Psychology Research and Behavior Management

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Reliability and Validity of a Chinese Version of Adaptive Cognitive Evaluation Tool in College Students

Ze Yu[1,2,*], Youyang Wang[1,2,*], Yiyun Li[3], Wei Feng[1,2]

[1]Department of Psychological Medicine, Fudan University Shanghai Cancer Center, Shanghai, 200032, People's Republic of China; [2]Department of Oncology, Shanghai Medical College of Fudan University, Shanghai, 200032, People's Republic of China; [3]Stomatological Hospital and Dental School of Tongji University, Shanghai, 200072, People's Republic of China

*These authors contributed equally to this work

Correspondence: Wei Feng, Department of Psychological Medicine, Fudan University Shanghai Cancer Center, Shanghai, 200032, People's Republic of China, Email ffww06@163.com

Background: Cognition is central to acquiring knowledge and learning new experiences, critical for social behavior and quality of life. Despite its importance, traditional cognitive assessment tools face limitations, including high labor costs and human error, underscoring an urgent need for cost-effective, precise tools to assess cognitive functions.

Objective: This study aims to address this gap by evaluating the reliability and validity of the Chinese version of the Adaptive Cognitive Evaluation (ACE) tool among college students, thereby contributing to the advancement of cognitive research and disease management strategies in China.

Methods: We collected data from 150 participants (72 males, 78 females) with an average age of 20.97 ±3.36 years. A baseline assessment was conducted using the ACE Chinese version, Digit Span Memory Test (DSMT), and Line-trailing Test-A & B (LTT-A&B). After one week, the ACE tests were administered again to assess test–retest reliability.

Results: The results indicated no significant correlations between age, sex, and the outcomes of the sub-tests. However, a significant association was found between educational level and the results of the sub-tests. The Cronbach's α for each sub-test exceeded 0.8, indicating high reliability. Both the I-CVI and S-CVI indexes were 1.00, demonstrating strong content validity. When DSMT, LTT-A, and LTT-B were used as criteria, most sub-tests showed satisfactory criterion validity. The factor-loading coefficient for each dimension of cognitive control was greater than 0.4, and the cumulative variance explanation rate was 64.84%.

Conclusion: The Chinese version of the ACE tool demonstrated satisfactory reliability and validity, making it an efficient tool for cognitive function assessment among college students.

Keywords: cognitive control, adaptive cognitive evaluation, reliability, validity

Introduction

Cognition, a fundamental neuropsychological process, is essential for acquiring knowledge and learning new experiences, which are vital for our social behavior and quality of life.[1–3] It encompasses multiple processing components such as attention, judgment, memory, reasoning, and calculation.[4,5] The cognitive process leverages existing knowledge to derive new insights.[1,4,6] Cognitive control, an advanced brain function, refers to our ability to adjust thoughts and actions according to specific goals or tasks.[7,8]

To perform cognitive control, the brain learns new rules and conflicts,[9] combining an initial impression of the external environment with actual circumstances to develop an adaptive response plan for emergencies and/or challenging environments.[10] Cognitive control mainly includes working memory, goal management, attention, and other cognitive domains.[11] Working memory actively maintains and processes associated information when performing cognitive tasks. It plays a crucial role in acquiring knowledge and learning new skills.[12] Goal management refers to the ability to flexibly

Received: 19 September 2024
Accepted: 25 December 2024
Published: 16 January 2025

adjust related behaviors based on changes in the surrounding environment.[13] Sustained attention is a basic process of attention management in which one's attention is focused on a specific object for a considerable period.[14] Whereas selective attention comes into play when a current goal is set and ignores irrelevant information.[15,16] Divided attention is an individual's ability to handle multiple tasks at the same time.[17] Task Switch refers to a combination of cognitive skills that require to accomplish attention-Switch between tasks and activities.[18]

Despite the significance of cognition, traditional cognitive assessment tools have limitations, including high labor costs for document management, data acquisition, and human error.[19–23] Addressing these gaps is crucial for advancing cognitive research and improving disease management strategies. The limitations of traditional "hard copy" cognitive assessment tools highlight the need for a more efficient and accurate method.[19–23] An electronic version could overcome these challenges, offering large-scale data analysis with precision and reducing subjective factors.[22,23] While traditional tools like MMSE,[19] MoCA,[20] Stroop test,[21] and WCST[18] have been valuable, there is a need to explore the potential of electronic cognitive assessment tools, which are less explored but could offer significant advantages.[19–23]

Previous studies have utilized traditional cognitive assessment tools, but the advent of electronic tools presents a new frontier in cognitive research.[19–23] Our study differs from previous ones by introducing the ACE software, an electronic cognitive assessment tool designed to cover cognitive domains such as working memory, goal management, and attention, which have not been fully explored in the electronic format.[22,23] Concurrently, we have also conducted studies on the reliability and validity of the ACE in populations with depression and schizophrenia.

The primary objective of this study is to evaluate the reliability and validity of the Chinese version of the ACE tool among college students in China. This research contributes to the field by providing insights into the performance of the ACE tool in a Chinese population, offering valuable data on its reliability and validity, and potentially transforming cognitive assessment practices.

Subjects and Methods

The study sample was sourced from three higher education institutions in China. Between December 2017 and December 2019, a stratified sampling approach was employed to select 180 new students from undergraduate, master's, and doctoral programs as participants. A baseline assessment was first conducted using the Chinese version of the ACE, Digit Span Memory Test (DSMT), and Line-trailing Test (LTT) for cognitive function evaluation. One week later, a retest with the ACE was performed. Based on the results of a pilot study, the ACE selected 8 out of 12 subtests for formal assessment. The correlation between demographic data and the results of each subtest was analyzed. Cronbach's α coefficient was used to evaluate the internal consistency of the ACE, and the correlation between the average response time and the number of correct answers at baseline and retest was analyzed to assess retest reliability. Content validity was evaluated using expert scoring, and the DSMT and LTT were used as criteria to evaluate criterion validity. Exploratory factor analysis and confirmatory factor analysis were conducted to evaluate structural validity. Normality analysis was performed, and skewness was used to assess floor/ceiling effects. After completing all assessments, a questionnaire survey on the use of ACE was conducted.

Participants

The study participants were students from three universities in Shanghai. From December 2017 to December 2019, a total of 180 first-year students were selected from the undergraduate, master's, and doctoral programs by stratified sampling method. After screening for eligibility criteria, 167 students were finally recruited to the study. Signed informed consent was collected from all the participants. The ACE, LTT A & B, DSMT were used for cognitive evaluations. The subject inclusion criteria were as follows: (1) ages ranging from 18 to 30 years; (2) must be receiving undergraduate education or above at the time of enrollment, and (3) able to perform right-handed actions. While the exclusion criteria were (1) suffering from any neurodegenerative diseases, brain trauma, epilepsy, or other serious neurological disorders; (2) previously diagnosed with the spectrum of schizophrenia, bipolar disorder, major depressive disorder, obsessive-compulsive disorder, or any other psychiatric disorders; (3) had a history of major physical diseases; (4) scores on PHQ-9 or GAD-7 scale ≥10; and (5) had a history of anerythrochloropsia.

Research Scheme

This study included two steps, namely the baseline assessment and retesting. At baseline, scores of ACE, DSMT, and LTT-A & B tests were recorded. One week later, the ACE test was retested. After retesting, the participants completed an appraisal questionnaire for the users.

Measurements

This study mainly focused on the reliability and validity of the Chinese version of ACE software, which was developed to evaluate cognitive functions using adaptive algorithms and stimulating feedback. All the verbal and audio information of ACE, including instructions, the content of the setting interface, and program characters were translated from English to Chinese by a Chinese psychiatrist who was a visiting scholar in the US at that time. The Chinese name of the ACE package was "Jiyibao".

There were 12 sub-programs in the ACE: Basic Reaction Time (BRT), Stroop Test (ST), Flanker, Delayed Working Memory (DWM), Mental Rotation (MR), Spatial Cueing (SC), Task Switch (TS), Tap & Trace (TNT), Boxed, SAAT, Spatial Span (SS), and Discrimination. The standard of adaptive adjustment had a response accuracy of 80%. The most important moderator was the response time window. Based on an algorithm, the system could automatically extend the display of stimuli if the participant provided an incorrect or overtime response. However, if the participant gave a correct response, the system would automatically shorten the time window.

1) *BRT*: In this program, when a heart sign showed up on the screen, the participant needed to click the button on the same side of the sign with their non-dominant, and dominant hand, respectively. Data on response time and correctness were automatically collected by the program. When undertaking the test, reactions should be made as quickly as possible. This program aimed to appraise and train the basic response speeds of the participants.
2) *ST*: In this program, the subjects needed to concentrate on the color of the text on the screen while ignoring its meaning. For example, if the white-colored word "red" appeared on the screen, the participant should click the "white" button, rather than 'red'. Data on response time and correctness were collected automatically by the program. Another 'consistent/inconsistent' judgment was also included in the system depending on whether the content was consistent with the color to evaluate and train the participant's ability to focus under substantial distractions.
3) *Flanker*: Here, the participants needed to identify the content of the letter in the center of the screen. If it was "A" or "B", they had to click the bottom left button with their left hand as quick as possible. But if it was "C" or "D", then they had to click the bottom right button with their right hands as quickly as possible. During the process, a random display of one or five letters appeared on the screen. When there were five letters, the participants were required to focus on 'A/B/C/D', and not be disturbed by any other letters. Likewise, data on response time and correctness were collected automatically by the program. Another 'consistent/inconsistent' judgment was also included in the system depending on whether the content was consistent with the color to assess and train the participant's ability of target management.
4) *DWM*: In the first part of this program, the participants were asked to remember and make a judgment of whether the first and third faces were the same. In the second part, the subjects were required to choose whether the second face was greater than 40 years of age. Data on response time and correctness were automatically collected by the program to evaluate and train the participants' short-term memory functions under distractions.
5) *MR*: In this program, the participants were shown two figures successively. Then, the participants were asked to rotate the first figure at 90° clockwise in their minds, and the second figure was shown up shortly on the screen. If it is identical to the rotated figure, click "Yes" as quickly as possible. If not, click "No". This program evaluated the working memory function of the participants.
6) *SC*: In this program, the participants were directed to keep pressing the two moon signs at the bottom of the screen with their two forefingers. "Flying saucers" would then show up on the screen. A "radar" would alert their occurrences in specific areas. When a green flying saucer (not a red one) appeared, the participants needed to click it quickly with their forefingers of dominant hands. During the process, participants should keep pressing the "moon" with their forefingers of non-dominant hands all the time. Data on response time and correctness were automatically recorded by the program to assess and train the participants' abilities to maintain their short-term memory functions.

7) *TS*: In this program, the participants confirmed the notice first. If the clue was a “shape”, they had to discriminate the shape of the subsequent figure (triangle/square). If the clue was a “color”, they were required to recognize the color of the subsequent figure (red or green). Similarly, data on response time and correctness were collected automatically by the program for evaluating the participants’ ability of target management.
8) *TNT*: This program consisted of three parts. First, when a target figure (a green circle) appeared on the screen, the participants were required to click the button as fast as possible with their dominant hands. Then, they had to trace the line of green trailing with their non-dominant hands to follow the green trace of the white line on the screen. Finally, the participants needed to complete the above two tasks spontaneously: the dominant hand “taps”, while the non-dominant hand “traces”. Data on response time and correctness were gathered automatically by the program to evaluate and train the participants’ ability in target management.
9) *Boxed*: Here, the participants needed to distinguish between locations of the breech (top/bottom) of the green box and click the “top” or “bottom” button accordingly as soon as possible with their dominant hands. During the process, red boxes and side-breeched green boxes appeared as interferences. It was a four-stage test based on the difficulty level: “4 items without interference” (4 boxes in total, without interference signs), ”12 items without interference” (12 boxes in total, without interference signs), ”4 items with interference” (4 boxes in total, with interference signs) and ”12 items with interference” (12 boxes in total, with interference signs). Data on response time and correctness were collected automatically by the program for the evaluation of the participant’s ability to direct the target under distractions.
10) *SAAT*: In this program, the participants were asked to click the button with their dominant hands when the target pattern (a heart sign) appeared at the top of the screen but not to make no reaction when the sign appeared at the bottom. Data on response time and correctness are collected automatically by the program. This program included two sub-categories: “continuous” (when the heart sign mostly occurred at the top of the screen) and “impulsive” (when the heart sign mostly occurred at the bottom of the screen). An integrated analysis was obtained to evaluate and train the participants’ ability to maintain attention.
11) *SS*: Initially, there were many irregularly distributed static circles on the screen. Then, some of them were filled with the green color and subsequently faded away in order, one by one. The process started with three colored circles, and the participants were asked to remember the color-filling order and restore them after the color faded. The level of difficulty (number of circles) was escalated if they succeeded twice in the current stage. Data on response time and correctness are recorded automatically by the program for the assessment of the spatial memory functions of the participants.
12) *Discrimination*: In this program, only when green circles appeared on the screen, the participants were directed to click the button as quickly as possible with their dominant hands, while ignoring all the other patterns. Data on response time and correctness were similarly collected by the program to evaluate and train the subjects’ ability to distinguish between different objects.

Results

Demographic Characteristics

A total of 167 subjects met the inclusion criteria for enrollment, but 17 of them did not complete the retest before the given time. The remaining 150 subjects completed all tests and scales within the time limit. The average age of the participants was 20.97±3.36 years. There were 72 males and 78 females in this cohort, including 95(63.30%) undergraduates, 34(22.70%) master’s, and 21(14.00%) doctoral students. Except for 11 minorities, the rest of this cohort were all of Han nationality (see Table 1).

There were no significant correlations between the sub-program results and the subjects’ ages and gender (p>0.05). The absolute values of the Kendall tau-b correlation coefficients between the participants’ educational levels and the results of BRT (right-handed), Flanker, TS, Boxed, SAAT, and TNT were between 0.4 and 0.7, suggesting that the above sub-programs had a moderate negative correlation with the educational level. The above correlational relationship was not found for other sub-programs (see Table 2).

 https://doi.org/10.2147/PRBM.S491302

Table 1 Demographic Data

Variable	Group	n	Percentage (%)
Gender	Male	72	48.00%
	Female	78	52.00%
Age (years old)	18–22	95	63.30%
	23–25	34	22.70%
	26–29	21	14.00%
Level of Education	Undergraduate Student	95	63.30%
	Master Student	34	22.70%
	Doctoral Student	21	14.00%
Nationality	Han	139	92.67%
	Minorities	11	7.33%

Table 2 Correlations of Gender, Age, and Educational Level with the Results of Sub-Programs

Program/Sub-Test Name	Correlation Coefficient with Age (tau-b)	Correlation Coefficient with Gender (tau-b)	Correlation Coefficient with Educational Level (tau-b)
BRT (Left hand)	0.061	0.027	−0.248
BRT (Right hand)	0.066	0.012	−0.419*
Stroop Test	0.090	−0.026	−0.213
Flanker	−0.032	0.043	−0.288
Task Switch	0.035	0.150	−0.707**
Boxed (without distraction-12 items)	0.003	−0.002	−0.762**
Boxed (without distraction-4 items)	0.009	0.106	−0.425*
Boxed (with distraction-12 items)	−0.009	0.011	−0.630**
Boxed (with distraction-4 items)	−0.031	0.129	−0.606**
SAAT (impulsive)	−0.089	0.052	−0.625**
SAAT (sustained)	−0.024	0.090	−0.513**
TNT	0.005	0.073	−0.490**
TNT (Tap Only)	0.048	0.016	−0.574**
Spatial Span	0.078	0.005	−0.537**

Notes: ** $p<0.01$, * $p<0.05$.

Internal Consistency Reliability

The reliability analysis was executed using the SPSS software. The overall Cronbach's α value was 0.846 ($p<0.01$), and Cronbach's α of each sub-program was above 0.8 (ranging from 0.816 to 0.887, $p<0.01$; see Table 3), indicating the ideal reliability of this ACE software. The value of Cronbach's α could detect if any sub-program was deleted from the analysis. However, the overall internal consistency of the ACE did not increase, indicating satisfying internal consistency reliability of each sub-program. No sub-program was required to delete.

Test-Retest Reliability

Scatterplots for the baseline and retest results of each sub-program were generated by SPSS. A tendency of linear correlation was observed between the test and retest results of each sub-program. Pearson's correlation coefficients were analyzed between the pairing results. The Pearson's correlation coefficients (r) of the ST, Boxed, Flanker, TS, SAAT, and TNT ("Tap" only) were all greater than 0.7, suggesting satisfying retest reliability of the above programs. The results were statistically significant ($p<0.01$). The Pearson's correlation coefficients (r) of BRT (bilateral), TNT, and SS ranged from 0.4 to 0.7, indicating statistically significant ($p<0.01$), and relatively satisfying results of the retest reliability (see Table 4).

Content Validity

After retrieving the scores of the panel, content validity was calculated using a series of indicators. In the panel, there were five psychiatrists to evaluate the content validity (see the scores in Tables 5 and 6). All five members believed that

Table 3 Internal Consistency Reliability of Sub-Programs and the Overall Results

Program/Sub-Test Name	CITC	Cronbach's α if Item Deleted	Cronbach's α
Boxed (with distraction-4 items)	0.766	0.816**	0.846**
Boxed (with distraction-12 items)	0.315	0.887**	
Boxed (without distraction-4 items)	0.769	0.818**	
Boxed (without distraction-12 items)	0.708	0.823**	
BRT (Left hand)	0.567	0.835**	
BRT (Right hand)	0.482	0.840**	
Flanker	0.708	0.824**	
SAAT (impulsive)	0.507	0.837**	
SAAT (sustained)	0.385	0.843**	
Task Switch	0.580	0.830**	
TNT (Tap Only)	0.585	0.836**	
TNT	0.480	0.837**	
Spatial Span	−0.335	0.851**	
Stroop Test	0.608	0.828**	

Note: ***p*<0.01.

Table 4 Retest Reliability of the Sub-Programs

Program/Sub-Test Name	Pearson Correlation (r)
BRT (Left hand)	0.699**
BRT (Right hand)	0.623**
Stroop Test	0.882**
Flanker	0.849**
Task Switch	0.747**
Boxed (without distraction-12 items)	0.869**
Boxed (without distraction-4 items)	0.889**
Boxed (with distraction-12 items)	0.740**
Boxed (with distraction-4 items)	0.841**
SAAT (impulsive)	0.736**
SAAT (sustained)	0.707**
TNT	0.692**
TNT (Tap Only)	0.774**
Spatial Span (Score)	0.532*

Notes: ***p<0.01*, **p<0.05.*

Table 5 Scores of the Panel on ACE

Program Name	Member 1	Member 2	Member 3	Member 4	Member 5
BRT	4	4	3.5	4	4
Boxed	3.5	4	3.5	4	4
Flanker	4	4	3.5	4	4
SAAT	3.5	4	3	4	3
Task Switch	4	4	4	3.5	4
TNT	3.5	4	4	3.5	4
Spatial Span	4	3	4	3.5	4
Stroop Test	4	4	4	3.5	4

the translation of ACE was accurate and appropriate, and the Chinese version of the ACE was operation friendly to Chinese users. Besides, sub-programs of the ACE could effectively evaluate their cognitive domains. I-CVC of each sub-project was 1.00, and K* was equaled to 1.00. S-CVI/Ave was also 1.00 (>0.90), indicating that as a method for assessment, the content validity of ACE was excellent.

Table 6 Index of Content Validity

Program Name	Numbers of Scores Above 3	I-CVC	Pc	K*
BRT	5	1.00	0.041	1.00
Boxed	5	1.00	0.041	1.00
Flanker	5	1.00	0.041	1.00
SAAT	5	1.00	0.041	1.00
Task Switch	5	1.00	0.041	1.00
TNT	5	1.00	0.041	1.00
Spatial Span	5	1.00	0.041	1.00
Stroop Test	5	1.00	0.041	1.00

Criterion Validity

In this study, the criterion validity was obtained by conducting Spearman correlation test or Pearson's correlation test as appropriate. When the correlation coefficient (r) was greater than 0.7, the criterion validity was satisfying. A value of $0.4 \leq r \leq 0.7$ indicated that the criterion validity was relatively ideal, and an r lower than 0.4 suggested a poor criterion validity (see Tables 7 and 8).

(1) Taking Digit Span as the criterion, the results exhibited that the Spearman's r of BRT, TS, and Boxed were between 0.4 and 0.7 ($p<0.05$). The criterion validities were ideal for these sub-programs, while the r values of other sub-programs were lower than 0.4.
(2) Taking LTT-A as the criterion, the results showed that the Spearman's r of Flanker, TS, Boxed, SAAT, and TNT ("Tap" only) were between 0.4 and 0.7 ($p<0.01$). The criterion validities were ideal. The other programs, including BRT, ST, and TNT, had poor criterion validities.
(3) Taking LTT-B as the criterion, the Pearson's r values of BRT (left hand) and LTT were lower than 0.4, indicating that the criterion validity was poor. The r values of TS and block testing (with or without interference – 4 items) were higher than 0.7 ($p<0.01$), indicating very good criterion validity. The r values of other sub-items were all between 0.4 and 0.7 ($p<0.01$), again indicating good criterion validity.

Table 7 Correlation of Sub-Programs Baseline Correction and Response Timing of DSMT

Program/Sub-Test Name	Correlation with Digit Span (r)
BRT (Left hand)	0.549**
BRT (Right hand)	−0.059
Stroop Test	0.172
Flanker	0.171
Task Switch	0.450*
Boxed (without distraction-12 items)	0.448*
Boxed (without distraction-4 items)	0.330
Boxed (with distraction-12 items)	0.023
Boxed (with distraction-4 items)	0.282
SAAT (impulsive)	−0.057
SAAT (sustained)	0.273
TNT	0.197
Spatial Span	0.142

Notes: ** $p<0.01$, * $p<0.05$.

Table 8 Correlation of Sub-Programs and Line-Trailing Test (Response Time as Standard)

Program/Sub-Test Name	Correlation with Line-Trailing Test-A (r)	Correlation with Line-Trailing Test-B (r)
BRT (Left hand)	0.154	0.161
BRT (Right hand)	0.367*	0.435*
Stroop Test (Consistent)	0.369*	0.539**
Stroop Test (Inconsistent)	0.376*	0.535**
Flanker (Consistent)	0.544**	0.657**
Flanker (Inconsistent)	0.551**	0.649**
Task Switch	0.463**	0.716**
Boxed (without distraction-12 items)	0.657**	0.664**
Boxed (without distraction-4 items)	0.598**	0.747**
Boxed (with distraction-12 items)	0.588**	0.661**
Boxed (with distraction-4 items)	0.598**	0.710**
SAAT (impulsive)	0.472**	0.523**
SAAT (sustained)	0.518**	0.542**
TNT	0.250	0.383*
TNT (Tap Only)	0.565**	0.645**

Notes: ** $p<0.01$, * $p<0.05$.

Construct Validity

In this study, the construct validity referred to whether the test result of each sub-program of ACE reflected the exact cognitive control ability. Participants' maximum correct answers were selected for the analysis of the SS sub-program, and their average time of each correct response was considered for other sub-program analyses.

First, the feasibility of exploratory factor analysis was tested by SPSS 24.0. When the KMO was above 0.8, the construct validity was found to be ideal. A value between 0.7 and 0.8 indicated that the construct validity was relatively satisfying. While a KMO score between 0.6 and 0.7 represented an acceptable construct validity. The structural validity was considered poor if the KMO value was less than 0.6.[24] For this cohort, the KMO was 0.724, and the χ^2 value of Bartlett's test was 3939.84 ($p<0.001$), indicating relatively satisfying construct validity. The principal component analysis (PCA) was used for the correlation matrix rotated by a maximum variance method. The maximum number of convergence iterations was 25 by default. The results are presented in descending order in Table 9.

Floor and Ceiling Effects

The response rate of all valid participants of each sub-program was 100% in baseline assessment and retesting. See Table 10 for the sub-programs' average response time, standard deviation (SD), and skewness of the LTT-A & B, as well as the average score, SD, and skewness of the DSMT, and SS test.

The absolute value of skewness of all sub-programs of ACE was less than 1, indicating that there were no significant floor and ceiling effects, and the software was able to carry out appraisals of cognitive functions of various levels.

Participants' Rating on the ACE

After completing the baseline test and retesting, the participants were invited to rate ACE voluntarily. The results were as follows: (1) the difficulty of each sub-program: low to average (15.33%), average (71.33%), and average to high (6.67%); (2) the entertaining aspect of each sub-program: low to average (17, 11.33%), average (76, 50.67%), and average to high (57, 38%); (3) total time for the test: low to average (5, 3.33%), average (91, 60.67%), and average to high (54, 36%); (4) willingness for using similar software in the future: low to average (28, 18.67%), average (80, 53.33%), and average to high (42, 28%). See Table 11 for detailed descriptions.

 https://doi.org/10.2147/PRBM.S491302

Table 9 Results of the Construct Validation

Program/Sub-Test Name	Factor Loading			Communality
	Attention	Target Management	Working Memory	
Boxed (with interference-4 items)	0.800**	0.171	0.151	0.693
Boxed (without interference-4 items)	0.804**	0.182	0.223	0.730
Boxed (without interference-12 items)	0.849**	0.181	0.177	0.785
Boxed (with interference-12 items)	0.888**	0.234	0.021	0.844
Flanker	0.868**	0.227	0.244	0.864
Stroop Test	0.760**	0.209	0.304	0.714
SAAT (sustained)	0.775**	0.193	−0.012	0.637
SAAT (impulsive)	0.736**	0.152	−0.058	0.568
TNT (Tap Only)	0.243	0.796**	0.227	0.744
TNT	0.159	0.785**	0.268	0.714
Task Switch	0.260	0.759**	0.277	0.720
Spatial Span	0.104	0.187	0.839**	0.749
Eigen Value (Unrotated)	8.698	2.777	2.142	–
% of Variance (Unrotated)	41.417%	13.222%	10.198%	–
Cumulative % of Variance (Unrotated)	41.417%	54.639%	64.837%	–
Eigen Value (Rotated)	6.480	4.351	2.785	–
% of Variance (Rotated)	30.856%	20.719%	13.263%	–
Cumulative % of Variance (Rotated)	30.856%	51.574%	64.837%	–
KMO	0.724			–
Bartlett's Test	3939.838			–
df	210			–
p value	0.000			–

Notes: **absolute value of factor loading>0.4.

Discussion

ACE software was designed and developed based on classic neurocognitive assessment paradigms such as ST, Flanker, and visual scanning.[25,26] It was primarily used for assessing cognitive control in attention, working memory, and goal management. The integration of the adaptive step-transformation algorithm ensured that the results reflected participants' actual cognitive functions, eliminating age-associated deceleration and bias due to test parameters. Our findings indicate no correlation between sub-program results and age or gender, aligning with previous studies that observed more significant age differences in elderly and depressed populations,[25,26] and no significant difference in cognitive control abilities between males and females.[27]

The results of each sub-program did not show any correlation with gender, consistent with Schirmer's reports.[27] There was no significant difference between males and females regarding their cognitive controlling abilities. However, many studies suggest that the influence of gender on cognitive control function could be significant. For example, Muller et al[28] found that men had higher flexibility than women; Ai[29] found that females can perform better under behavioral inhibitions than their male counterparts, especially under negative emotions. Clayson et al[30] revealed that women are more likely to be interfered with by distractions under inconsistent conditions in conflict control. Under inconsistent conditions, women exhibit longer reaction times than men, while under consistent conditions, there was no significant difference for gender.[30]

The absolute value of Kendall's tau-b correlation coefficients between the participants' educational levels and the results of BRT (right hand), Flankers, TS, Boxed, SAAT, and TNT were between 0.4 and 0.7, suggesting that the above sub-programs negatively correlated with educational level at a moderate level. The higher the educational level, the shorter the correct response time, and the faster the speed of response. Multiple empirical studies have revealed that higher education is a protective factor against cognitive impairment.[31] The finding of this study was also in agreement with the theory that education could promote the development of an individual's cognitive function.[31]

https://doi.org/10.2147/PRBM.S491302

Table 10 M, SD, Skewness and Response Rate of Baseline Assessment and Retest (Time of Response/Score)

Baseline					Retest				
Program/Sub-test Name	**Average (Millisecond)**	**Variance (Millisecond)**	**Skewness**	**Response Rate**	**Program/Sub-Test Name**	**Average (Millisecond)**	**Variance (Millisecond)**	**Skewness**	**Response Rate**
BRT (Left hand)	355.07	40.49	0.540	100%	BRT (Left hand)	328.28	22.42	0.012	100%
BRT (Right hand)	339.46	39.79	0.496	100%	BRT (Right hand)	328.76	21.25	0.372	100%
Stroop Test	602.77	58.68	0.746	100%	Stroop Test	575.34	29.36	0.343	100%
Flanker	570.90	59.49	0.756	100%	Flanker	550.54	33.08	0.897	100%
Task Switch	502.22	71.90	0.445	100%	Task Switch	489.78	34.07	0.251	100%
Boxed (without interference-12 items)	640.37	64.44	0.617	100%	Boxed (without interference-12 items)	625.06	32.47	0.430	100%
Boxed (without interference-12 items)	613.52	65.66	0.636	100%	Boxed (without interference-12 items)	599.46	33.01	0.325	100%
Boxed (with interference-12 items)	879.68	143.09	−0.118	100%	Boxed (with interference-12 items)	857.05	118.36	0.008	100%
Boxed (with interference-4 items)	715.50	76.15	0.113	100%	Boxed (with interference-4 items)	684.76	41.54	−0.114	100%
SAAT (impulsive)	412.01	54.72	0.868	100%	SAAT (impulsive)	418.78	20.33	0.405	100%
SAAT (sustained)	392.41	44.44	0.265	100%	SAAT (sustained)	390.95	18.41	0.683	100%
TNT	620.69	118.67	0.725	100%	TNT	573.42	32.73	0.815	100%
TNT (Tap Only)	467.39	38.61	0.337	100%	TNT (Tap Only)	458.51	20.96	−0.132	100%
Spatial Span	6.09 (score)	0.99	−0.531	100%	Spatial Span	6.12(score)	0.55	−0.370	100%
Digit Span	16.45 (score)	1.99	−0.15	100%					
Line-trailing Test-A	32.16	8.80	0.226	100%					
Line-trailing Test-B	63.77	12.59	0.323	100%					

Table 11 The Participants' Rating on the ACE

Item Name	Low to Average (n)	Percentage (%)	Average (n)	Percentage (%)	Average to High (n)	Percentage (%)
Overall difficulty	23	15.33	107	71.33	10	6.67
Entertaining Aspect	17	11.33	76	50.67	57	38.00
Test Time in Total	5	3.33	91	60.67	54	36.00
Willingness for using a similar software	28	18.67	80	53.33	42	28.00

Reliability

Cronbach's α of every sub-program and the overall result were all above 0.8 ($p<0.01$), indicating a high consistency level of ACE and its sub-programs for cognitive control assessment. There was no significant difference between the "Item deleted α" and the overall Cronbach's α, suggesting that the overall internal consistency reliability of ACE did not increase with any sub-program being deleted. It also indicated that the internal consistency reliability of each sub-program was satisfying, and none of them should be deleted.

In this study, the retest reliabilities of the ST, Boxed, Flanker, TS, SAAT, and TNT ("Tap" only) sub-programs were extremely satisfying and statistically significant ($p<0.01$). The retest reliabilities of BRT (non-dominant and dominant hand), TNT, and SS were also satisfying and statistically significant ($p<0.01$). Therefore, the above-mentioned sub-programs can be used for the evaluation of goal management, working memory, and attention ability.

Validity

In this study, a 4-point method was used for scoring. Combining with the definition of I-CVI, 4 points were summarized into two categories: if scored between 1 and 2, the sub-program was not related to its dimension and had poor correspondence; but if scored between 3 and 4, then the sub-program was related to its dimension and had ideal representativeness. The scoring panel of this study included 5 members. According to the discipline, when the number of panel members was less than or equal to 5, all members must reach a consensus on the representativeness of each sub-program (ie, I-CVI=1.00). The results of this study met the above requirement. Besides, S-CVI/Ave was also equal to 1.00 (>0.90),[32] and the K* value was 1.00 (>0.74),[33] indicating an extremely satisfying content validity. The assessment results of ACE were highly matched with the actual cognitive control level.

The criterion validity test using LTT-A, Flanker, TS, Discrimination, Boxed, SAAT, and TNT showed a satisfying validity, which was associated with the attention-evaluating nature of the above sub-programs. When completing cognitive control tasks, the participants needed to convert task requirements into short-term endogenous goals to guide their behaviors and abilities of goal management and attention/working memory interaction and complementation.

The LTT-B requires a combination of multiple cognitive processes. In the criterion validity test, except for TNT, the validities of other sub-programs of ACE were relatively similar. It might be because TNT requires both dominant and non-dominant hands to complete different operations at the same time. In addition to the function of cognitive control, this process also examined the capability of the left and right brain's cooperative work.

The construct validity was analyzed by factor analysis. In this study, the KMO value was 0.724 (>0.6), suggesting the validity of the data. Moreover, the commonality of all the subtests was higher than 0.4, indicating that the function of the corresponding cognitive domains can be effectively evaluated by ACE sub-programs. On the other hand, programs of ACE can be divided into three dimensions—attention, goal management, and working memory. Therefore, three factors were extracted. Rates of variation of these three factors were 30.86%, 20.72%, and 13.26%, respectively. The cumulative rate of variance (rotated) was 64.84% (>50%), revealing that the corresponding function can be effectively evaluated by each sub-program. Finally, from the perspective of the factor-loading coefficient analysis, when the absolute value of the coefficient is greater than 0.4, it suggests that there is a relationship between the sub-program and its corresponding dimensions. In this study, the coefficients of all sub-programs under their matching dimension were greater than 0.4.

User Feedback on ACE

The participants' comments on ACE included the level of difficulty of each sub-item, the level of entertainment of each sub-item, and time to complete the test, and the possibility of using similar software in the future. A 76.4% of the participants thought that the sub-programs of ACE were of medium difficulty levels; 54.2% of the participants commented that the degree of entertainment of each sub-program was moderate, which induced their curiosity and enthusiasm; 65% of the participants reported that the test time was reasonable and will not cause impatience, and 81.3% of the participants expressed their willingness to use similar software for assessment in the future. Generally, the ACE software was moderately difficult, entertaining, and user-friendly.

Broader Applications

The findings of this study serve as evidence for the future application of ACE to a broader demographic. In the future, we aim to expand the reach of the ACE software, providing a new method for early screening of cognitive control impairments and identification of high-risk groups for specific diseases.

Theoretical Implications

Our findings indicate that the cognitive control functions assessed by the ACE software are not influenced by age or gender, suggesting that cognitive control functions may be more stable across different populations than previously thought. This provides new insights into the stability of these functions.

Practical Implications

The ACE software's ability to accurately assess cognitive control functions holds practical significance for clinical work. It has the potential to assist in the development of personalized treatment plans and in evaluating the cognitive function of patients in the future.

Limitations

The primary limitation of this study is the singular nature of the sample, which was exclusively composed of healthy college students. Enhancing sample diversity in future research will be essential to broaden the applicability of our findings. Furthermore, the timing of the tests, conducted at various times throughout the day, may have influenced participants' performance. In subsequent studies, we will endeavor to mitigate the potential influence of such variables on the research outcomes.

Conclusions

The ACE software offers a reliable and valid tool for assessing cognitive control functions, with implications for both theoretical understanding and practical application in diverse populations. Our study lays the groundwork for further research and potential clinical applications of this technology.

Ethics Approval

This is an observational study among normal college students. The Tongji University Research Ethics Committee has confirmed that no ethical approval is required.

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

Consent to Publish

The authors affirm that human research participants provided informed consent for publication of the data in all figures and tables.

Funding

This study was funded by Talent Introduction Fund of Fudan University Shanghai Cancer Center (YJRC202105), Shanghai Anticancer Association EYAS PROJECT (SACA-CY22C06) and Xisike Clinical Oncology Research Foundation (CSCO-Haosen) (Y-HS202202-0165); Three-year action plan for the construction of Shanghai public health system (2023–2025), academic leaders cultivating project (GWVI-11.2-XD232).

Disclosure

The authors have no relevant financial or non-financial interests to disclose.

References

1. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci*. 2013;36(3):181–204. doi:10.1017/S0140525X12000477
2. Crawford KM, Choi K, Davis KA, et al. Exposure to early childhood maltreatment and its effect over time on social cognition. *Dev Psychopathol*. 2022;34(1):409–419. doi:10.1017/S095457942000139X
3. Ladegaard N, Larsen ER, Videbech P, Lysaker PH. Higher-order social cognition in first-episode major depression. *Psychiat Res*. 2014;216(1):37–43. doi:10.1016/j.psychres.2013.12.010
4. Siew CSQ, Wulff DU, Beckage NM, Kenett YN. Cognitive Network Science: a Review of Research on Cognition through the Lens of Network Representations, Processes, and Dynamics (vol 2019, 2108423, 2019). *Complexity*. 2020;2020:1.
5. Collins A, Koechlin E. Reasoning, Learning, and Creativity: frontal Lobe Function and Human Decision-Making. *PLoS Biol*. 2012;10(3):e1001293. doi:10.1371/journal.pbio.1001293
6. Diamond A. Executive Functions. *Annu Rev Psychol*. 2013;64:135–168.
7. Amer T, Campbell KL, Hasher L. Cognitive Control As a Double-Edged Sword. *Trends Cognit Sci*. 2016;20(12):905–915. doi:10.1016/j.tics.2016.10.002
8. Washburn DA. The Stroop effect at 80: the competition between stimulus control and cognitive control. *J Exp Anal Behav*. 2016;105(1):3–13. doi:10.1002/jeab.194
9. Inzlicht M, Bartholow BD, Hirsh JB. Emotional foundations of cognitive control. *Trends Cognit Sci*. 2015;19(3):126–132. doi:10.1016/j.tics.2015.01.004
10. Scheibel RS. Functional Magnetic Resonance Imaging of Cognitive Control following Traumatic Brain Injury. *Front Neurol*. 2017;8:352. doi:10.3389/fneur.2017.00352
11. Koster EHW, Hoorelbeke K, Onraedt T, Owens M, Derakshan N. Cognitive control interventions for depression: a systematic review of findings from training studies. *Clin Psychol Rev*. 2017;53:79–92. doi:10.1016/j.cpr.2017.02.002
12. Jaeggi SM, Buschkuehl M, Shah P, Jonides J. The role of individual differences in cognitive training and transfer. *Mem Cognit*. 2014;42(3):464–480. doi:10.3758/s13421-013-0364-z
13. Stamenova V, Levine B. Effectiveness of goal management training(R) in improving executive functions: a meta-analysis. *Neuropsychol Rehabil*. 2019;29(10):1569–1599. doi:10.1080/09602011.2018.1438294
14. Zuberer A, Kucyi A, Yamashita A, et al. Integration and segregation across large-scale intrinsic brain networks as a marker of sustained attention and task-unrelated thought. *Neuroimage*. 2021;229:117610. doi:10.1016/j.neuroimage.2020.117610
15. Darby KP, Deng SW, Walther DB, Sloutsky VM. The Development of Attention to Objects and Scenes: from Object-Biased to Unbiased. *Child Dev*. 2021;92(3):1173–1186. doi:10.1111/cdev.13469
16. Gazzaley A, Nobre AC. Top-down modulation: bridging selective attention and working memory. *Trends Cognitive Sciences*. 2012;16(2):129–135. doi:10.1016/j.tics.2011.11.014
17. Jans B, Peters JC, De Weerd P. Visual Spatial Attention to Multiple Locations at Once: the Jury Is Still Out. *Psychol Rev*. 2010;117(2):637–682. doi:10.1037/a0019082
18. Vermeylen L, Braem S, Notebaert W, Ruitenberg MFL. The subjective evaluation of task switch cues is related to voluntary task switching. *Cognition*. 2022;224:105063. doi:10.1016/j.cognition.2022.105063
19. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189–198. doi:10.1016/0022-3956(75)90026-6
20. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695–699. doi:10.1111/j.1532-5415.2005.53221.x
21. Gratton G, Cooper P, Fabiani M, Carter CS, Karayanidis F. Dynamics of cognitive control: theoretical bases, paradigms, and a view for the future. *Psychophysiology*. 2018;55(3). doi:10.1111/psyp.13016
22. Kessels RPC. Improving precision in neuropsychological assessment: bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *Clin Neuropsychol*. 2019;33(2):357–368. doi:10.1080/13854046.2018.1518489
23. Öhman F, Hassenstab J, Berron D, Schöll M, Papp KV. Current advances in digital cognitive assessment for preclinical Alzheimer’s disease. *Alzheimers Dement*. 2021;13(1):e12217.
24. Haruna H, Hu X, Chu SKW, Mellecker RR. Initial Validation of the MAKE Framework: a Comprehensive Instrument for Evaluating the Efficacy of Game-Based Learning and Gamification in Adolescent Sexual Health Literacy. *Ann Glob Health*. 2019;85(1). doi:10.5334/aogh.1110
25. Hammar A, Ronold EH, Rekkedal GA. Cognitive Impairment and Neurocognitive Profiles in Major Depression-A Clinical Perspective. *Front Psychiatry*. 2022;13. doi:10.3389/fpsyt.2022.764374
26. Lovden M, Fratiglioni L, Glymour MM, Lindenberger U, Tucker-Drob EM. Education and Cognitive Functioning Across the Life Span. *Psychol Sci Publ Int*. 2020;21(1):6–41. doi:10.1177/1529100620920576

27. Schirmer A, Zysset S, Kotz SA, Yves von Cramon D. Gender differences in the activation of inferior frontal cortex during emotional speech perception. *Neuroimage*. 2004;21(3):1114–1123. doi:10.1016/j.neuroimage.2003.10.048
28. Muller J, Dreisbach G, Brocke B, Lesch KP, Strobel A, Goschke T. Dopamine and cognitive control: the influence of spontaneous eyeblink rate, DRD4 exon III polymorphism and gender on flexibility in set-shifting. *Brain Res*. 2007;1131(1):155–162. doi:10.1016/j.brainres.2006.11.002
29. Ai CT: The research about gender difference in behavioral inhibitory control under anger emotion. *Master Thesis*. Southwest University; 2014.
30. Clayson PE, Clawson A, Larson MJ. Sex differences in electrophysiological indices of conflict monitoring. *Biol Psychol*. 2011;87(2):282–289.
31. Borella E, Carretti B, Riboldi F, De beni R. Working memory training in older adults: evidence of transfer and maintenance effects. *Psychol Aging*. 2010;25(4):767–778.
32. Walz CF, Strickland OL, Lenz ER. Measurement in nursing and health research. *Nephrol Nurs J*. 2006;33(1):76.
33. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007;30(4):459–467. doi:10.1002/nur.20199