Check for updates

Original Research Article

American Journal of Alzheimer's Disease & Other Dementias®
Volume 39: 1–11
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15333175241271910
journals.sagepub.com/home/aja
Sage

Development and Validation of an Automatic Computerized Neurocognitive Battery in Chinese

Ji Zhang, MM[1,2], Ze-yu Hong, BA[1], Liu Yang, MM[1], Xiao-Jia Li, MM[1], and Fang Ye, MD[1]

Abstract

Objectives: Neuropsychological test batteries, which accurately and comprehensively assess cognitive functions, are a crucial approach in the early detection of and interventions for cognitive impairments. However, these tests have yet to gain wide clinical application in China owing to their complexity and time-consuming nature. This study aimed to develop the Computerized Neurocognitive Battery for Chinese-Speaking participants (CNBC), an autorun and autoscoring cognitive assessment tool to provide efficient and accurate cognitive evaluations for Chinese-Speaking individuals.

Methods: The CNBC was developed through collaboration between clinical neurologists and software engineers. Qualified volunteers were recruited to complete CNBC and traditional neurocognitive batteries. The reliability and validity of the CNBC were evaluated by analyzing the correlations between the measurements obtained from the computerized and the paper-based assessment and those between software-based scoring and manual scoring.

Results: The CNBC included 4 subtests and an autorun version. Eighty-six volunteers aged 51-82 years with 7-22 years of education were included. Significant correlations (0.256-0.666) were observed between paired measures associated with attention, executive function, and episodic memory from the CNBC and the traditional paper-based neurocognitive batteries. This suggests a strong construct validity of the CNBC in assessing these cognitive domains. Furthermore, the correlation coefficients between manual scoring and system scoring ranged from 0.904-1.0, indicating excellent inter-rater reliability for the CNBC.

Interpretation: A novel CNBC equipped with automated testing and scoring features was developed in this study. The preliminary results confirm its strong reliability and validity, indicating its promising potential for clinical utilization.

Keywords

validation study, computerized neurocognitive battery, Chinese-speaking, automated testing and scoring, episodic memory, attention, executive function, visual-spatial abilities

Introduction

Cognitive impairment refers to deficits in neurocognitive domains—including learning and memory, attention, language, perceptual-motor development, execution function, and social cognition—that may lead to the decline of an individual's daily or social ability.[1] The underlying etiologies may be neurodegenerative diseases (such as Alzheimer's disease and frontotemporal lobe degeneration) or other causes (such as cerebrovascular disease, trauma, poisoning, and

[1]Department of Neurology, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China
[2]Department of Neurology, The Affiliated Chengdu 363 Hospital of Southwest Medical University, Chengdu, China

Corresponding Author:
Fang Ye, MD, Department of Neurology, Sichuan Provincial Academy of Medical Sciences and Sichuan Provincial People's Hospital, 32 West Second Section of First Ring Road, Chengdu 610072, China.
Email: yfxh00@163.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

emotional disorders). Regarding severity, cognitive impairment can be classified into mild cognitive impairment (MCI) and severe cognitive impairment (dementia). Subclinical cognitive changes not evident on formal testing—termed subjective cognitive decline (SCD)—have also been found to be important in predicting future cognitive impairment. It has been widely recognized that the interventions for cognitive impairment should be advanced to at least the MCI stage—or even the SCD stage—to preserve individuals' cognitive function as long as possible and reduce the burden on families and society.[2] Early intervention should be based on the accurate evaluation of cognitive functions. It usually requires the usage of neuropsychological test batteries, which combine multiple tests to assess the functions of several neurocognitive domains. Traditionally, these tests are administered by professional neuropsychologists—who instruct the participants face-to-face on completing various tasks and make assessments based on their performances—and take 1 to 2 hours for each participant. Such neuropsychological resources and services are widely available and regarded as standard procedures for clinical practice and relevant research in Western countries.

Fully developed computerized neuropsychological batteries complementary to traditional batteries have also been developed for English-speaking participants, such as the Penn Computerized Neurocognitive Battery (Penn CNB) from the University of Pennsylvania,[3] the Cambridge Neuropsychological Testing Automated Battery (CANTAB) from the University of Cambridge,[4] and the Cogstate computerized cognitive assessment battery developed by the Cogstate Cognitive Science Company.[5] The Penn CNB focuses on the following 4 cognitive domains: executive control, episodic memory, complex cognition, and social cognition and is frequently utilized in functional neuroimaging research and large-scale clinical and genomic studies. In order to enhance its sensitivity in detecting cognitive impairments, Penn CNB encompasses a comprehensive range of test content and presents high levels of difficulty, which poses challenges for Chinese middle-aged and older adults with relatively lower levels of education. The CANTAB serves as a comprehensive tool for evaluating multiple cognitive subdomains and is accompanied by specialized hardware, allowing for standardized testing. However, although widely utilized in scientific research, clinical trials, and health care settings in Europe and the United States, CANTAB requires a fee for usage and currently lacks a Chinese version and normative data for the Chinese population. The Cogstate battery, which comprises 8 test tasks, also involves a fee. The initial validation studies of the Chinese version of the Cogstate battery primarily targeted patients with schizophrenia.[6] However, the memory assessment component of the battery lacks a delayed recall task, limiting its ability to assess impairment in episodic memory, since a decline in delayed recall of episodic memory is considered an early and specific neuropsychological change in Alzheimer's disease.[7] Moreover, the subtest has not been automated and requires the assistance of test administrators to facilitate participant completion. Hence, despite their comprehensive assessment of various cognitive domains and wide utilization internationally, the 3 computerized cognitive assessment batteries are not suitable for evaluating and diagnosing cognitive impairments in Chinese middle-aged and older adults.

As for the cognitive assessment of Chinese-speaking participants, although there are applicable cognitive assessment batteries standardized to local populations, such neuropsychological resources are only available in a limited number of tertiary hospitals owing to the unbalanced economic levels in different areas of China, resulting in the lack of financial support for memory clinics and limited health personnel trained to administer cognitive assessments in less developed areas of China. Since more than 50 million people in China have been estimated to have a cognitive impairment, among which nearly 40 million have MCI,[8] there are substantial gaps in neuropsychological assessment resources to meet actual needs.

To address this problem, the present study aimed to develop and initially validate a computerized cognitive assessment battery with autorun and autoscoring systems for Chinese-speaking participants. To our knowledge, no similar study has previously been published.

Methods

Participants and Procedures

Chinese-speaking participants aged 50 years or older with reasonable vision, hearing, and ability to communicate in verbal or written Chinese were recruited from the Memory Clinic and the Inpatient Department of Neurology, Sichuan Provincial People's Hospital, Chengdu, China and nearby communities. Patients with a history of major depression, schizophrenia, epilepsy, significant head injury, substance abuse, alcoholism, or other severe physical disorders that would hinder the completion of the study procedures were excluded. Informed consent was obtained from all participants, and the study protocol was approved by the Institutional Ethics Committee.

The demographic information and past history of the participants were collected during the first interview. Then, neuropsychological test batteries were administered by certified psychometrists. All the tests were translated and adapted from Western countries harmonized with Chinese culture and validated in the Chinese population. Each participant received the following battery of neuropsychological tests: Auditory Verbal Learning Test (AVLT)[9,10]; shape-Trail Making Test (STT)[11]; Rey-Osterrieth Complex Figure Test (ROCF)[12,13]; Digit Span Test (DST)[14] and Boston Naming Test (BNT)[15] and Animal Fluency Test (AFT).[16] Raw scores from these neuropsychological tests were extracted to evaluate the functions of the following 4 cognitive domains: memory,

attention, visuospatial function, language and executive function. The *Z*-score (Z = (raw score - mean)/standard deviation [SD]) of each rated item was also calculated to evaluate the cognitive state of the participants. Our study employed the Jak/Bondi diagnostic criteria for MCI[17] to assess the cognitive function of the participants, using 6 indicators from 4 basic tests or the score of Functional Activity Questionnaire (FAQ) for diagnosis. The specific criteria are as follows: Impairment in 2 tests within the same cognitive domain (>1SD); or 1 test score impaired (>1SD) in 3 cognitive domains (memory, executive, language); or a FAQ scale score greater than 9.

Afterward, the participants completed the CNBC tasks under the auditory or written guidance of the software system; the raw scores of each task were recorded automatically and could be exported when necessary. Meanwhile, 1 researcher rated the performance of each participant manually. To minimize distractions, CNBC was administrated in a separate room using a standard Windows computer with Java. The screen resolution of the test computers should be set to 1920 × 1080 pixels. The participants were asked to turn off their phones and wear headsets to listen to the instructions for CNBC. The operation of the CNBC test interface is only done by clicking with the mouse (including press numbers, select pictures, or drag graphics). Before the formal test, a brief video and pre-test practice were provided to familiarize the participants with the software and mouse operations. Thus, the formal test required minimal manual guidance.

Description of the CNBC

The computerized battery comprised 4 tests targeting attention, processing speed, working memory, episodic memory, visuospatial function, and executive function, including computerized DST (cDST), computerized STT (cSTT), computerized Complex Figure Puzzle Test (cCFPT), computerized Picture Memory and Sorting Test (cCOMS) and an autorun version of CNBC. These tests were developed in Chengdu, China under the joint efforts of clinical neurologists (Sichuan Provincial People's Hospital) and software engineers (YiWei Medical Technology Co.Ltd., Shenzhen, China), which could be accessed for free through the website https://scale.drbraintele.com after applying and obtaining authorization. The instructions for CNBC were simultaneously displayed in text on the screen and presented as voice recordings in Chinese Mandarin. Animation demonstration and practice items were presented before the test items to ensure that the examinee fully understood the instructions. The operation of CNBC can be completed by clicking the mouse (including pressing numbers, selecting pictures or dragging graphics), without input through the touchpad and keyboard. Test items were designed to be completed and scored without the need for qualitative ratings or verbal analysis. The design and development of CNBC draw on the principles and formats of traditional tests to assess various cognitive functions. However, due to the inherent differences between electronic devices and traditional media, there are significant differences in the specific operations, answer methods, and scoring methods. In addition, the traditional test scales used in this study are all open-source and have been localized for Chinese-speaking participants. An automated scoring system was also established, enabling the system to automatically calculate the raw scores and *Z* scores of each test indicator and generate a cognition assessment report. CNBC's norm calculation is based on the test results of a reference group with normal cognitive function. However, due to the high requirements of rolling basis on software development, such as server and data storage capabilities, CNBC's norms are uploaded after calculation. We plan to collect more test data and update the norms regularly. Upon the completion of software development, the system underwent rigorous pretesting by the development team with extensive feedback obtained from experts in neuropsychology, neurologists, and the software development team. Through this iterative process, numerous issues were identified and resolved, leading to continuous improvements in interface design, scale content, voice prompts, and playback speed.

Some of the iterative processes of CNBC are as follows:

1. Optimization in voice broadcasting: We initially used real human recordings as the guide for participants to answer questions, but the recording process was time-consuming and expensive. If the text of the announcement needed to be modified, it required re-recording. Therefore, we changed the broadcasting method to use iFlytek's voice-over package to announce text. This package allows you to choose the gender, speech rate, tone, and pitch of the announcer. During the iteration process, based on user experience, we ultimately chose a female voice with clear enunciation, speaking Mandarin; an appropriate speech rate (about 200 words per minute) and moderate tone; a smooth and natural delivery with appropriate pauses, which helps participants better understand the questions.
2. Optimizations of scale content: In the cDST test, after the participant listens to the digital voice broadcast, they only have 1 chance to click on the number. If they click correctly, they move on to the next question. This operation avoids subjects from changing their answers repeatedly and extending the response time. In the cCFPT pre-test, when the scoring criteria were set to within 30 pixels of the component center from the original coordinates, there were sometimes inconsistencies between software scoring and manual scoring. Through iterative optimization, we adjusted the scoring criteria to within 15 pixels, which brought the software scoring closest to manual scoring. On the report page of the cCFPT test, the original image and the completed image by the subject need to be displayed to facilitate manual review of the accuracy of the computer scoring.

3. Optimization in interface design: Considering that the participants included in this study are middle-aged and elderly individuals aged 50 and above, many of them find it difficult to read due to vision problems, which leads to a longer time spent answering questions. In the pre-test phase, the font size displayed on our answer interface was small, and the mouse cursor was not clearly visible on the screen, which slowed down the participants' response speed. Therefore, we appropriately increased the font size and highlighted the mouse position with a bright red color, which further helped participants to answer questions smoothly.

This optimization process was conducted until a relatively refined and applicable level of usability was reached. The final version of CNBC was briefly described as follows:

1) Computerized Digital Span Test (cDST): cDST was an adaptation of the Chinese version of DST, which measures the function of attention and working memory. The participants were auditorily exposed to a sequence of digits. Immediately afterward, they were asked to click the correct digits shown on the screen in the same order. The number of digits increased from 3 to 12 in the forward part. In the backward part, the participants were asked to click the digits in the opposite order and the number of digits increased from 2 to 10.
2) Computerized Shape Trail Test (cSTT): STT is a culture-fair variant of the Trail Making Test and is thus more suitable for Chinese-speaking participants.[11] The cSTT was adapted from the Chinese version of STT to provide a computerized measure of the ability of "set-shifting". In Part A of cSTT, participants were instructed to make connections by clicking the digit buttons randomly arranged and shown on the screen in the proper order from 1-25. In Part B of cSTT, 2 sets of numbers (from 1-25) were randomly arranged and displayed on the screen with each number of 1 set enclosed by a circle and the other set by a square. The participants were asked to make lines by clicking the digit buttons in alternating order between the 2 sets of numbers. The time in seconds for completing Part A and Part B were automatically recorded as raw scores. Auditory and visual reminders were presented for incorrect connections, the numbers of which were also recorded.
3) Computerized Complex Figure Puzzle Test (cCFPT): cCFPT was developed based on an equivalent Complex Figure Test to ROCF to provide a computerized tool to assess visuo-constructional ability and visual memory.[18] However, it was not feasible to score hand drawings by image recognition software, requiring comparative accuracy as manual scoring. Thus, the complex figure was deconstructed into 17 components, each matched with an interference component with a similar appearance. The examinees were first required to reproduce a figure identical to the 1 displayed on the screen by selecting the correct components and moving them to appropriate positions in the puzzle area (copy trial). Then, they were asked to repeat the above procedures twice without displaying the original figure 5 minutes later (short-term delayed recall, SDR) and 30 minutes later (long-term delayed recall, LDR). For each step, the maximum score was 34 points (17 points for selecting the correct components and 17 points for moving them to the correct positions).
4) Computerized Common Object Memory and Sorting Test (cCOMS): cCOMS was developed as a counterpart of the AVLT in CNBC. However, it was not feasible to collect examinees' answers verbally in cCOMS since existing speech recognition technologies failed to meet the CNBC requirement of scoring accuracy and automation due to the confounding of accents and dialects. Thus, Format-unified color photographs of everyday objects—including a comb, umbrella, tree, knife, cup, banana, chair, scissors, eyeglasses, and clock—were used as stimuli for cCOMS. Across 3 learning trials, each photo was shown for 2 seconds in the center of the screen 1 after another in a fixed order. The examinees were asked to remember both the objects and the sequence in which they were shown. After each round of display, the examinees were asked to identify the 10 objects among 20 photographs (10 original objects and 10 distracters) and drag them to the answer boxes successively in the order that they had been displayed. Five minutes (short-term delayed recall, SDR) and 30 minutes later (long-term delayed recall, LDR), the participants were asked to repeat the identification and sequencing tasks using another set of distracters. In the recognition trial, 20 objects were displayed 1 at a time in random order and the participants were asked to select "Yes" or "No" to determine whether the object had been shown in the learning trial. For each learning and recall trial, the full score was 20 points (10 points each for correct identification and correct ordering). For the recognition trial, 1 point was given when it was correctly determined whether an object was displayed in the learning trial. The maximum score was 20 points.
5) Autorun version of CNBC: To make the test battery more time-saving and efficient, an autorun version of CNBC was developed by organizing the test procedures as follows: cCFPT copy trial, cCOMS learning trials, cCFPT SDR, cCOMS SDR, cSTT, cDST, cCFPT LDR, and cCOMS LDR. The time limit for each procedure was determined after repeated pretests to make it performable for most examinees. The completion of the autorun CNBC was estimated to take

40-55 minutes, which is much shorter than the time to perform the 4 tests 1 after another.

Statistical Analyses and Sample Size Calculation

All statistical analyses were performed using the Statistical Package for Social Sciences (SPSS) version 22.0 (IBM Corporation, Armonk, NY), and a P-value of <0.05 was considered statistically significant. Continuous variables were presented as mean ± SD or median and interquartile range where appropriate and the Mann–Whitney U test or the t test was used to examine the differences between the 2 groups. Categorical variables were expressed as proportions, and a chi-square test was applied in the comparison between groups. Spearman's rank correlation or Pearson's correlation was adopted to analyze the strength of the correlations between CNBC subtests and traditional neuropsychological tests when investigating the validity and reliability of CNBC. Raw scores were used for both computerized and traditional neuropsychological measures. Correlation coefficients (r) were interpreted based on effect size using the convention proposed by Cohen for use in behavioral sciences[19]: small effect = .10, medium effect = .30, and large effect = .50. The accuracy requirements of the present study were set to be a type II error rate (false negatives) of ≤0.2 and effect size of 0.3 and a type I error rate (false positives) of ≤0.05. Thus, the sample size calculated via GPower 3.1.9.7 was ≥82 using the statistical test of correlation: point biserial model in the test family of t-tests (two-tailed).[20]

Results

Demographics and Performance Characteristics of CNBC

From October 2020 to December 2022, 86 volunteers (30 male and 56 female) were recruited for the study. Their ages ranged from 51-82 years, whereas the years of education ranged from 7-22 years. As per the results of traditional cognitive assessment batteries, 70 participants (24 male and 46 female) were classified as having normal cognition (NC), whereas 16 participants (6 male and 10 female) were diagnosed as having MCI (Table 1). No significant difference was observed in gender, age, and years of education between the NC and MCI groups ($P > 0.05$). Table 1 also shows the average test results of traditional cognitive batteries and CNBC. Significant differences were identified between the NC and MCI groups in the test performances of ROCF-copy, ROCF-recall, STT-B, AVLT-$N_{1\text{-}5,7}$, cCFPT-copy, cCFPT-SDR, cCFPT-LDR, cSTT-A, cSTT-B, cCOMS-$C_{1\text{-}5}$.

As shown in Table 2, cCOMS-C_1 ($r = 0.275$, $P = 0.023$) and cDST-backward ($r = 0.390$, $P = 0.001$) were positively correlated with the years of education. cCOMS -C_2 ($r = -0.371$, $P = 0.002$), cCOMS -C_T ($r = -0.287$, $P = 0.018$), cCOMS -C_5 ($r = -0.257$, $P = 0.035$), cCFPT-SDR ($r = -0.256$, $P = 0.037$), and cCFPT-LDR ($r = -0.271$, $P = 0.026$) were negatively correlated with the participants' age. Moreover, cSTT-A ($r = 0.385$, $P = 0.001$) and cSTT-B ($r = 0.366$, $P = 0.002$) were positively correlated with the participants' age. Correlations between cognitive indicators of traditional batteries and demographic parameters (age and education years) were also shown in Table 2.

Construct Validity and Inter-rater Reliability of CNBC

The construct validity of CNBC was investigated by analyzing the correlations between the main test results of the computerized measures and those of the corresponding standard examiner-administered tests. As shown in Table 3, the test results of cDST-forward ($r = 0.652$, $P < 0.001$), cDST-backward ($r = 0.626$, $P < 0.001$), cCFPT-copy ($r = 0.578$, $P < 0.001$), cCFPT-long recall ($r = 0.516$, $P < 0.001$), cSTT-A ($r = 0.620$, $P < 0.001$), and cSTT-B ($r = 0.666$, $P < 0.001$) were strongly correlated with their counterpart indicators in traditional test batteries. Furthermore, cCOMS-C_2 ($r = 0.449$, $P < 0.001$), C_3 ($r = 0.411$, $P < 0.001$), C_T ($r = 0.486$, $P < 0.001$), C_4 ($r = 0.364$, $P < 0.001$), and C_5 ($r = 0.402$, $P < 0.001$) showed medium correlations with AVLT-N_2, N_3, N_T, N_4 and N_5, respectively.

To investigate whether education years impact the correlations between the corresponding indicators of AVLT and cCOMS, they were analyzed among participants with 7-12 years of education and those with more than 12 years of education, respectively. As shown in Table 4, among participants with 7-12 years of education, cCOMS-C_2 ($r = 0.438$, $P = 0.002$), C_3 ($r = 0.411$, $P = 0.003$), C_T ($r = 0384$, $P = 0.004$), and C_5 ($r = 0.384$, $P = 0.006$) showed medium correlations with AVLT- N_2, N_3, N_T and N_5. Among participants with more than 12 years of education, cCOMS-C_2 ($r = 0.408$, $P = 0.018$), C_3 ($r = 0.309$, $P = 0.025$), C_T ($r = 0.547$, $P = 0.001$), C_4 ($r = 0.499$, $P = 0.003$), and C_5 ($r = 0.493$, $P = 0.004$) showed medium to strong correlations with AVLT-N_2, N_3, N_T, N_4 and N_5, respectively.

Inter-rater reliability was assessed by comparing the scoring results of the human rater and the autoscoring system of CNBC. As shown in Table 5, the CNBC performances of 70-72 participants were rated simultaneously in both ways and the scores of each pair of indicators were highly consistent with r values ranging from 0.9-1.0 (P values <0.001). The findings indicated that the system scoring of CNBC exhibited a high level of accuracy, serving as a reliable substitute for manual scoring.

Discussion

This study developed a CNBC comprising the following 4 distinct subtests: cDST, cSTT, cCFPT, and cCOMS. These subtests were primarily designed to assess attention, executive function, episodic memory, and visuospatial abilities. The development of the CNBC was a scientific and meticulous

Table 1. Baseline Characteristics of the Participants.

	All Participants	NC	MCI	*P* (NC vs MCI)
n	86	70	16	
Gender (male ratio,%)	30 (34.88%)	24 (34.29%)	6 (37.50%)	0.808
Years of education	12.44 ± 2.92	12.39 ± 3.06	12.63 ± 2.31	0.776
Age	65.27 ± 8.31	66.16 ± 8.19	61.38 ± 7.92	0.037
DST-forward	7.53 ± 1.20	7.60 ± 1.19	7.25 ± 1.18	0.293
DST-backward	5.23 ± 1.13	5.30 ± 1.17	4.94 ± 0.93	0.251
ROCF-copy	32.69 ± 3.26	33.30 ± 1.81	30.00 ± 5.97	0.044*
ROCF-recall	15.90 ± 6.01	17.25 ± 5.32	10.00 ± 5.42	0.000**
STT-A	57.57 ± 19.78	55.57 ± 16.98	66.35 ± 28.10	0.158
STT-B	151.71 ± 52.06	140.08 ± 38.15	201.89 ± 72.63	0.004**
AVLT-N_1	4.52 ± 1.61	4.73 ± 1.58	3.53 ± 1.41	0.008**
AVLT-N_2	6.76 ± 1.81	7.06 ± 1.73	5.40 ± 1.60	0.001**
AVLT-N_3	8.41 ± 1.98	8.71 ± 1.81	7.00 ± 2.17	0.002**
AVLT-N_T	19.69 ± 4.72	20.50 ± 4.41	15.93 ± 4.43	0.000**
AVLT-N_4	7.70 ± 2.57	8.16 ± 2.31	5.43 ± 2.62	0.000**
AVLT-N_5	7.12 ± 2.51	7.47 ± 2.35	5.43 ± 2.65	0.005**
AVLT-N_7	21.67 ± 2.73	22.35 ± 1.30	18.36 ± 4.83	0.000**
cDST-forward	7.35 ± 1.12	7.41 ± 1.12	7.00 ± 1.08	0.223
cDST-backward	5.44 ± 1.45	5.57 ± 1.46	4.79 ± 1.25	0.064
cCFPT-copy	28.98 ± 5.96	30.19 ± 4.37	23.53 ± 8.74	0.011*
cCFPT-SDR	19.67 ± 6.82	21.09 ± 6.02	13.33 ± 6.76	0.000**
cCFPT-LDR	20.37 ± 6.84	21.85 ± 5.97	13.73 ± 6.67	0.000**
cSTT-A	76.37 ± 31.76	70.90 ± 26.40	99.99 ± 41.91	0.001**
cSTT-B	149.74 ± 54.05	142.08 ± 49.78	182.76 ± 60.78	0.006**
cCOMS -C_1	15.57 ± 3.12	16.01 ± 2.70	13.53 ± 4.10	0.005**
cCOMS -C_2	17.84 ± 2.48	18.44 ± 1.79	15.13 ± 3.34	0.002**
cCOMS -C_3	19.01 ± 2.02	19.50 ± 1.11	16.80 ± 3.43	0.009**
cCOMS -C_T	52.33 ± 6.77	53.79 ± 4.84	45.67 ± 9.91	0.007**
cCOMS -C_4	18.58 ± 2.32	19.04 ± 1.67	16.47 ± 3.50	0.014*
cCOMS -C_5	18.55 ± 2.59	19.04 ± 1.61	16.33 ± 4.52	0.037*
cCOMS -C_6	19.69 ± 0.83	19.74 ± 0.70	19.47 ± 1.25	0.431

NC: Normal Control; MCI: Mild Cognitive Impairment; DST: Digit Span Test, including DST-forward and DST-backward; ROCF: Rey-Osterrieth Complex Figure, including ROCF-copy and ROCF-recall; STT: Shape Trails Test, including STT-A and STT-B; AVLT: Auditory Verbal Learning Test, including AVLT-N_{1-3}(the number of correct words recalled on trial 1, 2, and 3, respectively), AVLT-N_T(the sum of correct words recalled on trial 1, 2, and 3), AVLT-N_4(the number of correct words on trial 4), AVLT-N_5 (the number of correct words on trial 5) and AVLT-N_7(the number of words recognized correctly in the recognition trial); cDST: computerized Digit Span Test; cCFPT: computerized Complex Figure Puzzle test, including cCFPT-copy, cCFPT-SDR (short-term delayed recall) and LDR (long-term delayed recall); cSTT: computerized Shape Trails Test; cCOMS: computerized Picture Memory and Sorting Test, including cCOMS-C_{1-3} (the number of objects along with their orders correctly recalled on trial 1, 2, and 3, respectively), cCOMS-C_T(the sum of objects along with their orders correctly recalled on trial 1, 2, and 3), cCOMS-C_4(the number of objects along with their orders correctly recalled on trial 4), cCOMS-C_5(the number of objects along with their orders correctly recalled on trial 5) and cCOMS-C_6(the number of objects correctly identified as old or new in the recognition trial).
*$P < 0.05$, **$P < 0.01$.

process that integrated an automated testing and scoring system that reduced manual assistance as far as possible. The automatic administration of the battery typically took approximately 50 minutes per session. Its standardized procedures for voice prompts and question display ensured consistency throughout the assessment, optimizing efficiency while reducing potential scoring errors caused by tester intervention, operational variability, and inter-rater discrepancies. This approach saved valuable resources and enhanced the reliability and accuracy of the testing and scoring process.

In terms of the validation of the CNBC, it exhibits excellent content validity since the system was designed by adopting or referencing the testing format of well recognized paper-based cognitive assessment batteries with established norms in the field of cognitive evaluation in China,[12,21] combining with computer software development techniques. To investigate its construct validity and inter-rater reliability (Table 3; Table 5), the test performances of 86 participants were analysed, and the results revealed significant correlations (0.364-0.666) between key indicators assessing attention, executive function, and episodic memory. This indicates strong construct validity Furthermore, the high correlation coefficients (0.904-1.0) between manual scoring and system scoring demonstrate excellent inter-rater reliability.

Table 2. Correlations Between Cognitive Indicators of Participants With Normal Cognition and Demographic Parameters (Age and Education Years).

CNBC		cCOMS-C_1	cCOMS-C_2	cCOMS-C_3	cCOMS-C_T	cCOMS-C_4	cCOMS-C_5	cCOMS-C_6	cSTT-A	cSTT-B	cDST-F	cDST-B	cCFPT-C	cCFPT-SDR	cCFPT-LDR
Years of education	r	0.275*	0.034	0.212	0.224	0.111	0.102	−0.29	−0.178	−0.198	0.216	0.390**	0.140	0.093	0.052
	P	0.023	0.78	0.083	0.067	0.367	0.408	0.813	0.143	0.103	0.073	0.001	0.257	0.456	0.676
Age	r	−0.233	−0.371**	−0.045	−0.287*	−0.231	−0.257*	−0.014	0.385**	0.366**	−0.149	−0.182	−0.123	−0.256*	−0.271*
	P	0.056	0.002	0.719	0.018	0.059	0.035	0.909	0.001	0.002	0.218	0.131	0.321	0.037	0.026

Traditional Baterries		AVLT-N_1	AVLT-N_2	AVLT-N_3	AVLT-N_T	AVLT-N_4	AVLT-N_5	AVLT-N_7	STT-A	STT-B	DST-F	DST-B	ROCF-C	ROCF-R
Years of education	r	0.297*	0.304	0.21	0.312**	0.091	0.076	0.268*	−0.084	−0.298*	0.190	0.264*	0.132	0.076
	P	0.013	0.100	0.081	0.009	0.461	0.537	0.027	0.487	0.013	0.115	0.027	0.275	0.531
Age	r	−0.212	−0.33	−0.459**	−0.393**	−0.423	−0.525**	−0.167	0.595**	0.418**	−0.276**	−0.096	−0.198	−0.042
	P	0.78	0.05	0.000	0.001	0.000	0.000	0.172	0.000	0.000	0.021	0.431	0.100	0.729

Computerized Neurocognitive Battery for Chinese-Speaking participants (CNBC);DST: Digit Span Test, including DST-f (DST-forward) and DST-b (DST-backward); ROCF: Rey-Osterrieth Complex Figure, including ROCF-c (ROCF-copy) and ROCF-r (ROCF-recall); STT: Shape Trails Test, including STT-A and STT-B; AVLT: Auditory Verbal Learning Test, including AVLT-$N_{1\text{-}3}$(the number of correct words recalled on trial 1, 2, and 3, respectively), AVLT-N_T(the sum of correct words recalled on trial 1, 2, and 3), AVLT-N_4(the number of correct words on trial 4), AVLT-N_5 (the number of correct words on trial 5) and AVLT-N_7(the number of words recognized correctly in the recognition trial); cDST: computerized Digit Span Test, including cDST-f (DST-forward) and cDST-b (DST-backward); cCFPT: computerized Complex Figure Puzzle test, including cCFPT-c (cCFPT-copy), cCFPT-SDR (short-term delayed recall) and LDR (long-term delayed recall); cSTT: computerized Shape Trails Test; cCOMS: computerized Picture Memory and Sorting Test, including cCOMS-$C_{1\text{-}3}$ (the number of objects along with their orders correctly recalled on trial 1, 2, and 3, respectively), cCOMS-C_T(the sum of objects along with their orders correctly recalled on trial 1, 2, and 3), cCOMS-C_4(the number of objects along with their orders correctly recalled on trial 4), cCOMS-C_5(the number of objects along with their orders correctly recalled on trial 5) and cCOMS-C_6(the number of objects correctly identified as old or new in the recognition trial).
*$P < 0.05$, **$P < 0.01$.

Table 3. Correlations Between Corresponding Indicators of CNBC and Traditional Batteries.

Traditional Indicators	CNBC Indicators	Number of Subjects (n)	Correlation Ratio (r)	*P*
DST-forward	cDST-forward	83	0.652**	0.000
DST-backward	cDST-backward	84	0.626**	0.000
ROCF-copy	cCFPT-copy	82	0.578**	0.000
ROCF-recall	cCFPT-LDR	82	0.516**	0.000
STT-A	cSTT-A	84	0.620**	0.000
STT-B	cSTT-B	84	0.666**	0.000
AVLT-N_1	cCOMS- C_1	82	0.256*	0.020
AVLT-N_2	cCOMS- C_2	82	0.449**	0.000
AVLT-N_3	cCOMS- C_3	82	0.411**	0.000
AVLT-N_T	cCOMS- C_T	82	0.486**	0.000
AVLT-N_4	cCOMS- C_4	82	0.364**	0.000
AVLT-N_5	cCOMS- C_5	82	0.402**	0.000
AVLT-N_7	cCOMS- C_6	82	0.108	0.333

Computerized Neurocognitive Battery for Chinese-Speaking participants (CNBC); DST: Digit Span Test, including DST-forward and DST-backward; ROCF: Rey-Osterrieth Complex Figure, including ROCF-copy and ROCF-recall; STT: Shape Trails Test, including STT-A and STT-B; AVLT: Auditory Verbal Learning Test, including AVLT-$N_{1\text{-}3}$(the number of correct words recalled on trial 1, 2, and 3, respectively), AVLT-N_T(the sum of correct words recalled on trial 1, 2, and 3), AVLT-N_4(the number of correct words on trial 4), AVLT-N_5 (the number of correct words on trial 5) and AVLT-N_7(the number of words recognized correctly in the recognition trial); cDST: computerized Digit Span Test; cCFPT: computerized Complex Figure Puzzle test; cCFPT- LDR (long-term delayed recall); cSTT: computerized Shape Trails Test; cCOMS: computerized Picture Memory and Sorting Test, including cCOMS-$C_{1\text{-}3}$ (the number of objects along with their orders correctly recalled on trial 1, 2, and 3, respectively), cCOMS-C_T(the sum of objects along with their orders correctly recalled on trial 1, 2, and 3), cCOMS-C_4(the number of objects along with their orders correctly recalled on trial 4), cCOMS-C_5(the number of objects along with their orders correctly recalled on trial 5) and cCOMS-C_6(the number of objects correctly identified as old or new in the recognition trial).
*$P < 0.05$, **$P < 0.01$.

Table 4. Correlations Between AVLT and cCOMS Indicators in NC Participants With 7-12 years of Education and Those With with More than12 years of Education, Respectively.

AVLT	cCOMS	Participants with 7-12 years of Education	Correlation Ratios (r)	Participants with >12 years of Education	Correlation Ratios (r)
AVLT-N_1	cCOMS-C_1	49	0.120	33	0.263
AVLT-N_2	cCOMS-C_2	49	0.438**	33	0.408*
AVLT-N_3	cCOMS-C_3	49	0.411**	33	0.309*
AVLT-N_T	cCOMS-C_T	49	0.384**	33	0.547**
AVLT-N_4	cCOMS-C_4	49	0.247	33	0.499**
AVLT-N_5	cCOMS-C_5	49	0.384**	33	0.493**
AVLT-N_7	cCOMS-C_6	49	0.287	33	−1.02

NC: Normal Cognitive; AVLT: Auditory Verbal Learning Test, including AVLT-$N_{1\text{-}3}$(the number of correct words recalled on trial 1, 2, and 3, respectively), AVLT-N_T(the sum of correct words recalled on trial 1, 2, and 3), AVLT-N_4(the number of correct words on trial 4), AVLT-N_5 (the number of correct words on trial 5) and AVLT-N_7(the number of words recognized correctly in the recognition trial); cCOMS: computerized Picture Memory and Sorting Test, including cCOMS-$C_{1\text{-}3}$ (the number of objects along with their orders correctly recalled on trial 1, 2, and 3, respectively), cCOMS-C_T(the sum of objects along with their orders correctly recalled on trial 1, 2, and 3), cCOMS-C_4(the number of objects along with their orders correctly recalled on trial 4), cCOMS-C_5(the number of objects along with their orders correctly recalled on trial 5) and cCOMS-C_6(the number of objects correctly identified as old or new in the recognition trial).
*$P < 0.05$, **$P < 0.01$.

Altogether, these findings affirm the promising clinical potential of the CNBC.

Some previous studies have compared the performances of computerized and paper-based cognitive tests. Gur et al examined the correlation between the 2 in a sample of 92 healthy individuals. The findings indicated significant correlation coefficients of 0.52 ($P = 0.0001$) and 0.53 ($P = 0.0001$) for measures assessing executive function and

Table 5. Correlations Between Scoring Results of the Human Rater and the Autoscoring System of CNBC

CNBC Autoscoring	Human Rater Scoring	Number of Subjects (n)	Correlation Ratio (r)	*P*
cDST-forward	cDST-forward	72	1	
cDST-backward	cDST-backward	72	1	
cCOMS- C_1	cCOMS- C_1	71	1	
cCOMS- C_2	cCOMS- C_2	71	1	
cCOMS- C_3	cCOMS- C_3	71	1	
cCOMS- C_T	cCOMS- C_T	71	1	
cCOMS- C_4	cCOMS- C_4	71	1	
cCOMS- C_5	cCOMS- C_5	71	1	
cCOMS- C_6	cCOMS- C_6	71	1	
cCFPT-copy	cCFPT-copy	70	0.904**	0.000
cCFPT-SDR	cCFPT-SDR	70	0.992**	0.000
cCFPT-LDR	cCFPT-LDR	70	0.995**	0.000
cSTT-A	cSTT-A	72	1	
cSTT-B	cSTT-B	70	1	

DST: Digit Span Test, including DST-forward and DST-backward; ROCF: Rey-Osterrieth Complex Figure, including ROCF-copy and ROCF-recall; STT: Shape Trails Test, including STT-A and STT-B; AVLT: Auditory Verbal Learning Test, including AVLT-$N_{1\text{-}3}$(the number of correct words recalled on trial 1, 2, and 3, respectively), AVLT-N_T(the sum of correct words recalled on trial 1, 2, and 3), AVLT-N_4(the number of correct words on trial 4), AVLT-N_5 (the number of correct words on trial 5) and AVLT-N_7(the number of words recognized correctly in the recognition trial); cDST: computerized Digit Span Test; cCFPT: computerized Complex Figure Puzzle test, including cCFPT-copy, cCFPT-SDR (short-term delayed recall) and LDR (long-term delayed recall); cSTT: computerized Shape Trails Test; cCOMS: computerized Picture Memory and Sorting Test, including cCOMS-$C_{1\text{-}3}$ (the number of objects along with their orders correctly recalled on trial 1, 2, and 3, respectively), cCOMS-C_T(the sum of objects along with their orders correctly recalled on trial 1, 2, and 3), cCOMS-C_4(the number of objects along with their orders correctly recalled on trial 4), cCOMS-C_5(the number of objects along with their orders correctly recalled on trial 5) and cCOMS-C_6(the number of objects correctly identified as old or new in the recognition trial).
*$P < 0.05$, **$P < 0.01$.

memory function, respectively.[22] Zhong et al. examined the construct validity of the Chinese version of the Cogstate Battery (CSB-C). The results revealed a correlation coefficient of 0.28 between the CSB-C and paired measures assessing attention in traditional comprehensive assessment batteries.[6] Additionally, for measures evaluating memory function, the correlation coefficients ranged from 0.36-0.62. In the validation study of the Penn CNB memory tests, the researchers analyzed the correlation between the computerized tests and paper-based ones (eg, the California Verbal Learning Test), finding a moderate correlation (0.30-0.49).[23] One recently published study, which validates a pad-version Hopkins Verbal Learning Test-Revised (HVLT-R) in Chinese participants, finds that the correlation coefficients between the Pad-HVLT-R and its paper-and-pencil version are moderately significant (around 0.50).[24] Comparing to these previous studies, the indicators evaluating attention (cDST, cSTT-A) and executive function (cSTT-B) in CNBC exhibit stronger construct validity ($r > 0.60$). This might be attributed to the close adherence of cDST and cSTT to the design of their paper-based versions, resulting in relatively smaller changes in format and content.

Meanwhile, the indicators rating episodic memory (cCFPT-recall and cCOMS-recall, r: 0.40-0.50) in CNBC showed comparable construct validity to those in the previous studies, which was weaker than those evaluating attention and executive function in CNBC. It might be attributed to the following reasons: the delayed recall components of cCFPT and cCOMS cannot be implemented in the form of free recall commonly used in paper-based memory tests due to the inability to accurately recognize speech or hand-drawing input which posed a challenge in achieving automated testing. Alternative input methods (eg, typing) either impose high demands on the participants or require operator assistance, which also hinders the realization of automated testing. Consequently, for cCFPT, the recall task involved selecting and assembling components to recreate the target figure. Although there were distractor components, it still incorporates a form of cue-based recall, leading to the reduction of task difficulty. This alteration in test format, specifically in the nature of the recall tasks, may have impacted the correlation between cCFPT-recall and the RCOF test. Similar issues also applied to the correlation between cCOMS and AVLT recall test. Similar influencing factors have been reported in previous studies.[23,24] For example, the PennCNB study used a calibrated scale for the California Verbal Learning Test, which has a format similar to AVLT, and reported comparable correlations of corresponding indicators to our results.[23] Nevertheless, when designing cCOMS, in order to address the potential difficulty reduction associated with solely relying on image selection, which is akin to recognition tasks, the present study implemented a sequential selection task that requires participants to choose the target image from a set of 20 pictures in a specific order in cCOMS. This design aimed to ensure a comparable level of difficulty to the picture-based AVLT and *COMT*, which do not require sequential recall.[10,12,25]

Still, the correlation coefficients between cCOMS and AVLT recall tests (0.364 and 0.402) indicated a slightly lower construct validity compared to that between cCFPT and ROFT recall tests ($r = 0.516$). This discrepancy may be attributed to the different impacts of education levels on cCOMS and AVLT, resulting in the further reduction of the correlations. As shown in Table 4, compared to participants with lower education levels, participants with higher levels of education exhibited a greater number of significant and stronger correlations between the corresponding measures of cCOMS and AVLT, especially the key indicators (delayed recall). Previous research has shown that the education level substantially influences AVLT scores, resulting in its inapplicability and significant variations among individuals with lower education (eg, a primary school education) or those who are illiterate.[10,12] By contrast, picture-based memory tests are less susceptible to the impact of education level.[12,25] The stronger influences of education level on AVLT might lead to the reduction of the correlations between cCOMS and AVLT in lower educated participants in the present study. These results also indicated that cCOMS may exhibit greater generalizability and broader applicability compared to AVLT.

The present study successfully developed the CNBC, which improved the testing efficiency and reduced the occupation of personnel by implementing an automated testing and scoring system. To assess the accuracy of the software scoring, we evaluated the inter-rater reliability by examining the correlation between the automatically generated scores and raters' manually assigned scores. The study results demonstrated a high correlation coefficient of 1 between the system scores and the manual scores for cDST, cSTT, and cCOMS (Table 5). This indicates a strong agreement between the automated scoring and the manual scoring for these 3 tests. The consistency can be attributed to the identical scoring rules employed in both versions and the objective measures used such as counting the number of correct responses or measuring the time taken to complete the tasks. The correlation coefficients between the system scores and the manual scores for cCFPT range from 0.904-0.995, suggesting a minor variation between the 2 (Table 5). This discrepancy may arise owing to differences in how the system and human raters assess the accuracy of component positioning in cCFPT. To enable automated scoring, the software engineers devised a method wherein the position coordinates for each component of cCFPT were determined based on the original image. During the puzzle-solving process, participants were instructed to accurately place the components in their designated positions. The software then assigned a position score by assessing the deviation between the center of the component and the original coordinates, with a tolerance of 15 pixels. By contrast, manual judgment is unable to achieve the same level of precision. Consequently, minor discrepancies may occur among some participants, but their overall impact remains negligible. Overall, the software-based scoring of CNBC demonstrates high accuracy, making it a reliable substitute for manual scoring.

In addition, the participants in this study were divided into the following 2 groups based on their cognitive function: NC and MCI groups. The 2 groups showed no significant differences in general characteristics such as gender, age, and education level. However, the MCI group exhibited significantly lower scores in multiple cognitive function assessments compared to the normal group. Ten out of the 13 pairs of score indicators from the traditional scales and 11 out of 14 pairs from the CNBC showed significant differences between the 2 groups (Table 1). The results imply that CNBC might distinguish NC individuals from MCI with comparable sensitivity to the traditional scales. However, it is important to validate these findings in larger sample sizes to establish the discriminant validity of CNBC.

The present study has a few limitations. First, the sample size was relatively small; it is necessary to expand the sample size in future studies to further validate the reliability and validity of CNBC. Second, the accurate recognition of input speech is challenging owing to factors such as accents, dialects, and the current limitations of speech recognition technology. Consequently, CNBC has not yet developed tests specifically targeting language function, neither did other computerized neurocognitive batteries. Moreover, due to technique limitations and the design priority of test automation, the delayed recall tests of the CNBC did not implement the form of free recall; other electronic memory scales also have similar issues. However, a sequential selection task was designed in cCOMS to address the potential difficulty reduction as mentioned above.

In conclusion, this study has successfully introduced a novel full-automatic computerized cognitive assessment battery tailored to the middle-aged and older adult Chinese population and conducted an initial validation. The results have demonstrated excellent reliability and validity of the battery, with potentially heightened sensitivity in detecting cognitive impairments, indicating its strong potential for clinical application and wide-scale implementation. It is necessary to enhance the sample size to further corroborate the reliability and validity of the assessment battery and evaluate its diagnostic efficacy specifically in individuals with cognitive impairments.

Acknowledgments

The authors would like to thank Chunhui Luo, Jialing Zhao and other software engineers from YiWei Medical Technology Co.Ltd. For program writing and software architecture.

Author Contributions

F. Y. designed the study, supervised the data collection and wrote the paper. J. Z. collected the data, carrying out the statistical analysis and assisted with writing the article. Z. H., X. L. and L. Y. assisted in the data collecting.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was carried out in Chengdu, China, and funded by grants from the STI 2030-Major Projects 2022ZD0208500 and the Sichuan Provincial Cadre Health Research Projects (2023-205).

ORCID iDs

Ji Zhang https://orcid.org/0000-0001-9649-9608
Fang Ye https://orcid.org/0000-0002-2615-9074

References

1. Morley JE. An overview of cognitive impairment. *Clin Geriatr Med*. 2018;34(4):505-513. doi:10.1016/j.cger.2018.06.003.
2. Jessen F, Amariglio RE, Buckley RF, et al. The characterisation of subjective cognitive decline. *Lancet Neurol*. 2020;19(3):271-278. doi:10.1016/S1474-4422(19)30368-0.
3. Moore TM, Reise SP, Gur RE, Hakonarson H, Gur RC. Psychometric properties of the Penn computerized neurocognitive battery. *Neuropsychology*. 2015;29(2):235-246. doi:10.1037/neu0000093.
4. Kim HS, An YM, Kwon JS, Shin MS. A preliminary validity study of the cambridge neuropsychological test automated battery for the assessment of executive function in schizophrenia and bipolar disorder. *Psychiatry Investig*. 2014;11(4):394-401. doi:10.4306/pi.2014.11.4.394.
5. Mielke MM, Weigand SD, Wiste HJ, et al. Independent comparison of CogState computerized testing and a standard cognitive battery with neuroimaging. *Alzheimers Dement*. 2014;10(6):779-789. doi:10.1016/j.jalz.2014.09.001.
6. Zhong N, Jiang H, Wu J, et al. Reliability and validity of the CogState battery Chinese language version in schizophrenia. *PLoS One*. 2013;8(9):e74258. doi:10.1371/journal.pone.0074258. Published 2013 Sep 2.
7. Benejam B, Aranha MR, Videla L, et al. Neural correlates of episodic memory in adults with Down syndrome and Alzheimer's disease. *Alzheimer's Res Ther*. 2022;14(1):123. doi:10.1186/s13195-022-01064-x. Published 2022 Sep 3.
8. Jia L, Du Y, Chu L, et al. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. *Lancet Public Health*. 2020;5(12):e661-e671. doi:10.1016/S2468-2667(20)30185-7.
9. Ding D, Zhao Q, Guo Q, et al. Prevalence of mild cognitive impairment in an urban community in China: a cross-sectional analysis of the Shanghai Aging Study. *Alzheimers Dement*. 2015;11(3):300-309. doi:10.1016/j.jalz.2013.11.002.
10. Guo QH, Sun YW, Yu PM, et al. Norm of auditory verbal learning test in the normal aged in China community. *Chin J Clin Psychol*. 2007;15:132-134. https://api.semanticscholar.org/CorpusID:148363717
11. Zhao Q, Guo Q, Li F, Zhou Y, Wang B, Hong Z. The Shape Trail Test: application of a new variant of the Trail making test. *PLoS One*. 2013;8(2):e57333. doi:10.1371/journal.pone.0057333.
12. Guo QH. Neuropsychological assessment. In: *Chinese*. 3rd ed. Shanghai: Shanghai Science and Technology Press; 2020.
13. Zhou B, Zhao Q, Kojima S, et al. One-year outcome of Shanghai mild cognitive impairment cohort study. *Curr Alzheimer Res*. 2019;16(2):156-165. doi:10.2174/1567205016666181128151144.
14. Gong Y, Cai T. *Manual of Wechsler Intelligence Scale for Children. Chinese Revision (C-WISC)*. Changsha, Hunan Atlas Press, 1993.
15. Cheung RW, Cheung MC, Chan AS. Confrontation naming in Chinese patients with left, right or bilateral brain damage. *J Int Neuropsychol Soc*. 2004;10(1):46-53. doi:10.1017/S1355617704101069.
16. Mok EH, Lam LC, Chiu HF. Category verbal fluency test performance in Chinese elderly with Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2004;18(2):120-124. doi:10.1159/000079190.
17. Bondi MW, Edmonds EC, Jak AJ, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis*. 2014;42(1):275-289. doi:10.3233/JAD-140276.
18. Yamashita H, Yasugi M. Comparison of the medical college of Georgia complex figures and the rey-osterrieth complex figure tests in a normal sample of Japanese university students. *Percept Mot Skills*. 2008;107(1):45-50. doi:10.2466/pms.107.1.45-50.
19. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Academic Press; 1969.
20. Kang H. Sample size determination and power analysis using the G*Power software. *J Educ Eval Health Prof*. 2021;18:17. doi:10.3352/jeehp.2021.18.17.
21. Wang H, Fan Z, Shi C, et al. Consensus statement on the neurocognitive outcomes for early detection of mild cognitive impairment and Alzheimer dementia from the Chinese Neuropsychological Normative (CN-NORM) Project. *J Glob Health*. 2019;9(2):020320. doi:10.7189/jogh.09.020320.
22. Gur RC, Ragland JD, Moberg PJ, et al. Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*. 2001;25(5):766-776. doi:10.1016/S0893-133X(01)00278-0.
23. Busch RM, Hogue O, Ferguson L, Parsons MW, Kubu CS, Floden DP. Validation of computerized episodic memory measures in a diverse clinical sample referred for neuropsychological assessment. *Clin Neuropsychol*. 2019;33(3):557-570. doi:10.1080/13854046.2018.1488995.
24. Jiang L, Xu M, Xia S, et al. Reliability and validity of the electronic version of the Hopkins verbal learning test-revised in middle-aged and elderly Chinese people. *Front Aging Neurosci*. 2023;15:1124731. doi:10.3389/fnagi.2023.1124731. Published 2023 Jun 12.
25. Kempler D, Teng EL, Taussig M, Dick MB. The common objects memory test (COMT): a simple test with cross-cultural applicability. *J Int Neuropsychol Soc*. 2010;16(3):537-545. doi:10.1017/S1355617710000160.