairment [Mayo Criteria], and "other cognitive impairment" not meeting one of the above categories).

“Cognitively normal” cases were consensus reviewed and deemed to have performed within normal limits on the cognitive test battery. The neuropsychological tests measured functioning in many domains including memory, executive functioning, visuospatial and constructional abilities, intellectual functioning, motor abilities, attentional and verbal abilities. The consensus diagnosis procedures have been more thoroughly described elsewhere (O’Bryant et al., 2011). Because Rasch analysis requires a continuous latent variable to meet assumptions of the analysis, no exclusion criteria applied to the participants included in the Rasch analyses.

Measures

The Brief Smell Identification Test (BSIT; Doty, Marcus, & Lee, 1996) includes 12 odorants embedded on scent strips and released when scratched with a pencil tip. Participants were presented a four-category multiple-choice questionnaire and asked to identify which smell corresponded to the scent strip for each odorant. The BSIT is a forced-choice test, with participants being instructed to identify each odorant even if no particular smell is perceived. The total olfaction score using the BSIT was defined as the number of odorants correctly identified out of the 12 tested, with higher scores designating better olfactory performance and abnormal olfactory functioning being defined as correctly identifying fewer than nine odorants (Doty, 2001). This test has been found to have good internal reliability and validity.

Statistical Analyses

Independent samples *t*-tests and chi-square analyses were conducted to examine any systematic differences between the two groups (cognitively impaired versus healthy) on basic sociodemographic variables including age, education, gender, and ethnicity. Any significant differences found on the demographic variables prompted their inclusion as covariates in subsequent analyses. Linear regression models were created to further delineate the relationship between demographic factors and BSIT scores. This was done to determine if and how the sample should be stratified based on these factors. In an effort to test for any significant differences in BSIT performance across groupings, one-way analyses of variance were conducted among the different age groups and an independent *t*-test was used with gender categories. Cronbach’s α coefficient was obtained to examine the level of agreement among the BSIT test items. Corrected item-total correlations were also calculated as a complementary index of consistency, which involve examining the correspondence between each item score and the composite total score extracted from all other items. Correlations and chi-square analyses were also conducted to examine the relationship between key sociodemographic variables and specific BSIT items. Finally, receiver operating characteristic (ROC) curve analyses were conducted to identify the best cutoff scores on the BSIT to identify cognitive impairment (dichotomized as cognitively healthy versus any cognitive impairment [all categories]), as the only cutoffs available to date are for smell abilities. ROC curves are commonly utilized to plot sensitivity (the likelihood of correctly detecting positive cases) against rates of false positives (1 - specificity, designating the proportion of correctly identified negative cases) for all possible cut-scores. These analyses used the Statistical Package for the Social Sciences Software (SPSS), version 16.0, and significance was set at $p < .05$.

Following these analyses to provide normative data for BSIT, a Rasch analysis was conducted with all participants included as one group, using the ltm package (Rizopoulos, 2006) for R (R Development Core Team, 2006). Rasch analysis assumes that all items are identically related to an underlying latent trait (e.g., smell identification as it relates to smell identification ability), and identifies the difficulty of each item along the latent trait, as well as how well each item discriminates between varying levels of the latent trait (Thissen &

Steinberg, 2009). A two parameter logistic model, which allows for both a difficulty score and a discrimination score per item, was most appropriate for the data. A likelihood ratio test revealed that the two-parameter logistic model was preferable to both the constrained and unconstrained models ($p < .001$). Akaike Information Criterion (AIC) for the constrained model was 2544.34; AIC for the unconstrained model was 2541.50, and AIC for the two-parameter logistic model was 2512.76. Using the two-parameter logistic model, we examined the discrimination scores, difficulty scores, item characteristic curves, and item information curves for each of the 12 items on the BSIT.

A Rasch analysis was first run with only cognitively healthy participants, but the goodness of fit statistics indicated that models did not have acceptable fit, regardless of whether they were constrained, unconstrained, two-parameter logistic, or introduced a guessing parameter. This is likely because an arbitrary cut-point was introduced in the latent trait when participants were excluded for cognitive impairment, and this analysis requires a continuous latent trait that represents all levels of performance to provide appropriate difficulty and discrimination parameters. Therefore, all participants were included in the Rasch analysis; this was the only way to ensure that all levels of difficulty would be represented, and the fit of the models was acceptable when all participants were included.

Results

Table 1 presents the descriptive characteristics of the sample across ethnicities. In the total sample of 302 individuals, the mean age of participants was 61.71 ($SD = 13.17$), with a mean of 10.29 ($SD = 4.33$) years of education. Forty-three percent of the sample was Hispanic and 57% was non-Hispanic White. The majority of the sample was women (68.2%).

A total of 108 (35.76%) participants were excluded from the BSIT normative sample based on the exclusion criteria described previously. Among the group of cognitively impaired participants, 89 met diagnostic criteria for MCI, 15 for dementia, and 4 for “other cognitive impairment.” The final normative sample included 194 participants with a mean age and education of 58.92 ($SD = 12.25$, range = 40–96) and 10.75 ($SD = 4.44$, range = 0–20), respectively. Cognitively impaired cases were older and less educated compared to the healthy sample.

A hierarchical regression analysis was conducted using only the cognitively healthy sample to determine the relative relationship of key sociodemographic variables with BSIT performance and to aid in establishing proper stratification criteria. Age, gender, years of education, ethnicity, and language of test administration were entered simultaneously as a first step. The full model was significant, accounting for 12.0% of the total variance in BSIT total scores, $F(5, 271) = 7.22$, $p < .001$. Both age and gender were significantly related to BSIT total scores, $\beta = -.04$, $t = -4.70$, $p < .001$ and $\beta = .60$, $t = 2.80$, $p < .01$, respectively. In contrast, years of education, $\beta = .02$, $t = .06$, $p = .95$, ethnicity, $\beta = .17$, $t = .54$, $p = .60$, and language $\beta = -.35$, $t = -1.13$, $p = .26$, did not predict BSIT performance after controlling for age and gender.

Based on these results, BSIT total scores were stratified by age and gender, with four distinct age subsamples created with fairly equivalent sample size representation across groups. The four age groups consisted of participants between the ages of 40 and 49, 50 and 59, 60 and 69, or 70 or more years old. The mean scores for the BSIT in the different age and gender groupings are reported in Table 2. One-way ANOVAs indicated significant differences in BSIT performance among the different age groupings among women ($F(3, 104) = 3.28$, $p = .02$) and men ($F(3, 44) = 2.82$, $p = .05$). Independent t-tests also revealed

significant BSIT performance differences across gender groups, $t(154)=-3.31$, $p=.001$. Additionally, the ROC curves were generated for each of these cells, and a score of 11 maximized sensitivity and specificity across all cells in terms of identifying cognitive impairment in our sample (see Table 3).

Inter-item correlations and identification rates were also examined. Cronbach's α coefficient was .60 in our sample. Significant differences in item identification rates were noted, with 98% of the cohort correctly identifying onion, but only 66% identifying turpentine. Correlation coefficients between specific items also varied substantially, with the lowest correlation coefficient observed for cinnamon (.11), and the highest noted for onion (.36). Item-total correlations for four items were significantly higher for women than men, and identification rates for five items were significantly and negatively correlated with age (See Table 4).

The Rasch analysis conducted with both cognitively healthy and cognitively impaired participants identified four items with good discrimination scores (i.e., paint thinner, gasoline, soap, and onion; see Table 5). Item characteristic curves verified this information, as the lines representing these four items had the steepest curves, indicating that responses discriminate well between varying levels of the latent trait (see Figure 1 for item characteristic curves). These four items provided 59.44% of all information assessed by BSIT. However, all four of these items had similar difficulty scores (see Table 5), indicating that they each measure approximately the same level of the latent trait. Other items had difficulty scores that differed from these four items, but had poor discrimination scores, indicating that they do not discriminate well between varying levels of the underlying latent trait (i.e., smell identification ability). This is reflected in the item information curves (see Figure 2), which show that the four curves representing the items that discriminate well overlap significantly with each other, meaning that they all measure similar difficulty levels along the latent trait. The flat or nearly flat lines are those items that do not discriminate well, though slight elevations were noted at other points along the ability axis, indicating that they measure different difficulty levels.

Discussion

This study provided normative data for BSIT in a previously unexamined bilingual sample of Hispanic and non-Hispanic Whites. Significant differences in BSIT performance were noted between men and women and among different age groups, with higher BSIT scores found in women and younger individuals. In contrast, years of education, ethnicity, and language were not significantly associated with BSIT performance. This pattern of results corroborates previous findings showing that BSIT scores are relatively resistant to sociodemographic factors, and suggests that this test can be applied to these different groups.

Age had a significant negative influence on BSIT performance, which is consistent with previous studies indicating that aging is associated with heightened deficits in olfaction (Murphy et al. 2002; Bramerson et al. 2004; Strauss, Sherman, & Spreen, 2006). In part, the relationship between aging and olfactory impairment has been attributed to degeneration of the olfactory bulb and olfactory epithelium (Bhatnagar, 1987; Duda et al., 1999), ossification of the olfactory foramina (Kalmey, Thewissen, Dluzen, 1998), and neuropathological changes connected with neurodegenerative diseases (Hawkes & Doty, 2009). Although progress has been made in understanding the relationship between aging and olfactory decline, more research is needed to further delineate the respective influences of peripheral versus cortically-based olfactory pathology on BSIT scores, as this

differentiation is essential to ensuring valid clinical interpretations (Doty et al., 1984; Martzke et al., 1997).

Our findings indicate that gender is a significant predictor of BSIT performance, with women found to outperform men on this measure. This pattern of results is congruent with previous research suggesting that women tend to be less vulnerable to smell deficiencies across the lifespan (e.g., Economou, 2003; Doty et al., 1985; Hawkins et al., 2011). Although the exact cause of this gender gap is not fully understood, gender differences in circulating hormones have been proposed to account for females' better sense of smell throughout the lifespan (e.g., Doty & Cameron, 2009; Doty, 1986). Further, discrepancies in verbal abilities have been found to mediate the relation between gender and olfactory recollection (Larsson, Lovden, & Nilsson, 2003) and could also have contributed to this finding.

Educational attainment did not significantly affect BSIT performance in our cohort. The lack of association between these two factors offers support to the literature and suggests that the BSIT is relatively resistant to sociodemographic factors (Doty et al., 1996; Goudsmit et al., 2003; Hawkins & Pearlson, 2011). The fact that this sample comprised a greater range of educational attainment compared to previous examinations is a strength of this study, as it does not merely replicate but also expands on previous reports by demonstrating that the BSIT is a valuable assessment tool in both highly educated and non-educated individuals.

Consistent with previous research, ethnicity did not account for a significant proportion of variance in BSIT performance (Hawkins et al., 2011), which suggests that the BSIT is appropriate for use across these ethnic groups. Similarly, language of administration did not significantly affect BSIT scores. To our knowledge, this study is the first to examine these effects in a bilingual sample of Hispanics and non-Hispanic Whites. Although more research is needed to replicate these findings with a larger sample size, these preliminary results are encouraging in as much as they suggest that the BSIT has value in assessing for odor identification deficits in diverse populations of varying ethnic and linguistic backgrounds.

Nonetheless, our follow-up examination of internal consistency and item performance revealed that the reliability of the BSIT is less than ideal, and that fine-tuning of this instrument may be recommended. Specifically, the internal consistency of the BSIT was moderate in our cohort, with a Cronbach's α coefficient of .60, which is comparable to previous reports relying on samples of African-Americans and Greek participants (Economou, 2003; Hawkins et al., 2011) and represents a substantial decrease from reliability estimates attributed to the UPSIT (Doty et al., 1996). While the relative brevity of this measure may account for this pattern, disparities in item identification rates and specific item correlations could also play a role in this issue.

In particular, important differences were noted in the psychometric performance of certain BSIT items, with the lowest reliability coefficient identified for turpentine (.66) and the highest coefficient for onion (.98). Turpentine was one of the two items that had two changes to the distractor choices when extrapolated from the UPSIT, which could have contributed to the lower reliability coefficients for this item. The effects of gender and age were relatively inconsistent across items. In addition, only four items appeared to discriminate well between those with varying levels of the underlying trait of smell identification, though these items all measure approximately the same level of difficulty in this trait. These four items (paint thinner, gasoline, soap, and onion) account for a majority of the information assessed by BSIT, which suggests that these items may be used when a very brief measure is needed, and fine-tuning of this instrument may therefore be recommended. In particular, items that assess for and discriminate well among other levels

of difficulty should be developed to ensure that the BSIT assesses for the full continuum of ability. The current version of the BSIT is appropriate for use among adults and older adults, but may cause undue participant burden as eight items do not provide much information about the participants' ability. It is important to note that the sample used for the Rasch analyses included both cognitively healthy and cognitively impaired participants, allowing for a model that fit the data well but also potentially suggesting that these results may differ in samples that are only cognitively healthy or only cognitively impaired. However, objective measurement is an important component of data collected for Rasch analyses (Thissen & Steinberg, 2009) and previous research suggests that the difficulty estimates provided by Rasch analyses are invariant across samples (Slinde & Linn, 1979).

Results from the ROC analyses identified 11 as a recommended cutoff score for identifying cognitively impaired from cognitively healthy individuals in this sample. Sensitivity and specificity are commonly used to evaluate clinical measures, with a balance of both suggesting greater detection power. Blake et al. (2002)'s recommendations suggest sensitivity indices higher than 80% as good with specificity values above 60% to be appropriate, though the optimal balance is dependent on the use of the test in a given setting along with the base rate of the given condition. Although the diagnostic performance in this sample was generally appropriate based on these criteria, it should be noted that the sensitivity and specificity values were less than ideal among some of the age and gender subgroups, including among men in the 50 to 59 age range (sensitivity and specificity = .667) and among women in the 60 to 69 age range (sensitivity = .643). Thus, these results suggest that the BSIT alone would not be a sufficient screener of cognitive impairment among these specific subgroups. Furthermore, future studies focusing on the fine-tuning of this instrument should take into account the complex interplay of various sociodemographic variables including age and gender in determining the diagnostic performance and differential item functioning profiles of such abbreviated versions.

Future studies are needed to replicate these findings in other subgroups as well. Our results parallel previous studies (e.g., Kjelvik et al., 2007) and underscore the need for efforts to examine the utility of BSIT items among different subgroups. In particular, the current findings underscore the need to explore the potential of a shortened version of the BSIT excluding these items found to have relatively low discriminatory power (e.g., turpentine and lemon). Importantly, previous item analysis research indicated different patterns of BSIT item sensitivities for varying disease states (e.g., Double et al., 2003). Given this, it is possible that the item composition specified in this study may not apply well to other groups. Thus, future research should further examine and cross-validate the specific misidentification patterns associated with different disease states and sociodemographic domains before being utilized as a point of reference in clinical and research settings.

While the current study has a number of strengths and contributes significantly to the literature, this study has limitations. First, future studies should replicate this investigation with larger samples, as the sample size in each cell was relatively low. Additionally, it is possible that our specified subgroups encompassed significant heterogeneity in culture, primary language, immigration history, and socioeconomic status. As such, efforts are needed to replicate and expand on these findings while accounting for these potentially influential dimensions. In particular, it is possible that some of the participants had limited knowledge or previous exposure to certain target odors, such as turpentine, and that this lack of familiarity influenced their performance on certain items. Future research should seek to rely on corroborating assessments of relevant odor knowledge or exposure (e.g., the Picture Identification Test) in an effort to minimize this possible source of bias. In addition, future studies should examine the role of bilingualism in BSIT performance, as dual linguistic mastery has been consistently shown to influence neurocognitive functioning (Bialystok,

Craik, Green, & Gollan, 2010). Furthermore, the fact that the screening tests utilized to delineate our healthy sample were not primarily created or validated for use in this diverse cohort is also of concern, as it may have inadvertently introduced construct, item, or method biases (Ardila & Moreno, 2001; Greenfield, 1997; Rogoff & Chavajay, 1995; Van de Vijver & Tanzer, 1997). This underscores the need for continued efforts to cross-validate various neuropsychological assessments in heterogeneous populations.

Lastly, it is possible that discrepancies in quality of education were present among the subgroups of interest, as different sociodemographic subgroups may be exposed to very different educational contexts (Manly, Jacobs, Touradji, Small, & Stern, 2002). In turn, this may have resulted in a somewhat biased representation of the relationship between education and BSIT performance. While quantity of education was included in the present analyses, quality of education has been offered as a more precise operationalization of educational attainment as it accounts for variability in educational experiences (Cosentino, Manly, & Mungas, 2007; Manly et al., 2002). Thus, future research should examine the effect of quality of education on BSIT performance among diverse samples.

Despite these limitations, this study is to our knowledge the first to provide normative data for the BSIT in a sample of English- and Spanish-speaking Hispanics and non-Hispanic Whites, and was the first to apply a Rasch analysis to the BSIT. Data was also provided for a more comprehensive age and education range compared to previous normative studies on this measure (e.g., Hawkins et al., 2011). As noted earlier, this represents an important first step in moving towards more socioculturally adapted and valid diagnostic decision-making. The results of this study indicate that the BSIT appears to be well-suited for assessing odor identification deficits in older adults of diverse backgrounds, but that fine-tuning of this instrument may be recommended in light of its item's difficulty and discrimination parameters.

Acknowledgments

Funding: Research reported in this publication was supported by the National Institutes of Health under Award Number L60MD001849, R01AG039389, and P30AG12300. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research was also funded by the Environmental Protection Agency (RD834794) and, in part, by grants from the Hogg Foundation for Mental Health (JRG-040 & JRG-149) and the National Academy of Neuropsychology. Project FRONTIER is supported by the F. Marie Hall Institute for Rural and Community Health and Garrison Institute on Aging. The authors would additionally like to thank the Project FRONTIER participants and research team.

References

- Albert MS, Jones K, Savage CR, Berkman L, Seeman T, Blazer D. Predictors of cognitive change in older persons: MacArthur Studies of Successful Aging. *Psychological Aging*. 1995; 10:578–589.
- Anthony JC, LeResche L, Niaz U, von Korff MR, Folstein ME. Limits of the “Mini-Mental State” as a screening test for dementia and delirium among hospital patients. *Psychological Medicine*. 1982; 12:397–408. [PubMed: 7100362]
- Ardila A, Moreno S. Neuropsychological test performance in Arauco Indians: An exploratory study. *Journal of the International Neuropsychological Society*. 2001; 7:510–515.
- Barrie MA. Objective screening tools to assess cognitive impairment and depression. *Topics in Geriatric Rehabilitation*. 2002; 18(2):28–46.
- Bhatnagar KP, Kennedy RC, Baron G, et al. Number of mitral cells and the bulb volume in the aging human olfactory bulb: A quantitative morphological study. *Anatomical Record*. 1987; 218:73–87. [PubMed: 3605663]
- Blake H, McKinney M, Treece K, Lee E, Lincoln N. An Evaluation of Screening Measures for Cognitive. *Age and Ageing*. 2002; 31:451–456. [PubMed: 12446291]

- Bialystok E, Craik FIM, Green DW, Gollan TH. Bilingual minds. *Association for Psychological Science*. 2010; 10(3):89–129.
- Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*. 1991; 82:239–259. [PubMed: 1759558]
- Chaves ML, Camozzato AL, Godinho C, Kochhann R, Schuh A, de Almeida VL, Kaye J. Validity of the Clinical Dementia Rating Scale for the detection and staging of dementia in Brazilian patients. *Alzheimer Disease and Associated Disorders*. 2007; 21:210–217. [PubMed: 17804953]
- Cosentino S, Manly J, Mungas D. Do reading tests measure the same construct in multilingual and multiethnic elders? *Journal of International Neuropsychological Society*. 2007; 13:228–236.
- Devanand DP, Michaels-Marston KS, Liu X, et al. Olfactory deficits in patients with mild cognitive impairment predict Alzheimers disease at follow up. *American Journal of Psychiatry*. 2000; 157:1399–1405. [PubMed: 10964854]
- Double KL, Rowe DB, Hayes M, Chan DK, Blackie J, Corbett A, Halliday GM. Identifying the pattern of olfactory deficits in Parkinson disease using the Brief Smell Identification Test. *Archives of Neurology*. 2003; 60:545–549. [PubMed: 12707068]
- Doty, RL. *The Brief Smell Identification Test Administration Manual*. Philadelphia, PA: Sensonics, Inc; 2001.
- Doty, RL. Gender and endocrine-related influences upon olfactory sensitivity. In: Meiselman, HL.; Rivlin, RS., editors. *Clinical measurement of taste and smell*. New York, NY: MacMillan; 1986. p. 377–413.
- Doty RI, Applebaum S, Zusho H, et al. Sex differences in odor identification ability: A cross-cultural analysis. *Neuropsychologia*. 1985; 23:667–672. [PubMed: 4058710]
- Doty RL, Cameron EL. Sex differences and reproductive hormone influences on human odor perception. *Physiology & Behavior*. 2009; 97:213–228. [PubMed: 19272398]
- Doty RL, Marcus A, Lee WW. Development of the 12-Item Cross-Cultural Smell Identification Test (CC-SIT). *Laryngoscope*. 1996; 106:353–356. [PubMed: 8614203]
- Doty RL, Shaman P, Applebaum S, Giberson R, Sikorski L, Rosenberg L. Smell identification ability: Changes with age. *Science*. 1984b; 226:1441–1443. [PubMed: 6505700]
- Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Test: Standardized microencapsulated test for olfactory function. *Physiological Behaviour*. 1984a; 32:489–502.
- Duda JE, Shah U, Arnold SE, et al. The expression of alpha-, beta-, and gamma-synucleins in olfactory mucosa from patients with and without neurodegenerative diseases. *Experimental Neurology*. 1999; 160:515–522. [PubMed: 10619569]
- Economou A. Olfactory identification in elderly Greek people in relation to memory and attention measures. *Archives of Gerontology and Geriatrics*. 2003; 37:119–130. [PubMed: 12888225]
- Espino DV, Lichtenstein MJ, Palmer RF, Hazuda HP. Ethnic differences in Mini-mental State Examination (MMSE) scores: Where you live makes a difference. *Journal of Geriatric Society*. 2001; 49:538–548.
- Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12(3):189–198. [PubMed: 1202204]
- Good KP, Martzke JS, Daoud M, Kopala LC. Unirhinal Norms for the University of Pennsylvania Smell Identification Test. *The Clinical Neuropsychologist*. 2003; 17(2):226–234. [PubMed: 13680430]
- Goudsmit N, Coleman E, Seckinger RA, Wolitzky R, Stanford AD, Corcorana C, Goetz RR, Malaspina D. A brief smell identification test discriminates between deficit and non-deficit schizophrenia: Olfactory tests in the diagnosis of essential tremor. *Parkinsonism & Related Disorders Psychiatry Research*. 2003; 120(2):155–164.
- Graves A, Bowen JD, Rajaram LL, McCormick WC, McCurry SM, Schellenberg GD, Larson EB. Impaired olfaction as a marker for cognitive decline: Interaction with apolipoprotein E ϵ 4 status. *Neurology*. 1999; 53(7):1480–1487. [PubMed: 10534255]
- Greenfield PM. You can't take it with you. Why ability assessments don't cross cultures. *American Psychologist*. 1997; 52:1115–1124.

- Hawkes, CH.; Doty, RL. The neurology of olfaction. Cambridge, MA: Cambridge University Press; 2009.
- Hawkins KA, Pearlson GD. Age and gender but not common chronic illnesses predict odor identification in older African Americans. *American Journal of Geriatric Psychiatry*. 2011; 19(9): 777–82. [PubMed: 21873833]
- Kalmey JK, Thewissen JG, Dluzen DE. Age-related size reduction of foramina in the cribriform plate. *Anatomical Record*. 1998; 251:326–329. [PubMed: 9669759]
- Kjelvik G, Sando SB, Aasly J, Engedal KA, White LR. Use of the Brief Smell Identification Test for olfactory deficit in a Norwegian population with Alzheimer's disease. *International Journal of Geriatric Psychiatry*. 2007; 22:1020–1024. [PubMed: 17299804]
- Lange R, Donathan CL, Hughes LF. Assessing olfactory abilities with the University of Pennsylvania Smell Identification Test: A Rasch scaling approach. *Journal of Alzheimer's Disease*. 2002; 4:77–91. Retrieved from <http://www.j-alz.com/>.
- Larsson M, Lovden M, Nilsson LG. Sex differences in recollective experience for olfactory and verbal information. *Acta Psychologica*. 112:89–103. [PubMed: 12423901]
- Manly JJ, Mayeux R. Ethnic differences in dementia and Alzheimer's disease. Critical perspectives on racial and ethnic differences in health in late life. 2004
- Strauss, E.; Sherman, EMS.; Spreen, O. A Compendium of Neuropsychological Tests: Administration, norms, and commentary. 3. Oxford, England: Oxford University Press; 2006.
- Manly JJ, Jacobs DM, Sano MM, Bell KK, Merchant CA, Small SA, Stern YY. Cognitive test performance among nondemented elderly African Americans and Whites. *Neurology*. 1998; 50(5): 1238–1245. [PubMed: 9595969]
- Manly JJ, Jacobs DM, Touradj P, Small SA, Stern Y. Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*. 2002; 8:341–348. [PubMed: 11939693]
- Martzke JS, Kopala LC, Good KP. Olfactory dysfunction in neuropsychiatric disorders: Review and methodological considerations. *Biological Psychiatry*. 1997; 42:721–732. [PubMed: 9325566]
- Minor KL, Wright BD, Park S. The Smell Identification Test as a measure of olfactory identification ability in schizophrenia and healthy populations: A Rasch psychometric study. *Journal of Abnormal Psychology*. 2004; 113:207–216.10.1037/0021-843X.113.2.207 [PubMed: 15122941]
- Murphy C, Jernigan TL, Fennema-Notestine C. Left hippocampal volume loss in Alzheimer's disease is reflected in performance on odor identification: A structural MRI study. *Journal of International Neuropsychological Society*. 2003; 9:459–471.
- O'Bryant SE, Schrimsher GW, Johnson LA, Zhang Y. The impact of sociodemographic factors on neuropsychological test performance in a rural sample: A Project FRONTIER study. *Journal of Rural Community Psychology*. 2011; E14(1)
- Oquendo-Jiménez I, Mena R, Antoun MD, Wojna V. Linguistic adaptation of the Clinical Dementia Rating Scale for a Spanish speaking population. *Puerto Rico Health Sciences Journal*. 2010; 29(2): 102–108. [PubMed: 20496524]
- Pedersen NL, Reynolds CA, Gatz M. Sources of variation among Mini-Mental State Examination scores, education, and cognitive abilities. *Journal of Gerontology: Psychological Sciences*. 1996; 51B:55–63.
- R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2006.
- Rizopoulos D. ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*. 2006; 17:1–25. Retrieved from <http://www.jstatsoft.org/v17/i05/>.
- Rodriguez-Violante M, Lees AJ, Cervantes-Arriaga A, Corona T, Silveira-Moriyama. Use of smell test identification in Parkinson's disease in Mexico: a matched case-control study. *Movement Disorders*. 2011; 26(1):173–6. [PubMed: 20842690]
- Rogoff B, Chavajay P. What's become of the research on the cultural bias of cognitive development? *American Psychologist*. 1995; 50:859–877.
- Rosselli M, Ardila A. The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain And Cognition*. 2003; 52(3):326–333. [PubMed: 12907177]

- Silveira-Moriyama L, Azevedo AS, Ranvaud R, Barbosa ER, Doty RL, Lees AJ. Applying a new version of the Brazilian-Portuguese UPSIT smell test in Brazil. *Arquivos De Neuro-Psiquiatria*. 2010; 68(5):700–705. [PubMed: 21049178]
- Suzuki Y, Yamamoto S, Umegaki H, et al. Smell identification test as an indicator for cognitive impairment in Alzheimer's disease. *International Journal of Geriatric Psychiatry*. 2004; 19:727–733. [PubMed: 15290695]
- Thissen, D.; Steinberg, L. Item response theory. In: Millsap, RE.; Maydeu-Olivares, A., editors. *The SAGE handbook of quantitative methods in psychology*. Thousand Oaks, CA: SAGE; 2009. p. 148-177.
- Van de Vijver F, Tanzer N. Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*. 1997; 47:263–279.
- Wang QS, Tain L, Huang YL, Qin S, He LQ, Zhou JN. Olfactory identification and apolipoprotein E epsilon 4 allele in mild cognitive impairment. *Brain Research* 27. 2002; 951(1):77–81.
- Westervelt H, Stern RA, Tremont G. Odor identification deficits in diffuse Lewy body disease. *Cognitive And Behavioral Neurology*. 2003; 16(2):93–99. [PubMed: 12799595]
- Wilson RC, Arnold SE, Tang Y, Bennett DA. Odor identification and decline in different cognitive domains in old age. *Neuroepidemiology*. 2006; 26:61–67. [PubMed: 16352908]
- Wilson RS, Arnold SE, Schneider JA, et al. The relationship between cerebral Alzheimer's disease pathology and odour identification in old age. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2007; 78:30–35.
- Wilson RS, Schneider JA, Arnold SE, Tang Y, Boyle PA, Bennett DA. Olfactory identification and incidence of mild cognitive impairment in older age. *Archives Of General Psychiatry*. 2007; 64(7): 802–808. [PubMed: 17606814]
- Yuan B-C, Lee P-L, Lee Y-L, Lin S-H, Shu C-H. Investigation of the Sniffin' Sticks Olfactory Test in Taiwan and Comparison With Different Continents. *Journal of Chinese Medical Association*. 2010; 73(9)

Key Points

Adequate normative data is needed for neuropsychological assessments with diverse population. The BSIT appears to be well-suited for assessing odor identification deficits in older adults of diverse backgrounds, although fine-tuning may be recommended based on its item profiles.

Figure 1.

Item characteristic curves, showing discrimination scores for each BSIT item. Top plot provides curves for selected items only (for clarity) and bottom plot provides curves for all 12 items (for comparison to non-selected items).

Figure 2.

Item information curves, showing difficulty scores for each BSIT item. Top plot provides curves for selected items only (for clarity) and bottom plot provides curves for all 12 items (for comparison to non-selected items).

Table 1

Descriptive Statistics of BSIT

Attributes	Complete Sample (<i>n</i> = 302)	Cognitively Impaired Sample (<i>n</i> = 108)	Total Healthy Sample (<i>n</i> = 194)	Healthy Subgroup Hispanic (<i>n</i> = 81)	Healthy Subgroup Non-Hispanic Whites (<i>n</i> = 113)	Comparison of Total Healthy Sample to Cognitively Impaired	<i>p</i>
Age, <i>M</i> (<i>SD</i>), <i>rg.</i>	61.71(13.17), 40–96	66.7(13.36), 40–94	58.92(12.25), 40–96	54.60(9.82), 40–79	63.0(13.0), 40–96	$t(300) = -5.11$	<.001
Education, <i>M</i> (<i>SD</i>), <i>rg.</i>	10.29(4.33), 0–20	9.46(4.0), 0–16	10.75(4.44), 0–20	6.99(3.70), 0–15	13.62(2.61), 6–20	$t(300) = 2.49$	<.05
Hispanics (%)	43.0	45.4	41.8	--	--	$\chi^2(1, N = 277), d = 1.54$	0.52
Females, %	68.2	68.5	68.0	69.1	66.7	$\chi^2(1, N = 302), d = 0.01$	0.13
MMSE, <i>M</i> (<i>SD</i>), <i>rg.</i>	27.1 (3.10), 12–30	22.85 (2.74), 12–23	28.05 (2.24), 24–30	27.10 (2.70), 24–30	29.0 (1.50), 24–30	$t(300) = 7.77$	<.001
CDR, <i>M</i> (<i>SD</i>), <i>rg.</i>	0.2 (0.2), 0–1	0.30 (0.25), 0–1	0.12 (0.21), 0–0.5	0.09 (0.20), 0–0.5	0.14 (0.23), 0.0.5	$t(300) = -6.63$	<.001
BSIT, <i>M</i> (<i>SD</i>), <i>rg.</i>	10.2 (2.00), 2–12	9.40 (2.24), 2–12	10.63 (1.42), 4–12	11.00 (1.30), 7–12	10.66 (1.50), 4–12	$t(300) = 5.97$	<.001

Notes: *M* = mean; *SD* = standard deviation; *rg.* = range; MMSE = Mini Mental Status Exam; CDR = Clinical Dementia Rating scale; BSIT = Brief Smell Identification Test. χ^2 = chi-square; *d* = Cohen's *d*. *p*-values obtained for comparison of those excluded and the normative sample using *t*-test or chi-square analysis.

Table 2

Normative Performance of BSIT

	Age							
	40-49		50-59		60-69		70+	
	Women	Men	Women	Men	Women	Men	Women	Men
n	37	13	35	18	29	17	31	14
Mean	10.67	10.76	10.94	10.61	11.14	10.06	10.32	9.93
SEM	0.20	0.28	0.21	0.38	0.14	0.42	0.35	0.24
SD	1.27	1.01	1.28	1.64	1.00	1.75	1.94	1.00
Range	8-12	9-12	7-12	7-12	10-12	6-12	4-12	8-11
Median	11	11	11	11	11	11	11	10
5%	8	9	8	7	10	6	5	8
10%	9	9	9	7	10	7	7	8
25%	10	10	10	10	10	9	10	9
50%	11	11	11	11	11	11	11	10
75%	12	12	12	12	12	11	11	11
90%	12	12	12	12	12	12	12	11
95%	12	12	12	12	12	12	12	11

Table 3

ROC Curve Analyses Results.

	Age							
	40–49		50–59		60–69		70+	
	Women	Men	Women	Men	Women	Men	Women	Men
Recommended Cutoff	11	11	11	11	11	11	11	11
Sensitivity	.800	1.000	.813	.667	.643	1.000	.846	.947
Specificity	.684	.692	.600	.667	.621	.875	.800	1.000

Table 4

Item Identification Rates, Gender and Age Effects in BSIT Total Scores

Item #	Odor	Correct Identification Rates	Corrected Item-Total Correlation	Gender Differences	Correlation with Age
12	Onion	98.0	0.36	$\chi^2(1, N=194), d = 1.68, p = .19$	$-0.13, p = 0.40$
11	Soap	98.0	0.20	$\chi^2(1, N=194), d = 6.49, p < .01^{**}$	$-0.12, p = .11$
10	Gasoline	98.0	0.18	$\chi^2 = (1, N=194), d = 6.49, p < .01^{**}$	$-0.21, p < .01^{**}$
7	Paint Thinner	96.0	0.15	$\chi^2 = (1, N=194), d = 9.65, p < .01^{**}$	$-0.20, p < .05^*$
5	Chocolate	95.0	0.15	$\chi^2 = (1, N=194), d = 1.58, p = .21$	$-0.20, p < .05^*$
1	Cinnamon	95.0	0.11	$\chi^2 = (1, N=194), d = 0.41, p = .52$	$-0.03, p = 0.66$
4	Smoke	95.0	0.22	$\chi^2 = (1, N=194), d = 0.67, p = .41$	$-0.24, p < .01^{**}$
9	Pineapple	90.0	0.25	$\chi^2 = (1, N=194), d = 0.31, p = .58$	$-0.12, p = .13$
3	Lemon	81.0	0.21	$\chi^2 = (1, N=194), d = 0.04, p = .84$	$-0.17, p < .05^*$
8	Banana	75.0	0.25	$\chi^2 = (1, N=194), d = 0.01, p < .90$	$-0.14, p = .07$
6	Rose	73.0	0.36	$\chi^2 = (1, N=194), d = 4.92, p < .05^*$	$-0.04, p = .62$
2	Turpentine	66.0	0.19	$\chi^2 = (1, N=194), d = 0.38, p = .53$	$-0.01, p = .92$

Note: χ^2 -chi-square; d = Cohen's d ; p -values obtained for comparison of gender groups using chi-square analyses, and correlation analyses.* $p < .05$, two-tailed;** $p < .01$, two-tailed;*** $p < .001$, two-tailed.

Table 5

Rasch Analysis Item Scores.

Item	Discrimination Score	Difficulty Score
Cinnamon	0.8412	-4.1786
Turpentine	0.5678	-0.8465
Lemon	0.7599	-1.7447
Smoke	1.3347	-2.3434
Chocolate	1.3804	-2.1187
Rose	1.2183	-0.7658
Paint Thinner	2.1871 *	-1.8082
Banana	1.0006	-1.0559
Pineapple	1.1506	-2.0754
Gasoline	3.7693 *	-1.9386
Soap	2.7896 *	-1.9152
Onion	3.2838 *	-1.9497

Note:

*
Selected items