DOI:
10.1080/13854046.2015.1034182

The Equivalence and Difference between the English and Chinese language versions of the Repeatable Battery for the Assessment of Neuropsychological Status

Rachel Phillips[1*], Yin Bun Cheung [2,3], Simon Lowes Collinson[4, 5], May-Li Lim[4, 6], Audrey Ling[4, 6], Lei Feng[4, 6], Tze-Pin Ng[4, 6].

1. Biostatistics Department, Singapore Clinical Research Institute, 31 Biopolis Way Nanos #02-01 Singapore, 138669.
2. Centre for Quantitative Medicine, Academia, 20 College Road, Duke-National University of Singapore Graduate Medical School, Singapore.
3. Department of International Health, University of Tampere, Finland 33014.
4. Gerontological Research Programme, Faculty of Medicine, National University of Singapore, Singapore.
5. Department of Psychology, National University of Singapore, Singapore 117570.
6. Department of Psychological Medicine, National University of Singapore, Singapore 119228.

*Corresponding author

Contact details of authors

RP: Telephone: (65) 6508 6754
Email: rachel.phillips@scri.edu.sg

YBC: Telephone: (65) 6576 7379
Email: yinbun.cheung@duke-nus.edu.sg

SLC: Telephone: (65) 6516 8117
Email: psycsl@nus.edu.sg

MLL: Telephone: (65) 6772 3494
Email: may_li_lim@nuhs.edu.sg

AL: Telephone: (65) 6772 3492
Email: pcmlhla@nus.edu.sg

LF: Telephone: (65) 6772 3491
Email: lei_feng@nuhs.edu.sg

TPN: Email: (65) 6772 3478
Telephone: tze_pin_ng@nuhs.edu.sg

Abstract word count: 249
Manuscript word count: 5100

This work has been support by the Biomedical Research Council (NTP, 03/1/21/17/214, 08/1/21/19/567). The authors report there are no conflicts of interest.

Abstract

Objective: Chinese is the most commonly spoken language in the world. The availability of Chinese translations of assessment scales is useful for research in multi-ethnic and multinational studies. This study aimed to establish whether each of the Chinese translations (Mandarin, Hokkien, Teochew and Cantonese) of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) achieved measurement equivalence to the English version.

Method: Participants included 1,856 ethnic Chinese, older adults. The RBANS was administered in the language/dialect according to the participants' preference by interviewers who were fluent in that language/dialect. Multiple regression analysis was used to adjust for demographic and clinical differences between participants who spoke different languages/dialects. Equivalence (practical equivalence) was declared if the 90% confidence interval for the adjusted mean difference fell entirely within the pre-specified equivalence margin, ±0.2 (±0.4) standard deviations (SD).

Results: The delayed memory index was at least practically equivalent across languages. The Mandarin, Hokkien and Teochew versions of the immediate memory, language and total scale score were practically equivalent to the English version; the Cantonese version showed small differences from the English version. Equivalence was not established for the Hokkien and Teochew versions of the visuospatial/constructional index. The attention index was different across languages.

Conclusions: Data from the English and Chinese versions for the total scale score, language, delayed and immediate memory indexes may be pooled for analysis. However, analysis of the attention and visuospatial/constructional indexes from English and Chinese versions should include a covariate that represents the version in the statistical adjustment.

Keywords: Chinese**,** language, equivalence, RBANS, cognitive decline.

Introduction

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) was designed to identify and characterise abnormal cognitive decline in older adults. It was also intended for use in younger adults as a neuropsychological screening tool (Randolph, 1998). The RBANS was later shown to be sensitive to impairments in people with mild cognitive impairment (MCI) (Badenes, Casas, Cejudo & Aguilar, 2008; Mahncke et al., 2006; Kotani et al., 2006) and in an array of other patient groups (http://www.rbans.com/clinicalvalidity.html). The battery is comprised of 12 subtests that assess the domains of immediate and delayed memory, language, attention and visuospatial/construction (Randolph, 1998). The conciseness of the RBANS and its ease of use make it an attractive tool for the assessment of older patients, especially for clinicians evaluating those with abnormal cognitive decline, compared to lengthier and more difficult neuropsychological assessments (Randolph, Tierney, Mohr & Chase, 1998; Lim, Collinson, Feng & Ng, 2010). The RBANS was originally developed in English and has since been translated into over 20 languages by the developers (http://www.rbans.com/translations.html). Several other translations have been conducted and reported in peer-review articles including a Japanese (Takaiwa et al., 2006; Yamashima et al., 2002), Hungarian (Juhász, Kemény, Linka, Sántha & Bartkó, 2003) and French (Holzer et al., 2007) translation. Since the initiation of this study a Chinese translation has also been conducted in Shanghai, China (Cheng et al., 2011).

At the time of writing approximately 885 million (Grimes, 1996; Dalby, 1999; World Health Organisation, 1998) to 955 million (Nationalencyklopedin, 2010) of the world's population named the Chinese language as their primary language. This number rose to over 1 billion when second language speakers were included (Grimes, 1999). The estimates for primary speakers of English varied between approximately 335 and 470 million depending on the source (Grimes, 1999; Dalby, 1999; World Health Organisation, 1998; Nationalencyklopedin, 2010). Singapore is a multi-ethnic society with the 2010 census revealing that the majority of the population were of Chinese ethnicity (74.1%) (Singapore Department of Statistics, 2010). Mandarin, the official Chinese language, was the language most frequently spoken at home in Singapore (35.6%). A further 14.3% of the population most frequently spoke a Chinese dialect (mainly Hokkien, Teochew or Cantonese) and 32.3% most frequently spoke English at home (Singapore Department of Statistics, 2010). Furthermore, approximately 79.9% of the Singaporean population were literate in English, 60.8% were literate in Chinese and 45.2% were literate in both English and Chinese. Given the large global and local Chinese speaking population, its expected importance into the future and the predicated future necessity for people to be multilingual (Graddol, 2004), it is useful that Chinese translations of assessment scales are available. This is not only relevant to research within a society with a primarily Chinese population but also research in multi-ethnic societies and in multinational studies (Cheung et al., 2004).

Thorough and careful translation procedures do not guarantee measurement equivalence between a translated version and the original of any instrument. The Chinese and English languages are based on pictogram and alphabet, respectively. The two languages have very different linguistic structures (Cheung et al., 2004). It is not clear whether the original English version and a Chinese translation of the same instrument will provide equivalent test scores if the English and Chinese respondents have the same level of the target functions (Cheung et al., 2004). To date there has been no assessment of the equivalence or differences between the test scores on the English RBANS and a Chinese translation.

We have previously reported normative data for the RBANS on a total of 1,165 Chinese people aged 55-91 with varying levels of education and a range of languages in the Singapore Longitudinal Ageing Study - II (SLAS-II). Statistically significant associations between age, education, language and gender on performance on the RBANS subtests, index and total scale scores were found (Collinson, Fang, Lim, Feng & Ng, 2014). However, this study did not examine significant practical differences in the RBANS test performance between the language versions. In the present study, we evaluated the RBANS in a large sample of older Chinese people. The aim was to establish whether each of the Chinese translations (Mandarin, Hokkien, Teochew and Cantonese) of the RBANS achieved measurement equivalence to the English version in terms of similar mean scores between users who had similar demographic, clinical and cognitive profiles.

Method

Design and Recruitment

The design and methods of SLAS-II have been described previously (Lim et al., 2010). Trained research nurses conducted recruitment via door-to-door visits between March 2008 and October 2013. Consenting participants arranged to visit the research site at a later date to perform the RBANS, among other instruments (Collinson et al., 2014). Those with a Mini-Mental State Examination (MMSE) score below the cut-off of 26 underwent further assessments including the Clinical Dementia Rating (CDR) scale. The small proportion of non-Chinese participants in the SLAS-II dataset (Malays, Indians, and others) was excluded from this analysis.

The RBANS (Form A) was administered to the participants as part of the battery of neuropsychological and clinical assessments of the SLAS-II. The RBANS was administered once in the language–dialect (English, Mandarin, Hokkien, Teochew or Cantonese) according to the participants' dominant or habitual preference by trained research nurses who were fluent in that particular language–dialect (Collinson et al., 2014). Using the forward–backward translation procedure, the RBANS (Copyright # 1998. NCS Pearson, Inc. Reproduced with permission. All rights reserved. Mandarin 2 S.L. Collinson et al. Archives of Clinical Neuropsychology Downloaded from http://acn.oxfordjournals.org/ by guest on June 8, 2014 Chinese translation copyright # 2012. NCS Pearson, Inc. Translated and reproduced with permission.) had previously been translated into Mandarin and Chinese dialects (Hokkien, Teochew and Cantonese) by a committee of trained multilingual research psychologists (Collinson et al., 2014). Words and phrases that were deemed culturally unfamiliar were replaced with suitable local equivalents. For example in the story memory subtest the item ‘in Cleveland, Ohio’ was replaced with ‘in Kuala Lumpur’. Words that were not translatable or meaningful in Chinese were replaced with understandable and appropriate alternatives that where possible were semantically and phonetically as close as possible to the original English word. For example, in the list recognition subtest ‘velvet’ was replaced with ‘cotton’ and ‘meadow’ was replaced with ‘forest’ (Collinson et al., 2014). Individual test performance was scored according to standardised instructions (Randolph, 1998) except for the figure copy and figure recall subtests, which were scored according to the modified criteria suggested by Duff et al. (2003).

The MMSE was also administered as part of the battery of neuropsychological and clinical assessments. The MMSE was developed as a concise tool to assess patients' level of cognitive impairment and to monitor cognitive changes in patients over time (Folstein, Folstein & McHugh, 1975). It is extensively used in clinical practice and clinical trials today (Pangman, Sloan & Guse, 2000). A Chinese translation has previously been developed (Zhang et al., 1990; and Zhang, Yu, & He, 1995) and validated in a Singaporean population which was utilised in this study (Ng, Niti, Chiam & Kua, 2007). Modifications have previously been reported and included, for example, the replacement of the question "What city/town are we in?" with "What area are we in?" since Singapore is a city-state and the only correct answer for the original question would be Singapore (Feng, Lim & Ng, 2012).

Participants with a MMSE score below the cut-off of 26 underwent further assessments including the CDR, which is used to stage the severity of dementia based on clinical information from semi-structured interviews with the patient and collateral source (Hughes, Berg, Danziger, Coben & Martin, 1982; Morris, 1993 and 1997). It was originally developed for use in people with Alzheimer's but can be used to stage dementia in other illnesses. In Singapore cultural modifications for the judgement and problem solving domain have previously been reported. For example the question "How many nickels in a dollar?" was modified to "How many ten cents in two Singaporean dollars?" (Li, Ng, Kua & Ko, 2005).

Both the MMSE and the CDR were administered in the language–dialect according to the participants' dominant or habitual preference by trained research nurses who were fluent in that particular language–dialect

Statistical Analysis

Participants were included in the analysis if they had completed all RBANS subtests and reported their age. Only RBANS assessments administered by a research nurse who conducted at least 10

RBANS interviews in both English and Chinese (Mandarin and/or dialects) were included. The latter inclusion criterion was intended to allow statistical control for prevention of interviewer effect confounding the language comparison.

For the purpose of scoring the RBANS, participants were categorised according to their age group (54 to 59 years old, 60 to 64 years old, 65 to 69 years old, 70 to 74 years old and 75 years and above). The RBANS index scores were calculated by first standardising the 12 raw subtest scores across age categories using the means and standard deviations produced by Collinson et al. (2014). Population specific norms were used as opposed to the standard US norms due to evidenced differences in neuropsychological performance between Asian and Caucasian populations (Boone, Victor, Wen, Razani & Ponto´n, 2007; Fuji, 2010; Hedden et al., 2002; and Shan, Chen, Lee & Su, 2008). The age standardised scores were then transformed to scores with a mean of 100 and standard deviation of 15. Individual index scores were then calculated by summing the transformed scores of the subtests that contribute to that specific index and taking the mean. The total scale score was derived by taking the mean of the sum of the five index scores. All analysis was conducted on scores after they had been both age standardised and transformed.

An analysis of variance was used to compare the age, Geriatric Depression Scale (GDS) score, MMSE score and physical activity score across language versions of the RBANS. Chi-squared tests were used to compare categorical variables, gender, education level and CDR scale score across language versions of the RBANS. P-values were from global statistical tests that tested the global null hypothesis of all language groups having the same mean (or proportion) of the variable of interest.

We used the equivalence study approach for the comparison between the English and Chinese RBANS versions (Cheung et al., 2004; Senn, 1997). In this approach, if the confidence interval of the difference between two groups totally fell within an equivalence margin, equivalence was declared. This is in contrast to the study of difference, which may conclude a difference even if the difference is clinically non-significant but statistically significant. 90% confidence intervals (CI) for the adjusted differences were calculated and equivalence declared if the 90% CI fell entirely within the specified equivalence margin. It is conventional that, for equivalence studies, 90% instead of 95% CIs are used (Senn, 1997; Machin, Campbell, Fayers & Pinol, 1997).

The equivalence margins were based on Cohen's definition of effect size (Cohen, 1998). Cohen considered an effect size smaller than 0.2 SD an immaterial difference. Accordingly, we defined a difference in mean RBANS score of no more than +/- 0.2 SD an 'equivalence margin'. Cohen considered an effect size between 0.2 and 0.5 a small difference. Accordingly, we defined a difference of +/- 0.4 SD a 'practical equivalence margin'. The margins were calculated by multiplying the SD of the English scale scores by 0.2 or 0.4.

Regression models for each of the five RBANS index scores and the RBANS total scale score were fitted to compare scores between languages, with English as the reference language. To remove potential confounding by demographic and clinical characteristics of participants, demographic (age, gender, and education level) and clinical/health (GDS, MMSE, CDR and physical activity score) covariates were included. Furthermore, to prevent confounding by interviewer effects, a mixed model approach was used to include the research nurse as a random-effect (Cheung, 2014). The model was of the form:

$$y_i = (\alpha + \mu_j) + \beta_1\, Mandarin_i + \beta_2\, Hokkien_i + \beta_3\, Teochew_i + \beta_4\, Cantonese_i + \beta_5\, Age_i + \beta_6\, Female_i + \beta_7\, Education1_i + \beta_8\, Education2_i + \beta_9\, Education3_i + \beta_{10}\, Education4_i + \beta_{11}\, GDS_i + \beta_{12}\, MMSE_i + \beta_{13}\, CDR1_i + \beta_{14}\, CDR2_i + \beta_{15}\, CDR3_i + \beta_{16}\, CDR3_i + \beta_{17}\, CDR4_i + \beta_{18}\, Physical_i + \varepsilon_i$$

where i indicated participant and y was the outcome (either one of the five RBANS index scores or the RBANS total scale score). β_1 through to β_{18} were included as fixed parts. *Mandarin*, *Hokkien*, *Teochew* and *Cantonese* coefficients represented the specific language compared to English. *Age* represented continuous variable for age. The *Female* coefficient represented female compared to male. *Education1*, *Education2*, *Education3* and *Education4* coefficients represented 1-3 years, 4-6 years, 7-10 years and >10 years of education compared to 0 years. *GDS* and *MMSE* represented continuous variable for GDS and MMSE scores. *CDR1*, *CDR2*, *CDR3* and *CDR4* coefficients represented CDR scale values 0.5, 1, 2 and 3 compared to 0. *Physical* coefficient represented continuous score for physical activity measurement. μ_j was the random effect for nurse and ε_i denoted residuals. The models gave us estimates of the adjusted differences and their standard errors (SE) for the RBANS index scores and total scale score for each Chinese language version compared to the English version, controlling for the demographic and clinical characteristics. The 90% CI was calculated as the estimated adjusted differences +/- z(0.90)×SE, where z(0.90) was the critical value at the 90th percentile of the standard normal distribution.

If the 90% CI for the language specific, adjusted difference from the regression model yielded a mean score of no more than 0.2 (0.4) SD higher or lower than the mean score of the English version, (practical) equivalence was declared. If equivalence and practical equivalence could not be declared for an index score (or the total scale score), the specific subtests that contributed to that index score were assessed for equivalence following the methods outlined above.

CDR was assumed to be 0 for those who had MMSE ≥ 26, which was the study's cut-off for including the participants in the CDR assessment. We ran sensitivity analyses by refitting the above models on participants at the upper end of cognitive performance based on both the CDR assessment and the MMSE, as alternative approach to prevent the potential confounding effect of cognitive function. The criteria we used to include participants in each of these analyses were: (1) those with CDR scores = 0 or MMSE ≥ 26 with no CDR assessment; (2) those with MMSE scores ≥ 24; and (3) those with MMSE scores ≥ 26.

All analyses were conducted using Stata version 13.1 (StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP).

Results

Descriptive Summary

Of 2,428 English and Chinese speaking participants that consented to participate in the SLAS-II study, 572 did not meet the pre-specified inclusion criteria (139 participants failed to complete the RBANS completely, 45 did not report age and 388 participants were interviewed by a research nurse who did not satisfy the criterion for number of interviews conducted), leaving a total of 1,856 (76.4%) participants available for analysis.

Table 1 shows the descriptive characteristics of the sample. Chinese speakers, particularly dialect speakers tended to be female (global $p<0.001$), older (global $p<0.001$), received less education (global $p<0.001$), had lower physical activity scores (global $p=0.002$), had lower MMSE scores (global $p<0.001$) and poorer CDR scores (global $p<0.001$). *[Include table 1 here.]*

Table 2 shows the mean and SD of the unadjusted RBANS index scores, total scale score and subtest scores by language version of the RBANS administered. English and Mandarin speakers tended to show higher RBANS scores across indexes and subtests, Cantonese speakers had marginally smaller scores than the English and Mandarin speakers, and the Hokkien and Teochew speakers had the lowest scores. *[Include table 2 here.]*

Measurement Equivalence

Table 3 shows the equivalence margins for each of the RBANS index scores and total scale score. The equivalence margins, defined as 0.2 of the English language SD, were ± 2.78, ± 2.18, ± 4.04, ± 2.69, ± 3.12 and ± 2.12 for the immediate memory index, visuospatial/constructional index, language index, attention index, delayed memory index and total scale score, respectively. The corresponding practical equivalence margins, defined as 0.4 of the English language SD, were ± 5.57, ± 4.36, ± 8.07, ± 5.38, ± 6.25 and ± 4.23, respectively.

English vs. Mandarin

After adjusting for covariates, the mean difference between the English and Mandarin visuospatial/constructional index (-0.42) and delayed memory index (-1.26) and the corresponding 90% CIs fell completely within the equivalence margins suggesting equivalence. Furthermore, the adjusted mean difference for the immediate memory index and total scale score (2.78 and 2.57, respectively) and the respective 90% CIs suggested that the English and Mandarin language versions were practically equivalent. For the language index, the 90% CI did not show conclusive evidence of equivalence or difference. The 90% CI for the attention index fell totally above the equivalence margin but included the practical equivalence margin, showing that the Mandarin language version gave at least slightly larger mean attention scores.

English vs. Hokkien

The adjusted mean difference between the English and Hokkien immediate memory index (2.24), language index (3.88), delayed memory index (-1.56) and the total scale score (1.42) and the 90% CIs were indicative of practical equivalence. The Hokkien language version gave at least slightly smaller mean visuospatial/constructional scores and at least slightly larger mean attention scores than the English language version. This was indicated by the 90% CIs for the visuospatial/constructional index and the attention index which fell entirely outside of the stricter equivalence margins.

English vs. Teochew

The results for the English and Teochew language versions indicated practical equivalence with regards to the immediate memory index, the language index, the delayed memory index and total scale score (adjusted mean differences 2.78, 4.25, -2.66 and 1.15, respectively). The 90% CIs for the visuospatial/constructional and attention indexes were wholly outside of the stricter equivalence margins but included the practical equivalence margins, showing that the Teochew language version gave at least marginally lower mean visuospatial/constructional scores and marginally higher attention scores.

English vs. Cantonese

The results for the Cantonese language version suggested equivalence to the English language version for the visuospatial/constructional index and delayed memory index (adjusted mean difference 0.07 and 0.26, respectively). The Cantonese language version gave at least marginally larger mean immediate memory, language and total scale scores than the English language version, as indicated by the 90% CIs that fell completely above their corresponding stricter equivalence margins but included their corresponding practical equivalence margins. The adjusted mean difference between the English language version and the Cantonese language version of the attention index (7.61) and 90% CI (6.21, 9.00) showed no overlap with the margin of practical equivalence (± 5.38), pointing to a difference.

[Include table 3 here.]

Table 4 shows the equivalence margins, adjusted mean differences and 90% CIs for the RBANS subtests that contribute to the indexes where equivalence could not be confirmed.

The immediate memory index is comprised of the list learning and story memory subtests. Results were indicative of practical equivalence for the story memory subtests between the English and Cantonese language versions; however the result for the list learning subtest was inconclusive.

The visuospatial/constructional index consists of the figure copy and line orientation subtests. The result for the comparison between the English and Hokkien figure copy subtest was inconclusive. However the comparison for the line orientation subtest showed that the Hokkien language version gave at the least marginally smaller mean line orientation scores than the English version. The comparison between the English and Teochew line orientation subtest was inconclusive. However the English and Teochew figure copy comparison showed that the Teochew language version gave at least slightly lower mean figure copy scores than the English language version.

The picture naming and semantic fluency subtests make up the language index. The results for the Mandarin and Cantonese language versions were suggestive of equivalence and practical equivalence, respectively, to the English language version for the semantic fluency subtest (adjusted mean difference -0.20 and 0.52, respectively). However, the comparison for the picture naming subtest revealed that the Mandarin and Cantonese language versions gave at least slightly higher mean picture naming scores than the English language version.

The attention index includes the digit span and coding subtests. The adjusted mean difference between the English language version and each of the Chinese language versions of the digit span subtest and 90% CIs all fell entirely outside of the margin of practical equivalence (± 1.11), pointing towards a difference for each language. In contrast, the evidence for the Mandarin and Cantonese language versions of the coding subtest were suggestive of practical equivalence to the English language version; the result for the Teochew language version was inconclusive; and the Hokkien language version gave at least slightly lower mean coding scores than the English language version.

[Include table 4 here.]

The unformatted, Stata output for the regression analysis results are provided in the online supplementary material.

Each of the sensitivity analyses restricted to cognitively healthy participants gave similar results with conclusions remaining largely unchanged. The only difference of note was that in each of the sensitivity analyses the evidence for the English and Mandarin language versions supported a conclusion of practical equivalence for the language index as compared to the inconclusive result previously obtained. The results of which are provided in the online supplementary material.

Discussion

The use of the Chinese language is on the rise (Grimes, 1999; Dalby. 1999; World Health Organisation, 1998; Nationalencyklopedin, 2010), as is the need for multilingualism and multinational clinical studies (Graddol, 2004; Cheung et al., 2004). For cross-cultural comparisons and multi-country studies, there is a demand for the assessment of measurement equivalence between Chinese translations of instruments to their originals (Thumbool et al., 2002). The RBANS was originally developed in English and is an important tool for neuropsychological assessments. An ongoing process is the development of the RBANS into different language versions, at the time of writing there were over 20 different language versions (http://www.rbans.com/translations.html) (Takaiwa et al., 2006; Yamashima, et al., 2002; Juhász et al., 2003; Holzer et al., 2007). Since the initiation of this study a team in Shanghai has also reported on the reliability and validity of a Chinese (Mandarin) translation in a community dwelling elderly sample in Shanghai, China (Cheng et al., 2011). However, it did not include an English version of the RBANS. So, unlike the present study, it was unable to investigate the measurement equivalence. To date there have been no studies looking at the equivalence of the English RBANS and a Chinese translation.

In this analysis we reported the equivalence and differences between the English RBANS and Mandarin, Hokkien, Teochew and Cantonese translations. These dialects exist in immigrant Chinese communities outside of mainland China as a consequence of the Chinese diaspora of the 19^{th} and 20^{th} centuries. We calculated the 90% CI of the adjusted mean differences of the index scores between the English version and each of the Chinese versions and evaluated them against pre-specified equivalence margins. We assessed the indexes first and then assessed the individual subtests that contributed to that index if equivalence could not be confirmed. This reduced the number of tests performed and helped guard against the risk of false findings due to chance (Cheung et al., 2004).

The Mandarin, Hokkien and Teochew versions of the total scale score all achieved practical equivalence with the English version. Equivalence could not be confirmed for the Cantonese version. Measurement equivalence in the total scale score across several language versions of the RBANS was an important finding. In situations where a primary endpoint is needed, such as in clinical trials of interventions to promote general cognitive functioning, the total scale may be the choice unless there are specific hypothesis pertaining to specific aspects of cognition.

The delayed memory index achieved at least practical equivalence across each language comparison. Therefore the RBANS is suitable for studies of delayed memory in this and other multi-language societies.

The Mandarin and Cantonese versions of the visuospatial/constructional index were equivalent to the English version. The Hokkien and Teochew languages were not. Examination of the subtests that make up the visuospatial/constructional index attributed this to the non-comparability across both the figure copy and line orientation subtests.

The Cantonese version of the attention index was confirmed to be different to the English version. The comparison of the English version to the Mandarin, Hokkien and Teochew versions suggested at least a small increase in mean attention index score on each of the Chinese versions. Examination of the attention subtests found this to be due to differences in the digit span subtest, where we were observed a difference to the English version for all Chinese language versions. Differences in performance on numerical based tasks between English and Chinese speakers have been well documented (Stigler, Lee & Stevenson, 1986; Hedden et al., 2002; Chen, Cowell, Varley & Wang, 2009; Ting, Hameed, Tan, Gabriel & Doshi, 2014). Specifically the literature supports our finding that Chinese speakers have larger digit spans than their English speaking counterparts (Hedden et al., 2002; Chen et al., 2009). This difference has been attributed to the shorter articulation time needed for Chinese digits compared to English digits, which in turn results in a lower processing load (Stigler, 1986; Hedden et al., 2002; Cheung & Kemp, 1993, 1994).

There were several limitations to our study. Firstly, there are many alternative ways to define equivalence. However, basing the conservative equivalence margin on Cohen's effect size of 0.2 and the practical equivalence margin on an effect size of 0.4 was in line with the literature (Cheung et al., 2004; Thumbool et al., 2002). More importantly, the reporting of SDs for all scores and CIs for adjusted mean differences allows readers to draw their own conclusion using their own equivalence margins. Secondly, there is no theoretical foundation for the use of the 90% CI. Its use stems from current accepted practice for equivalence studies and we used it for this reason alone.

Data for this analysis came from the SLAS-II study with each participant contributing at one time point using one language version of the RBANS. Cross-over designs, where each participant completes the assessment in both the original language and the translated version with a pre-defined period between assessments, are often considered a better design for equivalence studies (Thumbool et al, 2002). However, this would necessitate the use of bilingual respondents which would reduce the generalisability of findings because bilingual respondents tend to be younger and more educated (Cheung et al., 2004).

Observational studies are candidates for several sources of bias. These include: confounders, which relate to factors that may be associated with outcomes that have not been accounted for; selection bias, which relates to how participants are recruited or the type of participants recruited; and information bias, which relates to errors in the recording of data (Hammer, du Prel & Blettner, 2009).

Analysis was conducted by fitting mixed effects models allowing adjustments for differences in characteristics among participants, as well as controlling for interviewer effects by including them as random effects. One could argue that not all confounders were considered and there may still be the risk of residual confounding (Cheung et al., 2004). However, we controlled for key demographics and indicators of cognition. We also accounted for potential dependence between language versions and interviewers. Unfortunately most research nurses did not conduct assessments in more than two to three languages/dialects which resulted in a highly unbalanced incomplete block design. Therefore we were unable to examine a possible interviewer-language interaction i.e. did interviewers tend to obtain different scores when using different languages/dialects. Inclusion of covariates such as GDS, MMSE and CDR that were likely to be highly correlated could result in unstable and difficult to interpret coefficients for their respective coefficients. However the collinearity does not affect the coefficients of the other variables that are not highly correlated with them. Since we were only interested in the coefficients for each specific language, multicollinearity between GDS, MMSE and CDR was of no concern (Montgomery, Peck & Vining, 2001).

Recruitment was conducted via door to door visits and specifically targeted senior activity centres, elderly day care centres and elderly homes across different community neighbourhoods in Singapore. This was to ensure that the sample population were truly representative of the target population i.e. Singaporean, elderly population. Our sample age ranged from 54-94 years of age and the language preference of participants closely followed the distribution of languages spoken most frequently at home in those of Chinese ethnicity, aged 55 years and over according to the 2010 Singapore Census. However given that our analysis is about the characteristics of different language versions of the RBANS and not about the characteristics of the sample, representativeness of the sample is not critical.

Information bias would be present if the study was subject to errors in the recording of information. The research nurse performed the RBANS overseen by research psychologists according to research protocols. This level of supervision and monitoring would have kept inaccuracies in the recording of data to a minimum.

There may be concerns regarding the inclusion of participants with cognitive impairment or dementia in a psychometric equivalence study. However this is typical of the population the RBANS will be used on and therefore we felt it was important to include them so that the findings are applicable in a real world setting. Sensitivity analyses including only cognitively normal participants suggested the robustness of the findings.

The English and Chinese language versions of the RBANS showed a mixture of results in terms of equivalence. Overall, it is reasonable to consider the English and Chinese versions of the RBANS comparable in terms of measurement equivalence on the total scale score, delayed memory index, immediate memory index and language index, but not the attention index (for all Chinese dialects), and the visuospatial/constructional index (for Hokkien and Teochew dialects). Data for the comparison of neuropsychological performance for the equivalent indexes from English and Chinese speakers may be pooled and anyone of these language versions could be used for bilingual speakers. In the short term, analyses of attention and visuospatial/constructional performance that need to pool data from different RBANS language versions should include a covariate that represents the version in the statistical adjustment. The development of normative values should either be version specific, or remove the estimated (covariate-adjusted) difference from the English version scores as part of the normalisation process. In the long-term, the sources of the differences between language versions of the attention and visuospatial/constructional indexes should be examined in detail. A new round of translation/adaptation, taking into account the lessons learnt, should be considered.

Acknowledgements

References

Badenes, G.D., Casas, H.L., Cejudo, B.J.C., & Aguilar, B.M. (2008). Evaluation of the capacity to drive in patients diagnosed of mild cognitive impairment and dementia. (Article in Spanish). Neurologia, 23(9), 575-582.

Boone, K.B., Victor, T.L., Wen, J., Razani, J. & Ponto´n, 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. doi:10.1016/j.acn.2007.01.010.

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, 31(4), 385-391. doi:10.1080/13803390802195195.

Cheng, Y., Wu, W., Wang, J., Feng, W., Wu, X. & Li, C. (2011). Reliability and validity of the Repeatable Battery for the Assessment of Neuropsychological Status in community-dwelling elderly. Archives of Medical Science, 7(5), 850-857. doi:10.5114/aoms.2011.25561.

Cheung, H. and Kemper, S. (1993). Recall and articulation of English and Chinese words by Chinese-English bilinguals. Memory and Cognition, 21, 666-670. doi:10.3758/BF03197197.

Cheung, H. and Kemper, S. (1994). Recall and articulation of English and Chinese words under memory preload conditions. Language and Speech, 37(2), 147-161. doi:10.1177/002383099403700203.

Cheung, Y.B., Thumboo, J., Goh, C., Khoo, K.S., Che, W. & Wee, J. (2004). The equivalence and difference between the English and Chinese versions of two major, cancer-specific, health- related quality of life questionnaires. Cancer, 101, 2874-2880. doi:10.1002/cncr.20681.

Cheung, Y.B. (2014). Statistical analysis of Human Growth and Development. Chapman & Hall/CRC Biostatistics Series.

Cohen, J. (1998). Statistical power analysis for the behavioural sciences, 2nd ed. Erlbaum, Hillsdale, NJ.

Collinson, S.L., Fang, S.H., Lim, M.L., Feng, L., & Ng, T.P. (2014). Normative Data for the Repeatable Battery for the Assessment of Neuropsychological Status in Elderly Chinese. Archives of Clinical Neuropsychology, 29(5), 442-455. doi:10.1093/arclin/acu023.

Dalby, D. The Linguasphere Register of the World's Languages and Speech Communities. (1999). Wales: Linguasphere Press.

Duff, K., Patton, D., Schoenberg, M.R., Mold, J., Scott, J.G., & Adams, R.L. (2003). Age and education-corrected independent normative data for the RBANS in a community dwelling elderly sample. The Clinical Neuropsychologist, 17, 351-366. doi:10.1076/clin.17.3.351.18082.

Feng, L., Lim, M.S.C.W.S., & Ng, T.P. (2012) The Modified Mini-Mental State Examination test: normative data for Singapore Chinese older adults and its performance in detecting early cognitive impairment. Singapore Medical Journal, 53(7), 458-462.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12, 189-198. doi:10.1016/0022-3956(75)90026-6.

Fuji, D.E.M. (2010). The neuropsychology of Asian-Americans. London: Psychology Press.

Graddol, D. (2004). The future of language. Science, 303, 1329-1331. doi:10.1126/science.1096546.

Grimes, B., ed. Ethnologue: Languages of the World, 13th ed. (1996). Dallas: Summer Institute of Linguistics Inc. (Information updated in February 1999.)

Hammer, G.P., du Prel J.B. & Blettner, M. (2009). Avoiding bias in observational studies. Deutsches Arzteblatt International, 106(41), 664-668. doi:10.3238/arztebl.2009.0664.

Hedden, T., Park, D.C., Nisbett, R., Ji, L.J., Jing, Q. & Jiao, S. (2002). Cultural variation in verbal versus spatial neuropsychological function across life span. Neuropsychology, 16(1), 65-73. doi:10.1037//0894-4105.16.1.65.

Holzer, L., Chinet, L., Jaugey, L., Plancherel, B., Sofia, C., Halfon, O., Randolph, C. (2007). Detection of cognitive impairment with the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) in adolescents with psychotic symptomatology. Schizophrenia Research, 95 (1-3), 48-53. doi:10.1016/j.schres.2007.06.013.

Hughes, C.P., Berg, L., Danziger, W.L., Coben, L.A. & Martin, R.L. (1982) A new clinical scale for the staging of dementia. The British Journal of Psychiatry, 140, 566-572. doi:10.1192/bjp.140.6.566.

Juhász, L.Z., Kemény, K., Linka, E., Sántha, J., & Bartkó, G. (2003). The use of RBANS test (Repeatable Battery for the Assessment of Neuropsychological Status) in neurocognitive testing of patients suffering from schizophrenia and dementia. (Article in Hungarian). Clinical Neuroscience/Ideggyogyaszati Szemle, 56 (9-10), 303-308.

Kotani, S., Sakaguchi, E., Warashina, S., Matsukawa, N., Ishikura, Y., Kiso Y., … Yamashima, T. (2006). Dietary supplementation of arachidonic and docosahexaenoic acids improves cognitive dysfunction. Neuroscience Research, 56, 159-164. doi:10.1016/j.neures.2006.06.010.

Li, M., Ng, T.P., Kua, E.H., & Ko, S.M. (2005). Brief informant screening test for Mild Cognitive Impairment and early Alzheimer's Disease. Dementia and Geriatric Cognitive Disorders, 21, 391-402. doi:10.1159/000092808.

Lim, M.L., Collinson, S.L., Feng, L., & Ng, T.P. (2010). Cross-cultural application of the Repeatable Battery for the assessment of Neuropsychological Status (RBANS): Performances of elderly Chinese Singaporeans. The Clinical Neuropsychologist, 24(5), 811-826. doi:10.1080/13854046.2010.490789.

Machin, D., Campbell, M., Fayers, P. & Pinol, A. (1997). Sample size tables for clinical studies, 2nd ed. Oxford: Blackwell.

Mahncke, H.W., Conner, B.B., Appelman, J., Ahsanuddin, O.N., Hardy, J.L., Wood, R.A., … Merzenich, M.M. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. Proceedings of the National Academy of Sciences of the United States of America, 103(33), 12523-12528. doi:10.1073pnas.0605194103.

Montgomery, D.C., Peck, E.A. & Vining, C.G. (2001). Introduction to Linear Regression Analysis, 3rd ed. New York, Wiley.

Morris, J.C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology, 43(11), 2412–2414.

Morris, C. (1997). Clinical Dementia Rating: A reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. International Psychogeriatrics, 9 (Suppl. 1), 173-176.

Nationalencyklopedin 'Världens 100 största språk 2007' [The world's 100 largest languages in 2007]. (2010). In Nationalencyklopedin. Retrieved from http://www.ne.se/uppslagsverk/figur/tabell/världens-100-största-språk-2010.

Ng, T.P., Niti, M., Chiam, P.C., Kua, E.H. (2007). Ethnic and educational differences in cognitive test performance on Mini-Mental State Examination in Asians. American Journal of Geriatric Psychiatry, 15(2), 130-139. doi:10.1097/01.JGP.0000235710.17450.9a.

Pangman, V.C., Sloan, J. & Guse, L. (2000). An examination of psychometric properties of the Mini-Mental State Examination and the standardized Mini-Mental State Examination: implications for clinical practice. Applied Nursing Research, 13(4), 209-213. doi:10.1053/apnr.2000.9231.

Randolph, C. (1998). Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). San Antonio: Harcourt, TX: The Psychological Corporation.

Randolph, C., Tierney M.C., Mohr E., & Chase T.N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary Clinical Validity. Journal of Clinical and Experimental Neuropsychology, 20(3), 310-319. doi:1380-3395/98/2003-310$12.00.

Senn, S. (1997). Statistical issues in drug development. Chichester: John Wiley & Sons, Inc.

Shan, I.K., Chen, Y.S., Lee, Y.C., & Su, T.P. (2008). Adult normative data of the Wisconsin card sorting test in Taiwan. Journal of the Chinese Medical Association, 71(10), 517-522. doi:10.1016/S1726-4901(08)70160-6.

Singapore Department of Statistics. (2010). Census of Population 2010 Statistical Release 1 Demographic Characteristics, Education, Language and Religion. Retrieved from http://www.singstat.gov.sg/publications/publications_and_papers/cop2010/census_2010_release1/cop2010sr1.pdf.

Stigler, J.W., Lee, S.Y. & Stevenson, H.W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. Cognition, 23, 1-20. doi:10.1016/0010-0277(86)90051-X.

Takaiwa, A., Kuwayama, N., Hayashi, N., Kubo, M. Matsui, M., & Endo, S. (2006). Cognitive function in patients with severe carotid stenosis – evaluation of RBANS, WAIS-R and NART before treatment of carotid revascularization. (Article in Japanese). No To Shinkei, 58(8), 681-686.

Thumbool, J., Fong, K.Y., Chan, S.W., Machin, D., Feng, P.H., Thiol, S. & Boey, M.L. (2002). The equivalence of English and Chinese SF-36 versions in bilingual Singapore Chinese. Quality of Life Research 11, 495-503. doi:10.1023/A:1015680029998

Ting, S.K.S., Hameed, S., Tan, E.K., Gabriel, C. & Doshi, K. (2014). Digit span: a comparison of Chinese versus alphabetic language speakers in dysexecutive dementia patients. Journal of Neurological, Neurosurgery and Psychiatry, 85, 117-118. doi:10.1136/jnnp-2013-305828.

World Health Organization. Reaching out to the largest audience: languages for communication in WHO EB105/20. (1999). Geneva: World Health Organization.

Yamashima, T., Yoshida, M., Kumahashi, K., Matsui, M., Koshino, Y., Higashima, M., … Matsushita, M. (2002). The Japanese version of RBANS (Repeatable Battery for the Assessment of Neuropsychological Status). (Article in Japanese). No To Shinkei, 54(6), 463-471.

Zhang, M.Y., Katzman, R., Salmon, D., Jin H., Cai, G., Wang Z.,...... Liu, W.T. (1990). The prevalence of dementia and Alzheimer's disease in Shanghai, China: impact of age, gender, and education. Annals of Neurology, 27, 428-37. doi:10.1002/ana.410270412.

Zhang, M.Y., Yu, E. & He, Y.L. (1995). [Instruments for the epidemiological survey of dementia and their applications.] Shanghai Psychiatry, 7, 3-5. Chinese.

Tables

Table 1: Demographic characteristics by dominant language

Dominant language	English		Mandarin		Hokkien		Teochew		Cantonese		P-value [a]
Variable	n		n		n		n		n		
Age - years, mean (SD)	402	63.61 (7.04)	471	64.73 (6.75)	376	67.89 (7.42)	120	67.59 (8.02)	487	68.42 (7.65)	< 0.001
GDS score, mean (SD)	402	0.79 (1.78)	471	0.54 (1.10)	376	0.78 (1.30)	120	0.62 (0.84)	487	0.83 (1.74)	0.019
MMSE, mean (SD)	398	28.94 (1.61)	468	28.76 (1.49)	370	27.13 (2.76)	119	28.05 (2.66)	487	27.56 (2.76)	< 0.001
Physical Activity Score, mean (SD)	384	9.39 (4.34)	455	9.34 (4.31)	366	8.49 (4.03)	115	8.91 (4.03)	471	8.51 (4.13)	0.002
Gender, n (%)											
Male	191	(47.51)	153	(32.48)	130	(34.57)	35	(29.17)	182	(37.37)	< 0.001
Female	211	(52.49)	318	(67.52)	246	(65.43)	85	(70.83)	305	(62.63)	
Education, n (%)											
0 years	6	(1.49)	41	(8.70)	140	(37.23)	38	(31.67)	118	(24.23)	< 0.001
1-3 years	24	(5.97)	59	(12.53)	81	(21.54)	29	(24.17)	96	(19.71)	
4-6 years	49	(12.19)	167	(35.46)	105	(27.93)	39	(32.50)	164	(33.68)	
7-10 years	209	(51.99)	153	(32.48)	36	(9.57)	12	(10.00)	94	(19.30)	
More than 10 years	111	(27.61)	49	(10.40)	14	(3.72)	2	(1.67)	14	(2.87)	
CDR, n (%)											< 0.001
0 - No dementia	370	(92.04)	439	(93.21)	291	(77.39)	99	(82.50)	402	(82.55)	
0.5 - CIND	25	(6.22)	23	(4.88)	59	(15.69)	14	(11.67)	55	(11.29)	
1 - Mild dementia	1	(0.25)	1	(0.21)	8	(2.13)	2	(1.67)	11	(2.26)	
2 - Moderate dementia	0	(0.00)	1	(0.21))	0	(0.00)	0	(0.00)	1	(0.21)	

SD: standard deviation; GDS: geriatric depression scale; MMSE: mini-mental state examination; CDR: clinical dementia rating; CIND: cognitive impairment no dementia.
[a] Differences in mean age, GDS score and MMSE score were tested with analysis of variance and differences in categorical variables were tested by chi-squared analysis. P-values are from global statistical tests that tested the global null hypothesis of all language groups having the same mean (or proportion) of the variable of interest.

Table 2: Summaries of the unadjusted Repeatable Battery for the Assessment of Neuropsychological Status indexes, total scale score and subtests by language version administered

Dominant language	English		Mandarin		Hokkien		Teochew		Cantonese	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Indexes										
Immediate Memory	96.73	13.92	99.19	13.39	91.90	14.89	94.19	15.41	95.76	13.48
Visuospatial/Constructional	103.63	10.90	102.44	11.38	90.68	15.51	91.63	15.99	98.41	13.71
Language	94.88	20.18	101.40	11.34	94.74	21.77	97.43	28.01	99.03	13.38
Attention	98.37	13.45	101.10	11.28	94.94	11.64	96.24	13.59	99.29	11.71
Delayed Memory	97.79	15.62	96.21	14.16	90.42	15.95	91.58	22.36	93.63	16.33
Total Scale	98.28	10.58	100.07	8.40	92.54	11.31	94.21	15.49	97.22	9.90
Subtests										
List Learning	26.74	4.86	27.48	4.88	24.81	5.18	24.78	5.48	26.27	4.96
Story Memory	16.19	4.00	16.38	3.63	14.30	4.35	15.48	4.35	14.82	4.21
Figure Copy	18.02	2.23	17.69	2.58	15.63	3.85	15.83	3.48	16.83	3.23
Line Orientation	16.21	3.19	15.79	3.16	12.30	4.06	12.65	4.43	14.22	3.87
Picture Naming	9.68	0.69	9.94	0.30	9.73	0.76	9.79	0.97	9.86	0.47
Semantic Fluency	16.63	3.84	16.60	3.69	15.45	3.61	16.12	3.67	15.82	3.34
Digit Span	12.11	2.78	13.76	2.21	13.62	2.11	13.68	2.66	13.96	2.06
Coding	38.94	12.77	32.95	12.12	21.19	13.33	23.58	13.17	25.83	13.76
List Recall	6.91	2.34	6.92	2.16	5.63	2.72	5.99	2.80	6.14	2.61
List Recognition	19.50	1.15	19.36	1.14	19.14	1.35	19.12	2.08	19.14	1.44
Story Recall	9.12	2.42	9.04	2.08	7.39	2.86	7.99	2.92	7.87	2.78
Figure Recall	13.17	4.41	12.00	4.48	9.10	4.59	9.61	4.82	11.01	4.89

SD: standard deviation

Table 3: Repeatable Battery for the Assessment of Neuropsychological Status indexes and total scales equivalence margins and adjusted difference between language versions (n = 1732)

Index	Equivalence/Practical Equivalence [a]	Adjusted Difference [b]	90% CI	Conclusion [c]
Immediate Memory	± 2.78			
	± 5.57			
English vs. Mandarin		2.78	1.20, 4.35	PE
English vs. Hokkien		2.24	0.36, 4.13	PE
English vs. Teochew		2.78	0.33, 5.23	PE
English vs. Cantonese		4.86	3.25, 6.46	
Visuospatial/Constructional	± 2.18			
	± 4.36			
English vs. Mandarin		-0.42	-1.93, 1.09	E
English vs. Hokkien		-4.32	-6.11, -2.54	
English vs. Teochew		-5.66	-8.01, -3.32	
English vs. Cantonese		0.07	-1.47, 1.62	E
Language	± 4.04			
	± 8.07			
English vs. Mandarin		5.96	3.79, 8.12	
English vs. Hokkien		3.88	1.32, 6.44	PE
English vs. Teochew		4.25	0.89, 7.61	PE
English vs. Cantonese		9.30	7.09, 11.51	
Attention	± 2.69			
	± 5.38			
English vs. Mandarin		5.51	4.14, 6.88	
English vs. Hokkien		6.62	4.99, 8.25	
English vs. Teochew		6.88	4.75, 9.01	
English vs. Cantonese		7.61	6.21, 9.00	D
Delayed memory	± 3.12			
	± 6.25			
English vs. Mandarin		-1.26	-3.08, 0.55	E
English vs. Hokkien		-1.56	-3.65, 0.54	PE
English vs. Teochew		-2.66	-5.46, 0.14	PE

English vs. Cantonese		0.26	-1.61, 2.14	E
Total scale score	± 2.12 ± 4.23			
English vs. Mandarin		2.57	1.46, 3.68	PE
English vs. Hokkien		1.42	0.11, 2.72	PE
English vs. Teochew		1.15	-0.57, 2.87	PE
English vs. Cantonese		4.41	3.28, 5.54	

90% CI: 90% confidence interval; SD: standard deviation; GDS: geriatric depression scale; MMSE: mini-mental state examination; CDR: clinical dementia rating; E – Equivalence to the English language version confirmed; PE – Practical Equivalence to the English language version confirmed; D – Difference to the English language version confirmed.

Note: Cohen considered an effect size smaller than 0.2 SD an immaterial difference. Accordingly, we defined a difference in mean RBANS score of no more than +/- 0.2 SD an 'equivalence margin'. Cohen considered an effect size between 0.2 and 0.5 a small difference. Accordingly, we defined a difference of +/- 0.4 SD a 'practical equivalence margin'.

[a] The equivalence margin is the English scale score SD multiplied by 0.2. The practical equivalence margin is the English scale score SD multiplied by 0.4.

[b] The English mean scores minus the non-English scores adjusted for age, gender, education level, GDS score, MMSE score, CDR scale score and physical activity score. Research nurse included as a random effect.

[c] Blank space in the conclusion column indicates an inconclusive result.

Table 4: Repeatable Battery for the Assessment of Neuropsychological Status subtests equivalence margins and adjusted difference between language versions (n = 1732)

Index	Equivalence/ Practical Equivalence [a]	Adjusted Difference [b]	90% CI	Conclusion [c]
List Learning	± 0.97			
	± 1.94			
English vs. Cantonese		1.38	0.80, 1.95	
Story Memory	± 0.80			
	± 1.60			
English vs. Cantonese		1.05	0.59, 1.51	PE
Figure Copy	± 0.45			
	± 0.89			
English vs. Hokkien		-0.65	-1.03, -0.27	
English vs. Teochew		-1.05	-1.57, -0.54	
Line Orientation	± 0.64			
	± 1.28			
English vs. Hokkien		-1.20	-1.69, -0.71	
English vs. Teochew		-1.23	-1.87, -0.59	
Picture Naming	± 0.14			
	± 0.28			
English vs. Mandarin		0.26	0.19, 0.34	
English vs. Cantonese		0.34	0.26, 0.41	
Semantic Fluency	± 0.77			
	± 1.54			
English vs. Mandarin		-0.20	-0.62, 0.23	E
English vs. Cantonese		0.52	0.08, 0.95	PE
Digit Span	± 0.56			
	± 1.11			
English vs. Mandarin		2.05	1.76, 2.34	D
English vs. Hokkien		2.63	2.28, 2.98	D
English vs. Teochew		2.68	2.23, 3.13	D

English vs. Cantonese		2.54	2.25, 2.84	D
Coding	± 2.55 ± 5.11			
English vs. Mandarin		-2.98	-4.21, -1.76	PE
English vs. Hokkien		-4.29	-5.75, -2.82	
English vs. Teochew		-3.56	-5.47, -1.65	
English vs. Cantonese		-2.57	-3.82, -1.33	PE

90% CI: 90% confidence interval; SD: standard deviation; GDS: geriatric depression scale; MMSE: mini-mental state examination; CDR: clinical dementia rating; E – Equivalence to the English language version confirmed; PE – Practical Equivalence to the English language version confirmed; D – Difference to the English language version confirmed.

Note: Cohen considered an effect size smaller than 0.2 SD an immaterial difference. Accordingly, we defined a difference in mean RBANS score of no more than +/- 0.2 SD an 'equivalence margin'. Cohen considered an effect size between 0.2 and 0.5 a small difference. Accordingly, we defined a difference of +/- 0.4 SD a 'practical equivalence margin'.

[a] The equivalence margin is the English scale score SD multiplied by 0.2. The practical equivalence margin is the English scale score SD multiplied by 0.4.

[b] The English mean scores minus the non-English scores adjusted for age, gender, education level, GDS score, MMSE score, CDR scale score and physical activity score. Research nurse included as a random effect.

[c] Blank space in the conclusion column indicates an inconclusive result.