A retrospective comparison of cognitive performance in individuals with advanced Parkinson's Disease in Hong Kong and Canada

Abstract

A deeper understanding of the cross-cultural applicability of cognitive tests across countries and cultures is needed to better equip neuropsychologists for the assessment of patients from diverse backgrounds. Our study compared cognitive test scores in patients with advanced Parkinson's disease (PD) at the Prince of Wales Hospital ($n = 63$; Hong Kong) and the Foothills Medical Centre ($n = 20$; Calgary, Canada). The groups did not differ in age or sex ($p > 0.05$), but Western patients had significantly more years of education ($M = 14.2$, $SD = 2.7$) than Asian patients ($M = 10.33$, $SD = 4.4$). Cognitive tests administered to both groups included: digit span, verbal fluency (animals), the Boston Naming Test, and verbal memory (California Verbal Learning Test or Chinese Auditory Verbal Learning Test). Testing was completed before and 12 months after deep brain stimulation surgery. Results showed cognitive performance was similar across time, but significant group differences were found on digit span forward (longer among patients from Hong Kong; $F(1, 75) = 44.155$, $p < 0.001$) and the Boston Naming Test (higher percent spontaneous correct among patients from Canada; $F(1, 62) = 7.218$, $p = 0.009$, $\eta^2 = 0.104$), after controlling for age, sex, and years of education. In conclusion, our findings provide preliminary support for the similarity of Chinese versions of tests originally developed for Western populations. Also, we caution that some aspects of testing may be susceptible to cultural bias and therefore warrant attention in clinical practice and refinement in future test development for Asian patients.

Keywords: Parkinson's Disease, Neuropsychological Tests, Psychometrics, Cross-cultural Comparison

Vivian Kwan, David Shum, Angela Haffenden, Keith Owen Yeates, Alice Kwok, Herman Lau, Wai Sang Poon, Danny Chan, X. L. Zhu, David Chan, Vincent Mok, Anne Chan, Karen Ma, Jonas Yeung, Claire Lau, Yarema Bezchlibnyk, Zelma Kiss & Venus Tang (2022) A retrospective comparison of cognitive performance in individuals with advanced Parkinson's Disease in Hong Kong and Canada, Applied Neuropsychology: Adult, 29:6, 1562-1570.

A Retrospective Comparison of Cognitive Performance in Individuals with Advanced Parkinson's Disease in Hong Kong and Canada

Parkinson's disease (PD) is the second most common neurodegenerative disease, with a prevalence of 306-371 per 100,000 in Asia and 1267-1535 per 100,000 in Western countries (i.e., Europe, North America, and Australia; Pringsheim, Jette, Frolkis, & Steeves, 2014). The disease is known to affect a variety of cognitive functions, including executive functions, memory, and visual spatial skills (Kehagia, Barker, & Robbins, 2010). Approximately 40-70% of patients with Parkinson's disease will develop dementia during the course of their illness (Aarsland, Andersen, Larsen, Lolk, & Kragh-Sørensen, 2003). Given the risk for cognitive decline in this population, the Movement Disorder Society Task Force published recommendations for neuropsychological assessment, including guidance on which key cognitive tests to administer (Bronstein et al., 2011; Dubois et al., 2007; Goldman et al., 2018; Litvan et al., 2011, 2012; Winblad et al., 2004). However, most of the cognitive tests were designed and validated for use in Western populations. As a consequence, their applicability may be limited for patients who have low levels of acculturation to Western society because of differences in language, education, and other cultural factors (Boone, Victor, Wen, Razani, & Pontón, 2007). This is problematic because neuropsychological evaluation is key to surgical preparation, access to surgical treatment, and monitoring in patients with PD. Therefore, if cognitive testing is culturally biased or not available to patients from different backgrounds, potential inequities may result in health and treatment access.

Most cross-cultural cognitive studies have focused on healthy participants rather than clinical samples (Boone et al., 2007; T.-B. Chen et al., 2014; Z. Y. Chen, Cowell, Varley, & Wang, 2009; Kempler, Teng, Dick, Taussig, & Davis, 1998; Manly et al., 1998). A number of studies have examined cognitive test performance in patients with PD outside of North America, such as from Prague and the Czech Republic (Bezdicek et al., 2016, 2017) and non-Western locations such as Hong Kong (Tang et al., 2015; Zhu et al., 2014), Taiwan (Yu et al., 2020, 2012), and China (Wang et al., 2012; Wang, Zhang, Li, Wen, & Xu, 2014). Data from healthy participants as well as patients with PD from diverse regions is important, but more research to compare test performance from one culture to another culture where the test was originally designed/developed/normed is needed to provide support for concurrent validity. Such research will be key to gain confidence in the use of different versions of tests and to guide interpretation for assessment of patients from different cultural backgrounds in line with evidence-based practice.

To the best of our knowledge, only one intra-cross-cultural comparison (same nation but different cultures) of cognitive testing in patients with PD has been published (Statucka & Cohn, 2019). The study examined patients with PD living in Canada who were born in Anglosphere countries (*n* = 248; USA, UK, Canada) compared to other international countries (*n* = 167; Asia, Europe, Caribbean, Africa). Testing was conducted through an interpreter for those with low English proficiency (0.4% of the Anglosphere group, 32.4% of the international group). Results showed that groups differed in cognitive performance, with the international group performing lower on visuo-perceptual and non-verbal executive tasks. This finding underscores the need to incorporate the influence of culture when assessing individuals who originate from other countries. Despite the interesting results, the international group of the study consisted of patients with PD from a wide range of backgrounds, making the influences of cultural factors and language difficult to parse. In the current study, we built on these findings by performing an extra-cross-cultural comparison (different ethnic and national groups living in different countries) of Hong Kong and Canada, using tools that have been translated for use in the language specific to the participants (i.e., Chinese-Cantonese for Hong Kong, and English for Canada), and administered directly (rather than through an interpreter) in the patient's native language.

Many previous studies have shown an effect of culture and ethnicity on cognitive test performance (Boone et al., 2007; Razani, Burciaga, Madore, & Wong, 2007; Rosselli & Ardila, 2003). English as the first language and acculturation to Western society are clearly related to higher test performance on common tests assessing attention (Razani et al., 2007), memory (Chen et al., 2009), processing speed (Razani et al., 2007), language, visual construction (Statucka & Cohn, 2019), and executive functioning skills (Boone et al., 2007). For example, cognitively healthy Chinese individuals, both Cantonese and Mandarin speaking, tend to perform better than English speaking individuals from the United Kingdom on digit span (Chen et al., 2009; Dehaene, 2011). The Boston Naming Test is another test recommended for patients with PD. It is a confrontation naming task that requires patients to name drawings of objects and animals (Chen et al., 2014). The test has been criticized for using items that are common to North America, but unknown in other countries (Rosselli & Ardila, 2003). A recent study found that healthy English-speaking participants were more likely to be successful than Chinese-speaking participants when naming pictures of certain items (i.e., tongs, cactus, trellis, tripod, dart, harp, igloo). In contrast, healthy Chinese-speaking participants were more likely to correctly name other items (e.g., compass, abacus, protractor) (Chen et al., 2014). This demonstrates the potential for cultural influences on confrontation naming. Overall, accumulating evidence indicates that while basic cognitive processes may be universal, a person's culture influences how those processes are reflected in their cognitive test performance (Statucka & Cohn, 2019). However, as mentioned above, few studies to date have compared cognitive performance in patient groups across different countries/cultures.

The purpose of the present retrospective study was to compare cognitive performance, based on commonly used neurocognitive tests recommended by the Movement Disorder Society (Bronstein et al., 2011; Dubois et al., 2007; Goldman et al., 2018; Litvan et al., 2011, 2012), in two groups of patients with PD that are relatively homogenous in clinical presentation. Our aim was to clarify whether Chinese versions of these well-established tools yield performances comparable to the original Western versions, both before and after deep brain stimulation. Neurocognitive evaluation is the standard of care for PD patients and results assist in selecting patients for the DBS procedure. Patients with advanced PD have a unique neurocognitive profile compared to healthy controls, and are expected to show minimal changes in cognitive performance following DBS, based on Western research. It is therefore useful to explore cross-cultural similarities and differences from a clinical point of view. In the presence of any group differences in performance, we sought to identify potential sources of the bias, to advise on score interpretation and future test development for Asian populations. As emphasized in a previous review (Chan, Shum, & Cheung, 2003), cross-cultural studies comparing local findings with Western findings are needed to demonstrate concurrent validity and to understand the applicability of adapted tests in the local population. Based on previous findings (Chen et al., 2014; Chen et al., 2009), we hypothesized that Asian and Western patients would differ on some of the administered measures, especially digit span and the Boston Naming Test.

The present study will have implications for better understanding cultural differences to guide test interpretation when assessing patients from Hong Kong in Canadian Hospitals. An estimated 1,769,195 Chinese citizens live in Canada, with 649, 260 immigrants from China and 208,940 immigrants from Hong Kong documented by the most recent Canadian census (Canada Census, 2016). The top three cities with the largest Chinese populations include Toronto, Vancouver, and Calgary (where the current study is based). According to statistics from Immigration, Refugees and Citizenship Canada (IRCC), Canada has seen a surge in interest from individuals in Hong Kong because of political events since July 2019 (Szeto, 2020).

Methods

Participants

We retrieved data retrospectively that was collected between year 2009 and 2018 at two medical centers. They were the Prince of Wales Hospital in Hong Kong, which is a teaching hospital of the Chinese University of Hong Kong, and the Foothills Medical Centre in Calgary, which is affiliated with the University of Calgary. We included all adult patients with a diagnosis of Parkinson's disease in the advanced stage without a diagnosis of dementia in Parkinson's disease (PDD), who received either bilateral subthalamic nucleus deep brain stimulation (STN-DBS) or globus pallidus deep brain stimulation (GPi-DBS). All Hong Kong participants (PD-HK) were Chinese Cantonese-speaking adults, and all Canadian participants (PD-CAD) spoke English as their primary language. All participants received baseline and 12-month post-operative neurocognitive evaluations using their first language based on testing protocols that were comparable but slightly different between the two centers due to language differences. Ethics approval was obtained from the Hong Kong New Territories East Cluster Clinical Research Ethics Committee (CREC Reference Number: 2019.188), and the Conjoint Health Research Ethics Board in Calgary (Study ID Number: REB19-1243).

Test Battery

In both Calgary and Hong Kong, standard clinical procedure includes neurocognitive evaluations before surgery (baseline) and approximately 12 months post-operatively by a clinical psychologist with specialized training in neuropsychology. All participants were assessed during their optimal functioning (i.e., drug "on" at baseline, drug and DBS "on" post-DBS). All tests were administered to Calgary participants in English by a native speaker of English and to Hong Kong participants in Cantonese by a native speaker of Cantonese.

Although the testing protocols differed between the two centers, a number of neurocognitive tests were the same or comparable, including tests of attention (Digit Span), verbal fluency (animals), confrontation naming (Boston Naming Test), and verbal memory (California Verbal Learning Test or the Chinese Auditory Verbal Learning Test). Of note, the Boston Naming Test and the verbal memory measures were different at the two sites, as we elaborate below. These four tests are part of the recommended battery for assessing cognitive functions in individuals with PD (Bronstein et al., 2011; Dubois et al., 2007; Goldman et al., 2015; Litvan et al., 2011, 2012; Saint-Cyr, 2003). They were not the only tests administered to the participants, but were the only ones administered to both groups and collected as part of a dataset for research. The other tests typically selected for clinical purposes are similar in Hong Kong and Calgary. A full test list and sources of normative data typically used at each site can be viewed in Appendix A and B.

Digit Span. For digit span, the tests were identical in Calgary (Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012) and Hong Kong (Lee & Wing, 2010). On digit forward, participants were read a sequence of numbers at the rate of one digit per second, and then asked to repeat them immediately after presentation in the same order. On digit backwards, patients were asked to recall the digits in reverse order. We recorded the longest span forward and the longest span backwards.

Verbal Fluency. Instructions for verbal fluency (animals) were the same for both sites, with participants asked to name as many animals as possible in 1 minute (Chan & Poon, 1999). The measure used for this test was the total number of animals correctly named within the time limit.

Confrontation Naming. The Boston Naming Test (Goodglass, 1983) contains 60 pictured items, with participants asked to name the items. This test is known to be culturally sensitive. A Chinese version was established and used widely in the Hong Kong population (Cheung, Cheung, & Chan, 2004); it contains 30 picture items from the original Boston Naming Test selected based on cultural relevance and the order of presentation is rearranged. Research found the modified 30-item version to have satisfactory internal consistency (Cronbach's alpha of 0.70 to 0.83) and discriminant validity (73.1% sensitivity and 75.3% specificity for differentiating individuals with brain injury from neurologically healthy individuals), suggesting it is appropriate for use in the Chinese population (Cheung, Cheung, & Chan, 2004). Because the two sites differed in the number of items administered, we examined the total percent of correct responses (spontaneous naming) in our analyses for comparison. For the 60-item original version of the Boston Naming Test, participants start at item 30 and continue through to item 60, unless certain discontinue or reverse-order criteria are met.

Verbal Memory. The two sites used different, but somewhat similar, verbal memory measures. In Hong Kong, the Chinese Auditory Verbal Learning Test (CAVLT) was administered. The CAVLT is a modified version of the Rey Auditory Verbal Learning Test, which includes a list of 15 unrelated words (Cheung, Cheung, & Chan, 2004; Lee, Yuen, & Chan, 2003). At the 12-month post-DBS assessment, the same version of the CAVLT was used, because no alternate form is available. In Calgary, patients completed the California Verbal Learning Test (CVLT), which includes 16 words that allow for semantic clustering into categories (Delis, Kramer, Kaplan, & Over, 1987). At the 12-month post-surgical assessment, an alternate form of the CVLT was used. Both the CVLT and CAVLT include five learning trials, immediate recall after interference, delayed recall, and delayed recognition, and record the number of intrusion errors and false alarm errors. Because of the difference in number of words for the CVLT and CAVLT, we examined the total percent correct recalled, as well as raw scores for number of errors (intrusions, false alarms). A study comparing the CVLT and CAVLT found no statistically significant differences in the performance of healthy young adults (Crossen & Wiens, 1994).

Surgery

All patients underwent stereotactic insertion of DBS electrodes under local anesthesia, targeting the subthalamic nucleus (STN) or, in 4 cases, the globus pallidus. Removal of the latter four patients from analyses did not lead to significantly different results; therefore, results for all participants are reported. Methods for bilateral insertion of DBS into the STN have been previously described (Hutchison et al., 1998; Zhu et al., 2014). DBS for the treatment of PD was granted approval by the US Food and Drug Administration (FDA) in 2002, and has been well recognized to improve motor functions among patients with advanced PD. The surgery itself consists of electrodes inserted in deep brain structures (Limousin et al., 1998). No major adverse events related to the surgery occurred.

Statistical Analyses

SPSS 26.0 software was used for all statistical analyses, and cognitive test data are presented as means and *SDs*. For the Boston Naming Test and verbal memory subtests (total words recalled, recall after interference, delayed recall), we analyzed percent correct responses because the measures have different numbers of items at each site. For digit span and verbal fluency (animals), we used raw scores. We examined group differences (i.e., PD-HK *vs.* PD-CAD) and changes in performance over time using repeated measures ANCOVA, co-varying for age, sex, and years of education. We report results with adjustment for multiple comparisons using the Benjamini-Hochberg False Discovery Rate (FDR) (Benjamini & Hochberg, 1995). Results presented without adjustment for multiple comparisons can be viewed under supplemental materials. Effect size was measured using partial eta squared, interpreted as follows: 0.01 = small, 0.06 = medium, 0.14 = large (Vacha-Haase & Thompson, 2004).

Results

Demographics

A total of 83 patients ($n = 20$ PD-CAD, $n = 63$ PD-HK) were included in the analysis. As Table I shows, the PD-HK group had significantly fewer total years of education on average ($M = 10.3$, $SD = 4.4$) than the PD-CAD group ($M = 14.2$, $SD = 2.7$; $p < 0.001$). The groups did not differ significantly in age or sex ($p > 0.05$). Groups also did not significantly differ in disease stage, as rated using the Hoehn and Yahr Scale (H&Y) (Zhao et al., 2010) and the Unified Parkinson's Disease Rating Scale (Part 2 and 3 for activities of daily living and motor exam; Goetz, 2003) while on medications and on stimulation for the post-surgical assessment. However, significant differences were noted between groups for H&Y pre-DBS off medications and post-DBS off medications and stimulation.

Site by time interactions

Repeated-measures ANCOVA revealed no significant group by time interactions on the neuropsychological tests. A nominally significant group by time interaction was found on the CVLT/CAVLT percent of total words remembered, $F(1, 74) = 5.88$, $p = 0.018$, but did not remain significant after adjusting for multiple comparisons. No main effects of change over time from pre to post-DBS were significant. Across time, pooled within-groups correlations were significant for digit span ($r = 0.59$, $p < 0.001$), digits backward ($r = 0.64$, $p < 0.001$), Boston Naming Test ($r = 0.69$, $p < 0.001$), Verbal Fluency ($r = 0.63$, $p < 0.001$), percent of words recalled ($r = 0.64$, $p < 0.001$), percent of words remembered after interference ($r = 0.73$, $p < 0.001$), percent of words remembered after delay ($r = 0.73$, $p < 0.001$), percent of words recognized ($r = 0.24$, $p = 0.037$), number of intrusion errors ($r = 0.25$, $p = 0.005$), number of false alarms ($r = 0.56$, $p < 0.001$), and number of recognition hits ($r = 0.24$, $p < 0.001$).

Main effect of site

Group means of cognitive test scores are presented in Table II. Significant main effects for site were found for digit span forward. $F(1, 75) = 44.155$, $p < 0.001$, $\eta^2 = 0.371$, whereby the PD-HK group demonstrated a significantly longer digit span forward. The PD-HK group also showed a lower percentage of spontaneous correct naming of objects, $F(1, 62) = 7.218$, $p = 0.009$, $\eta^2 = 0.104$, and fewer intrusion errors on verbal memory testing, $F(1, 74) = 10.959$, $p = 0.001$, $\eta^2 = 0.129$, compared to the PD-CAD group. These group differences remained significant after adjustment for multiple comparisons.

Covariates

Years of education was significantly positively related ($p < 0.05$) to digit span forward ($\eta^2 = 0.156$), digit span backwards ($\eta^2 = 0.177$), Boston Naming percent correct ($\eta^2 = 0.244$), percent of words remembered ($\eta^2 = 0.134$), percent of words remembered after a delay ($\eta^2 = 0.061$), and percent of words recognized ($\eta^2 = 0.085$), after adjusting for multiple comparisons. Age was significantly negatively associated with performance on digit span backwards ($\eta^2 = 0.167$), percent of total words remembered ($\eta^2 = 0.074$), percent recalled after interference ($\eta^2 = 0.096$), percent recalled after delay ($\eta^2 = 0.098$), and positively associated with interference errors ($\eta^2 = 0.075$) after adjusting for multiple comparisons. Sex was only significantly associated with the percent of total words recalled in the CVLT/CAVLT ($\eta^2 = 0.115$), with females performing higher than males after adjusting for multiple comparisons.

Discussion

Our aim was to determine whether Chinese versions of the same or similar neurocognitive tests in Hong Kong show findings comparable to the original Western versions in Calgary, Canada, in patients with PD before and after DBS. To the best of our knowledge, this is the first extra-cross-cultural comparison of patients with PD, which provides guidance for cognitive assessment of patients with PD in Canada but originally from Hong Kong. The results demonstrated some significant differences between test performance of the PD-CAD and PD-HK groups. Consistent with previous literature, the PD-HK group performed better on digit span forward (Hoosain, 2011) and worse on the Boston Naming Test (Boone et al., 2007). The latter finding should be interpreted with caution, because the two sites differed in the number of items in the test (30 versus 60). Otherwise, cognitive performance was not significantly different between sites in digit span backwards, verbal fluency (animals), or the CVLT/CAVLT measures of verbal memory. The two groups also demonstrated similar changes over time, from baseline to post-operation, given that repeated-measures ANCOVA revealed no significant group by time interactions on the neuropsychological tests.

Our results are consistent with previous research involving patients with PD before and after DBS (Cernera, Okun, & Gunduz, 2019). A meta-analysis of 28 cohort studies (612 patients in total) reported small effects of surgery in most cognitive domains, except for verbal fluency (Cohen's $d = 0.73$), which showed a moderate decline following STN-DBS (Witt et al., 2008). Our current findings are also consistent with previous results by our Hong Kong team members, who reported diminished performance on category fluency at 6 and 12 month re-evaluations in 27 patients recruited from the Prince of Wales Hospital (Tang et al., 2015). The absence of group differences on animal fluency is also consistent with previous cross-cultural comparative studies of cognitively healthy Chinese, Caucasian, Vietnamese, and Spanish participants (Kempler et al., 1998). However, our results differ from those reported by Statucka and Cohn (Statucka & Cohn, 2019), who found lower verbal fluency among Canadian PD patients born in international countries compared to those born in Anglosphere countries. The difference in results may be because they measured fluency using both animals' and boys' names. In the current study, the Chinese test of verbal fluency measured fluency for animals and types of transportation, and because both sites had only animal fluency in common, only this score was compared.

A longer digit span forwards, but not backwards, has previously been reported in Asian individuals. On average, digit span in Chinese languages is approximately two digits longer than those in English (Hoosain, 2011), consistent with our findings. The difference is likely related to articulation speed—numbers can be pronounced more quickly in Chinese (Mandarin & Cantonese) than English, which contains more syllables (Chen et al., 2009; Hoosain, 2011; Stigler, Lee, & Stevenson, 1986). While rate of administration is controlled at a rate of 1 digit per second, patients are free to recall at any speed. Interestingly, faster articulation speed may not be advantageous for digit span backwards, where a longer temporal duration of spoken digits may also provide some cognitive advantages (Chen et al., 2009). This notion is consistent with a number of studies that have shown less robust effects of language on backward than forward digit span (C. Chen & Stevenson, 1988; Hoosain, 1984). Also consistent are studies demonstrating a link between longer syllables and better recall performance on tasks requiring mental manipulations (Tolan & Tehan, 2005), similar to reciting numbers backwards.

Only a few studies have examined potential cultural effects on the Boston Naming Test in Chinese individuals (Chen et al., 2014; Cheung et al., 2004). A recent study at Taipei Veterans General Hospital examined 264 cognitively healthy Chinese-speaking older individuals (> 60 years old) and compared their data to that from other studies in English-speaking countries (e.g. USA, Canada, Australia) (Chen et al., 2014). Similar to our findings, years of education, but not sex or age, were positively associated with test performance. Some items were universally likely to be answered correctly across cultures ('scissors', 'pencil'), while others were more prone to be answered correctly by Chinese-speaking ('abacus') versus English-speaking participants ('dart', 'harp', 'igloo', 'accordion', 'mushroom'). The unique properties of Chinese as a logographic language introduce additional complexities. While correct answers in English on the Boston Naming Test are single words, some require two-character words in Chinese. For example, the correct answer for one of the test items is "mushroom" in English (one word) and "*moh gu"* in Chinese (two-character words). From our experience, Cantonese-speaking patients sometimes may say "*dong gu"* instead of *"moh gu,"* which is close but not correct. Thus, the Boston Naming Test may be more challenging in the Chinese version because multi-character words may create more room for error. Nonetheless, past research has found that the Chinese version of the Boston Naming Test was a sensitive (73.1%) and specific (75.3%) measure for differentiating individuals with brain injury from neurologically healthy individuals (Cheung et al., 2004).

Direct comparisons of the CVLT and CAVLT are difficult because of differences between the test and forms used. As mentioned earlier, the CVLT consists of 16 semantically-related words, and different versions were used for the baseline and 12-month follow-up assessment for Canadian patients to minimize practice effects. The CAVLT consists of 15 non-related words (Lee et al., 2003), and the same version was used at baseline and 12 month follow-up assessments for Hong Kong patients because no alternate form is available. Previous studies that compared the English versions of CVLT and AVLT found no significant differences in raw scores in healthy individuals (n = 60)(Crossen & Wiens, 1994) and individuals with brain injury (n = 40) (Stallings, Boake, & Sherer, 1995). Scores were found to be almost equivalent, with approximately one half to one full word mean difference between the AVLT and CVLT, closely corresponding to the difference in the actual number of items on the two tests. This is consistent with our results suggesting no significant group differences in the percent of total words recalled, percent recalled after interference, and percent recalled after delay. Interestingly, we found that Hong Kong patients made significantly fewer intrusion errors on the CAVLT than did Canadian patients on the CVLT. This finding requires further investigation, but may be because the words are not semantically related on the CAVLT. In the absence of the categories for clustering, patients may be less likely to make intrusion errors involving other words that also belong in a category.

Not surprisingly, age and years of education emerged as significant covariates for performance on cognitive tasks. Our samples differed significantly, such that the PD-HK group had significantly fewer total years of education on average than the PD-CAD group. This is not surprising and may reflect historical differences between the two sites. In Hong Kong, the first law requiring compulsory education for children between ages of 6 and 11 was established relatively recently, in 1971. Education in both Hong Kong and Canada is largely modelled based on the United Kingdom. Therefore, the two education systems are comparable. Most school curriculums are designed with a balance, including but not limited to language, mathematics, science, arts, and physical education. To date, qualifications of public examinations in Hong Kong and Canada are mutually recognized (UNESCO Institute of Statistics, 2021).

In summary, we compared the cognitive performance of PD patients using English versions of several well-established neurocognitive tests in Canada versus Chinese-Cantonese versions in Hong Kong. The current study is one of the very few to examine cross-cultural differences in cognitive performance across groups of patients rather than healthy participants. Our findings contribute preliminary evidence for the similarity of Chinese versions of tests originally developed for Western populations (verbal fluency, California Verbal Learning Test/Chinese Auditory Verbal Learning Test), but also some areas of difference (digit span forward, Boston Naming Test), consistent with past research in diverse healthy samples (Boone et al., 2007; Hoosain, 2011). More research is needed to cross-validate, adapt, and develop tools for use in clinical settings for patients with diverse backgrounds. In the following sections, we discuss study limitations and suggestions for future research.

Limitations

The current findings are limited by a lack of data on ethnicity, country of birth, level of acculturation (Tan et al., 2020), and socio-developmental context, e.g. Historical Index of Human Development (De la Escosura, 2015), which are known to be related to cognitive performance (Statucka & Cohn, 2019). The tests we compared were not identical, especially for the verbal learning measures (CVLT, CAVLT) and the Boston Naming Test (60 items versus 30 items). We addressed this concern by examining percent correct rather than raw scores. The current study was also limited by relatively small and unequal sample sizes, which reflected the retrospective use of an existing database, rather than a priori power analyses. The Calgary sample had significantly higher years of education. Regarding targets of DBS for PD, it is reported that those who underwent GPi-DBS have fewer cognitive effects (Odekerken et al., 2013; Okun et al., 2009) and depression (Combs et al., 2015) than STN-DBS. We have included a small number of Canadian participants with GPi-DBS (n = 4), and our results did not change after removing those cases from the analyses. In addition, we did not have information available for Levodopa equivalents or any other medications. Our comparison showed that the Hong Kong participants had higher H&Y scores for pre-DBS off medications and post-DBS off medications and stimulation. This suggests that surgeries may be done for less advanced cases of PD in Calgary, Canada. The patients from Hong Kong in the current study may or may not be similar or representative of Chinese patients from other regions such as Macau and Taiwan or from other ethnic groups in China (e.g. Han Chinese, Zhuang, Hui, Manchu). However, Hong Kong has very close ties to China, and the majority of Hong Kong citizens descendants are of families from various parts of Southern China (Guangzhou, Siyi, Chaoshan, Fujian, and Shanghai) according to the most recent Hong Kong census in 2016 (Hong Kong Census and Statistics Department, 2016). In summary, the present study has several important limitations: lack of comprehensive neuropsychological test data, tests that were not standardized or identically administered across the two sites, absence of a healthy control group, and a small sample size that could potentially lead to underpowered analyses. Future research should be designed a priori with these key factors (comprehensive data, standardized tests, and inclusion of a control group) incorporated for the strongest empirical study possible.

Conclusion

In general, cognitive test performance was largely similar between the groups in Hong Kong and Canada, with both showing a profile of performance consistent with the large body of research on neurocognition in PD. That is, neurocognitive functions remained largely unaffected following DBS surgery. The current study provides evidence for the similarity of Chinese versions of tests recommended for use in individuals with PD compared to the original Western versions, before and after DBS. However, we caution that Digit Span Forward and the Boston Naming Test may be susceptible to cultural influences among Asian patients with PD. To be more specific, applying appropriate norms are important for digit span, given evidence that Chinese patients tend to have a longer forward span when tested in Cantonese Chinese. For the Boston Naming Test, we found that certain items have a higher risk of bias, favoring those with an English-speaking Western background. However, this warrants further study because non-identical versions of the Boston Naming Test were administered at each site. From a broader perspective, a better appreciation and understanding of cultural variations in cognitive testing should be emphasized in research and clinical practice. Further cross-cultural research will be essential to help guide services for patients from diverse backgrounds.

Acknowledgments

The authors thank the participants of this study who agreed for their information to be used in research. This study was funded by Mitacs Globalink as well as the Hong Kong Polytechnic University Incoming Visiting Ph.D. Student Attachment Programme awarded to V.K. The authors have no conflicts of interest to disclose.

References

Aarsland, D., Andersen, K., Larsen, J. P., Lolk, A., & Kragh-Sørensen, P. (2003). Prevalence and characteristics of dementia in Parkinson disease: An 8-year prospective study. *Archives of Neurology*. https://doi.org/10.1001/archneur.60.3.387

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. https://doi.org/10.2307/2346101

Bezdicek, O., Moták, L., Schretlen, D. J., Preiss, M., Axelrod, B. N., Nikolai, T., … Růžička, E. (2016). Sociocultural and language differences on the Trail Making Test. *Archives of Assessment Psychology*.

Bezdicek, O., Sulc, Z., Nikolai, T., Stepankova, H., Kopecek, M., Jech, R., & Růžička, E. (2017). A parsimonious scoring and normative calculator for the Parkinson's disease mild cognitive impairment battery. *Clinical Neuropsychologist*. https://doi.org/10.1080/13854046.2017.1293161

Boone, K. B., Victor, T. L., Wen, J., Razani, J., & Pontón, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology*, *22*(3), 355–365. https://doi.org/10.1016/j.acn.2007.01.010

Bronstein, J. M., Tagliati, M., Alterman, R. L., Lozano, A. M., Volkmann, J., Stefani, A., … DeLong, M. R. (2011). Deep brain stimulation for Parkinson disease an expert consensus and review of key issues. *Archives of Neurology*. https://doi.org/10.1001/archneurol.2010.260

Canada Census (2016). *Census Profile, 2016 Census. https://www12.statcan.gc.ca/census-recensement/2016*

Cernera, S., Okun, M. S., & Gunduz, A. (2019). A review of cognitive outcomes across movement disorder patients undergoing deep brain stimulation. *Frontiers in Neurology*. https://doi.org/10.3389/fneur.2019.00419

CHAN, A. S., & POON, M. W. (1999). Performance of 7- to 95-year-old individuals in a Chinese version of the category fluency test. *Journal of the International Neuropsychological Society*. https://doi.org/10.1017/s135561779956606x

Chan, A. S., Shum, D., & Cheung, R. W. Y. (2003). Recent Development of Cognitive and Neuropsychological Assessment in Asian Countries. *Psychological Assessment*, *15*(3), 257–267. https://doi.org/10.1037/1040-3590.15.3.257

Chen, C., & Stevenson, H. W. (1988). Cross-linguistic differences in digit span of preschool children. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/0022-0965(88)90027-6

Chen, T.-B., Lin, C.-Y., Lin, K.-N., Yeh, Y.-C., Chen, W.-T., Wang, K.-S., & Wang, P.-N.

(2014). Culture Qualitatively but Not Quantitatively Influences Performance in the Boston Naming Test in a Chinese-Speaking Population. *Dementia and Geriatric Cognitive Disorders Extra*. https://doi.org/10.1159/000360695

Chen, Z. Y., Cowell, P. E., Varley, R., & Wang, Y. C. (2009). A cross-language study of verbal and visuospatial working memory span. *Journal of Clinical and Experimental Neuropsychology*. https://doi.org/10.1080/13803390802195195

Cheung, R. W., Cheung, M. C., & Chan, A. S. (2004). Confrontation naming in Chinese patients with left, right or bilateral brain damage. *Journal of the International Neuropsychological Society*. https://doi.org/10.1017/S1355617704101069

Combs, H. L., Folley, B. S., Berry, D. T. R., Segerstrom, S. C., Han, D. Y., Anderson-Mooney, A. J., … van Horne, C. (2015). Cognition and Depression Following Deep Brain Stimulation of the Subthalamic Nucleus and Globus Pallidus Pars Internus in Parkinson's Disease: A Meta-Analysis. *Neuropsychology Review*. https://doi.org/10.1007/s11065-015-9302-0

Crossen, J. R., & Wiens, A. N. (1994). Comparison of the Auditory-Verbal Learning Test (AVLT) and California Verbal Learning Test (CVLT) in a Sample ofdm Normal Subjects. *Journal of Clinical and Experimental Neuropsychology*. https://doi.org/10.1080/01688639408402630

De la Escosura, L. P. (2015). World Human Development: 1870-2007. *Review of Income and Wealth*. https://doi.org/10.1111/roiw.12104

Dehaene, S. (2011). The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition. In *The number sense How the mind creates mathematics rev and updated ed*.

Delis, D. C., Kramer, J. H., Kaplan, E., & Over, B. A. (1987). California verbal learning test research edition manuaL. In *The Psychological Corporation*.

Dubois, B., Burn, D., Goetz, C., Aarsland, D., Brown, R. G., Broe, G. A., … Emre, M. (2007). Diagnostic procedures for Parkinson's disease dementia: Recommendations from the Movement Disorder Society Task Force. *Movement Disorders*, *22*(16), 2314–2324. https://doi.org/10.1002/mds.21844

E, Goodglass H, W. S. (1983). The Boston Naming Test. *Philadelphia: Lea and Febiger.*

Goetz, C. C. (2003). The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Movement Disorders*. https://doi.org/10.1002/mds.10473

Goldman, J. G., Holden, S., Ouyang, B., Bernard, B., Goetz, C. G., & Stebbins, G. T. (2015). Diagnosing PD-MCI by MDS task force criteria: How many and which neuropsychological tests? *Movement Disorders*. https://doi.org/10.1002/mds.26084

Goldman, J. G., Vernaleo, B. A., Camicioli, R., Dahodwala, N., Dobkin, R. D., Ellis, T., … Simmonds, D. (2018). Cognitive impairment in Parkinson's disease: a report from a multidisciplinary symposium on unmet needs and future directions to maintain cognitive health. *Npj Parkinson's Disease*. https://doi.org/10.1038/s41531-018-0055-3

Hoosain, R. (2011). Correlation between Pronunciation Speed and Digit Span Size. *Perceptual and Motor Skills*. https://doi.org/10.2466/pms.1982.55.3f.1128

HOOSAIN, R. (1984). LATERALIZATION OF BILINGUAL DIGIT SPAN FUNCTIONS. *Perceptual and Motor Skills*. https://doi.org/10.2466/pms.58.1.21-22

Hutchison, W. D., Allan, R. J., Opitz, H., Levy, R., Dostrovsky, J. O., Lang, A. E., & Lozano, A. M. (1998). Neurophysiological identification of the subthalamic nucleus in surgery for Parkinson's disease. *Annals of Neurology*. https://doi.org/10.1002/ana.410440407

Kehagia, A. A., Barker, R. A., & Robbins, T. W. (2010). Neuropsychological and clinical heterogeneity of cognitive impairment and dementia in patients with Parkinson's disease. *The Lancet Neurology*. https://doi.org/10.1016/S1474-4422(10)70212-X

Kempler, D., Teng, E. L., Dick, M., Taussig, I. M., & Davis, D. S. (1998). The effects of age, education, and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society*. https://doi.org/10.1017/S1355617798466013

Lee, T. M. C., Yuen, K. S. L., & Chan, C. C. H. (2003). Normative Data for Neuropsychological Measures of Fluency, Attention, and Memory Measures for Hong Kong Chinese. *Journal of Clinical and Experimental Neuropsychology*. https://doi.org/10.1076/jcen.24.5.615.1001

Limousin, P., Krack, P., Pollak, P., Benazzouz, A., Ardouin, C., Hoffmann, D., & Benabid, A. L. (1998). Electrical stimulation of the subthalamic nucleus in advanced Parkinsonian's disease. *New England Journal of Medicine*. https://doi.org/10.1056/NEJM199810153391603

Litvan, I., Aarsland, D., Adler, C. H., Goldman, J. G., Kulisevsky, J., Mollenhauer, B., … Weintraub, D. (2011). MDS task force on mild cognitive impairment in Parkinson's disease: Critical review of PD-MCI. *Movement Disorders*. https://doi.org/10.1002/mds.23823

Litvan, I., Goldman, J. G., Tröster, A. I., Schmand, B. A., Weintraub, D., Petersen, R. C., … Emre, M. (2012). Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Movement Disorders*. https://doi.org/10.1002/mds.24893

Manly, J. J., Jacobs, D. M., Sano, M., Bell, K., Merchant, C. A., Small, S. A., & Stern, Y. (1998). Cognitive test performance among nondemented elderly African Americans and whites. *Neurology*. https://doi.org/10.1212/WNL.50.5.1238

Odekerken, V. J. J., van Laar, T., Staal, M. J., Mosch, A., Hoffmann, C. F. E., Nijssen, P. C. G., … de Bie, R. M. A. (2013). Subthalamic nucleus versus globus pallidus bilateral deep brain stimulation for advanced Parkinson's disease (NSTAPS study): A randomised controlled trial. *The Lancet Neurology*. https://doi.org/10.1016/S1474-4422(12)70264-8

Okun, M. S., Fernandez, H. H., Wu, S. S., Kirsch-Darrow, L., Bowers, D., Bova, F., … Foote, K. D. (2009). Cognition and mood in Parkinson's disease in subthalamic nucleus versus globus pallidus interna deep brain stimulation: The COMPARE trial. *Annals of Neurology*. https://doi.org/10.1002/ana.21596

Pringsheim, T., Jette, N., Frolkis, A., & Steeves, T. D. (2014). The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Movement Disorders : Official Journal of the Movement Disorder Society*. https://doi.org/10.1002/mds.25945 [doi]

Razani, J., Burciaga, J., Madore, M., & Wong, J. (2007). Effects of acculturation on tests of attention and information processing in an ethnically diverse group. *Archives of Clinical Neuropsychology*. https://doi.org/10.1016/j.acn.2007.01.008

Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*. https://doi.org/10.1016/S0278-2626(03)00170-2

Saint-Cyr, J. A. (2003). Neuropsychology for Movement Disorders Neurosurgery. *Canadian Journal of Neurological Sciences / Journal Canadien Des Sciences Neurologiques*. https://doi.org/10.1017/s0317167100003280

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable Digit Span: A Systematic Review and Cross-Validation Study. *Assessment*. https://doi.org/10.1177/1073191111428764

Stallings, G., Boake, C., & Sherer, M. (1995). Comparison of the California Verbal Learning Test and the Rey Auditory Verbal Learning Test in Head-Injured Patients. *Journal of Clinical and Experimental Neuropsychology*. https://doi.org/10.1080/01688639508405160

Statucka, M., & Cohn, M. (2019). Origins Matter: Culture Impacts Cognitive Testing in Parkinson's Disease. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2019.00269

Stigler, J. W., Lee, S. Y., & Stevenson, H. W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition*. https://doi.org/10.1016/0010-0277(86)90051-X

Szeto, W. (2020, December, 05). Canada to ease immigration rules for Hong Kongers. The move couldn't come fast enough for these B.C. students. *CBC News. https://www.cbc.ca/news/canada/british-columbia/hong-kong-china-students-immigration-canada-1.5821554*

Tan, Y. W., Burgess, G. H., Green, R. J., Wen, Y., Burgess, G. H., & Green, R. J. (2020). The effects of acculturation on neuropsychological test performance : A systematic literature review The effects of acculturation on neuropsychological test. *The Clinical Neuropsychologist*, *0*(0), 1–31. https://doi.org/10.1080/13854046.2020.1714740

Tang, V., Zhu, C. X. L., Chan, D., Lau, C., Chan, A., Mok, V., … Poon, W. S. (2015).

Evidence of improved immediate verbal memory and diminished category fluency following STN-DBS in Chinese-Cantonese patients with idiopathic Parkinson's disease. *Neurological Sciences*, *36*(8), 1371–1377. https://doi.org/10.1007/s10072-015-2117-1

Tolan, G. A., & Tehan, G. (2005). Is spoken duration a sufficient explanation of the word length effect? *Memory*. https://doi.org/10.1080/09658210344000305

UNESCO Institute of Statistics. (2021). *China, Hong Kong Special Administrative Region.* http://uis.unesco.org/en/country/hk

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*. https://doi.org/10.1037/0022-0167.51.4.473

Wang, L., Nie, K., Zhang, Y., Wang, L., Zhao, J., Huang, Z., … Li, S. (2012). A pilot study of psychometric properties of the Beijing version of Montreal Cognitive Assessment in patients with idiopathic Parkinson's disease in China. *Journal of Clinical Neuroscience*. https://doi.org/10.1016/j.jocn.2011.11.039

Wang, Q., Zhang, Z., Li, L., Wen, H., & Xu, Q. (2014). Assessment of cognitive impairment in patients with Parkinson's disease: Prevalence and risk factors. *Clinical Interventions in Aging*. https://doi.org/10.2147/CIA.S47367

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., … Petersen, R. C. (2004). Mild cognitive impairment - Beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*. https://doi.org/10.1111/j.1365-2796.2004.01380.x

Witt, K., Daniels, C., Reiff, J., Krack, P., Volkmann, J., Pinsker, M. O., … Deuschl, G. (2008). Neuropsychological and psychiatric changes after deep brain stimulation for Parkinson's disease: a randomised, multicentre study. *The Lancet Neurology*. https://doi.org/10.1016/S1474-4422(08)70114-5

Yu, R. L., Lee, W. J., Li, J. Y., Chang, Y. Y., Chen, C. C., Lin, J. J., … Fuh, J. L. (2020). Evaluating Mild Cognitive Dysfunction in Patients with Parkinson's Disease in Clinical Practice in Taiwan. *Scientific Reports*. https://doi.org/10.1038/s41598-020-58042-2

Yu, R. L., Wu, R. M., Tai, C. H., Lin, C. H., Cheng, T. W., & Hua, M. S. (2012). Neuropsychological profile in patients with early stage of Parkinson's disease in Taiwan. *Parkinsonism and Related Disorders*. https://doi.org/10.1016/j.parkreldis.2012.06.002

Zhao, Y. J., Wee, H. L., Chan, Y. H., Seah, S. H., Au, W. L., Lau, P. N., … Tan, L. C. S. (2010). Progression of Parkinson's disease as evaluated by Hoehn and Yahr stage transition times. *Movement Disorders*. https://doi.org/10.1002/mds.22875

Zhu, X. L., Chan, D. T. M., Lau, C. K. Y., Poon, W. S., Mok, V. C. T., Chan, A. Y. Y., … Yeung, C. (2014). Cost-effectiveness of subthalmic nucleus deep brain stimulation for the treatment of advanced Parkinson disease in Hong Kong: A prospective study. *World Neurosurgery*. https://doi.org/10.1016/j.wneu.2014.08.051

Appendix A

Below is the full cognitive test list typically used at each site for assessment of patients with Parkinson's Disease. Tests are subject to change based on the individual patient.

Calgary, Canada, Foothills Medical Centre (typically 3-4 hours): Wechsler Adult Intelligence Scale- IV (Matrix Reasoning, Similarities, Digit Span, Letter Number Sequencing) Wechsler Memory Scale-III (Spatial Span and Mental Control), Wechsler Memory Scale -IV Symbol Span, Boston Naming Test, Benton Judgment of Line Orientation, Rey Complex Figure Test, Brief Test of Attention, California Verbal Learning Test-II, Memory for Intentions Test, Verbal Fluency, Wisconsin Card Sorting Test, and the Delis-Kaplan Executive Function System Color Word Interference Test.

Hong Kong, Prince of Wales Hospital (typically 1-2 hours): Hong Kong Montreal Cognitive Assessment, Chinese Auditory Verbal Learning Test, Benton Visual Retention Test Fifth Edition, Stroop Test (Chinese Translated Victoria Version), Digit Span, Boston Naming Test, Hooper's Visual Organization Test, Benton Judgment of Line Orientation, and Category Fluency (animal and transportation), Back Depression Inventory II, Beck Anxiety Inventory.

Appendix B

Test	Normative Data
Calgary	
Digit Span	Wechsler norms
Verbal Fluency	Heaton norms
Boston Naming Test	Heaton and Tombaugh norms
Verbal learning Test (California Verbal Learning Test)	California Verbal Learning Test norms
Hong Kong	
Digit Span	Lee and Wang (2010) norms
Verbal Fluency	Chan and Poon (1998) norms
Boston Naming Test	Cheung, Cheung and Chan (2004) norms
Verbal Learning Test (Chinese version of Rey Auditory Verbal Learning Test)	Lee and Wang (2010) norms

Table I. Participant Demographics

Variable	Calgary *M (SD)*	Hong Kong *M (SD)*	Statistic *t* or x^2	*p*-value
N	20	63		
Age in years	58.5 (6.9) (range 45-70)	57.7 (6.7) (range 38-72)	-0.482	0.631
Years of education	14.2 (2.7) (range 11-22)	10.3 (4.4) (range 2– 8)	-4.651	<0.001*
% Male	60	63	0.079	0.779
H & Y stage (pre-DBS, on medications)	2.395 (0.39)	2.429 (0.545)	0.251	0.148
UPDRS score (part 2 and 3, pre-DBS, on medications)	24.55 (13.17)	23.29 (12.48)	-0.39	0.698
H & Y stage (pre-DBS, off medications)	2.68 (0.47)	3.37 (0.85)	4.68	< 0.001*
UPDRS score (part 2 and 3, pre-DBS, off medications)	54.50 (16.24)	60.08 (19.09)	1.18	0.24
H & Y post-DBS (on meds/stim)	2.658 (0.502)	2.54 (0.51)	-0.88	0.377
H & Y post-DBS (off meds/stim)	2.91 (0.73)	3.54 (0.92)	2.96	0.006*
UPDRS post-DBS (on meds/stim)	23.79 (9.52)	18.37 (12.67)	-1.723	0.089
UPDRS score (part 2 and 3, post-DBS, off meds/stim)	58.26 (25.56)	65.51 (17.31)	1.42	0.159

Table I note: Abbreviations of *H & Y* for the Hoehn and Yahr scale; *UPDRS* for the Unified Parkinson's Disease Rating Scale; *DBS* for deep brain stimulation.

Table II. Cognitive Test Performance

Cognitive Measure	PD-CAD (*n*=20)		PD-HK (*n*=63)	
	Pre-DBS M (SD)	Post-DBS M (SD)	Pre-DBS M (SD)	Post-DBS M (SD)
Digit span forward*	6.70 (1.17)	6.65 (1.31)	8.25 (1.28)	8.22 (1.37)
Digit span backward	4.7 (1.34)	4.50 (1.05)	4.7 (1.48)	4.77 (1.47)
Category fluency (animals)	20.25 (5.61)	17.55 (6.52)	17.95 (3.43)	15.75 (4.85)
Boston Naming Test (% correct)*	93.33 (4.76)	93.25 (4.89)	85.10 (9.60)	82.08 (11.86)
CVLT/CAVLT (1-5 total) %	60.20 (11.00)	56.84 (13.28)	53.89 (13.13)	60.04 (13.72)
CVLT/CAVLT Intrusion Errors*	6.42 (7.23)	5.68 (6.83)	3.26 (3.35)	2.40 (3.43)
CVLT/CAVLT recall after interference %	60.20 (20.33)	59.21 (20.02)	56.34 (20.93)	57.67 (21.92)
CVLT/CAVLT delayed recall %	64.47 (18.76)	59.21 (22.86)	55.85 (21.51)	56.00 (24.74)
CVLT/CAVLT recognition %	92.43 (7.39)	90.13 (7.60)	91.69 (9.86)	89.33 (13.12)
CVLT/CAVLT False Alarm	2.68 (3.38)	3.68 (4.37)	2.39 (3.05)	1.65 (2.40)

Table II note: * indicates significant main effect of site (Canada *vs.* Hong Kong) after adjusting for multiple comparisons using the False Discovery Rate. No significant site by time (pre *vs.* post-DBS) interactions were found. Abbreviations: *CVLT* for the California Verbal Learning Test; *CAVLT* for the Chinese Auditory Verbal Learning Test; *DBS* for deep brain stimulation.

Supplemental materials

The following results were not adjusted for multiple comparisons, in light of the possibility that differences or effects may be undetected due to underpowered analyses. Comparisons of results with and without adjustment shows largely similar findings, except for three additional significant associations (bolded below) between: (1) site and false alarm errors on the CVLT/CAVLT, (2) sex and percent of words recognized on the CVLT/CAVLT, and (3) education and category fluency.

Site by time interactions

Repeated-measures ANCOVA revealed no significant group by time interactions on the neuropsychological tests. A significant group by time interaction was found on the CVLT/CAVLT percent of total words remembered, $F(1, 74) = 5.88$, $p = 0.018$. No main effects of change over time from pre to post-DBS were significant. Across time, pooled within-groups correlations were significant for digit span ($r = 0.59$, $p < 0.001$), digits backward ($r = 0.64$, $p < 0.001$), Boston Naming Test ($r = 0.69$, $p < 0.001$), Verbal Fluency ($r = 0.63$, $p < 0.001$), percent of words recalled ($r = 0.64$, $p < 0.001$), percent of words remembered after interference ($r = 0.73$, $p < 0.001$), percent of words remembered after delay ($r = 0.73$, $p < 0.001$), percent of words recognized ($r = 0.24$, $p = 0.037$), number of intrusion errors ($r = 0.25$, $p = 0.005$), number of false alarms ($r = 0.56$, $p < 0.001$), and number of recognition hits ($r = 0.24$, $p < 0.001$).

Main effect of site

Group means of cognitive test scores are presented in Table II. Significant main effects for site were found for digit span forward ($F(1, 75) = 44.155$, $p < 0.001$, $\eta^2= 0.371$) whereby the PD-HK group demonstrated a significantly longer digit span forward. The PD-HK group also showed a lower percentage of spontaneous correct naming of objects ($F(1, 62) = 7.218$, $p = 0.009$, $\eta^2 = 0.104$), fewer intrusion errors ($F(1, 74) = 10.959$, $p = 0.001$, $\eta^2= 0.129$)**, and fewer false alarm errors ($F(1, 73) = 4.00$, $p = 0.049$, $\eta^2= 0.052$)**, compared to the PD-CAD group.

Covariates

Years of education was significantly positively related ($p < 0.05$) to digit span forward ($\eta^2 = 0.156$), digit span backwards ($\eta^2 = 0.177$), Boston Naming percent correct ($\eta^2 = 0.244$), percent of words remembered ($\eta^2 = 0.134$), percent of words remembered after a delay ($\eta^2 = 0.061$), and percent of words recognized ($\eta^2 = 0.085$)**, and words generated for category fluency ($\eta^2 = 0.053$).** Age was significantly negatively associated with performance on digit span backwards ($\eta^2 = 0.167$), percent of total words remembered ($\eta^2 = 0.074$), percent recalled after interference ($\eta^2 = 0.096$), percent recalled after delay ($\eta^2 = 0.098$), and positively associated with interference errors ($\eta^2 = 0.075$) after adjusting for multiple comparisons. Sex was significantly associated with the percent of total words recalled in the CVLT/CAVLT ($\eta^2 = 0.115$) **and percent of words recognized ($\eta^2 = 0.077$), with females performing higher than males on both tasks.**