

RESEARCH ARTICLE

Working memory test battery for young adults: Computerized working memory assessment

Liang Ma¹, Lei Chang², Xiaoying Chen¹, Renlai Zhou^{3*}

1 School of Psychology, Beijing Normal University, Beijing, China, **2** Department of Psychology, University of Macau, Macau S.A.R., China, **3** Department of Psychology, School of Social and Behavior Sciences, Nanjing University, Nanjing, China

* rlzhou@nju.edu.cn



OPEN ACCESS

Citation: Ma L, Chang L, Chen X, Zhou R (2017) Working memory test battery for young adults: Computerized working memory assessment. PLoS ONE 12(3): e0175047. <https://doi.org/10.1371/journal.pone.0175047>

Editor: Jerson Laks, Universidade Federal do Rio de Janeiro, BRAZIL

Received: December 6, 2016

Accepted: March 20, 2017

Published: March 31, 2017

Copyright: © 2017 Ma et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was funded by the Key Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu (2015JDXM001). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interests exist.

Abstract

This study developed a battery of computerized working memory (WM) tests and a scoring system suitable for young adult users. The tests comprised five classic tasks derived from Baddeley's model of WM, reflecting each of the five WM functions. We recruited 115 undergraduate and graduate students from various academic fields and constructed a preliminary WM scoring norm for young adults. The scoring norm was used as a basis for developing a computerized assessment system. The results of correlation analysis show that the fluid intelligence of young adults is related to the memory function of WM, but not to the central executive system. The proposed working memory test battery for young adults comprehensively reflects the WM capacity of adults.

Introduction

1.1. State of the art

Working memory (WM) is a cognitive system with limited capacity that enables the temporary storage and manipulation of information. WM is necessary for such complex tasks as comprehension, learning, and reasoning [1, 2], and comprises the following three components: the phonological loop, visuospatial sketch pad, and central executive system. The phonological loop is a temporary storage system in which acoustic or speech-based information can be held as memory traces that spontaneously fade. The visuospatial sketch pad temporarily stores visual and spatial information. The central executive is responsible for attentional control and information processing pertaining to WM [3]. The phonological loop and visuospatial sketch pad comprise the information storage system of WM, and the central executive serves as the information processing system of WM. According to the operation of WM information in the storage system, the central executive is divided into the following three distinct yet strongly interactive functions: (1) controlling shifting between tasks or mental sets (i.e., shifting function); (2) controlling the inhibition of proponent responses (i.e., inhibition function); and (3) controlling the updating and monitoring of WM representations (i.e., updating function) [4]. WM correlates highly with fluid intelligence [5–7] and reasoning ability [8–10]. Jaeggi trained

the WM capacity of adults through a dual n-back task [11]. Zhao trained 9–11-year-old children through running WM tasks [12]. Rudebeck trained young adults using a spatial WM training task [13]. Wang trained children with a running WM task [14]. All of these studies have shown increases in fluid intelligence, thereby demonstrating the close relationship between WM development and fluid intelligence.

WM is also closely related to decreases in human cognitive abilities. Therefore, if we develop a WM assessment for adults, researchers can estimate the severity of disease from a cognitive perspective. In neuropathological research on Alzheimer's disease, the cognitive decline associated with dementia has provided subtle evidence of a decline in WM capacity [15]; the higher WM capacity a person has, the slower the decline in his or her cognitive ability during the early stage of the disease [16]. In addition, other cognitive diseases such as attention deficit hyperactivity disorder [17], schizophrenia [18], time estimation disorder [19], reading disorders [20], and emotional regulation disorder [21, 22] correlate highly with WM capacity. Because of the importance of WM, research on WM measurement has attracted considerable attention in recent years. Gathercole et al. developed a set of WM tests for children aged 6–7 years based on Baddeley's three-component model. However, of the aforementioned three WM components, the test was effective only for measuring the phonological loop [23]. In recent years, an increasing number of WM test batteries have been produced. Kenny and Hicks [24] developed a WM capacity test set called Online Working Memory Lab (OWL), which consisted of the following four complex span tasks in its first version: operation span, reading span, symmetry span, and running span [25]. In its most recent version [24], patterns were used as task material instead of letters, which were more difficult to write in unsupervised settings. Because the changing of memory material prevented cheating, the test results were as reliable as laboratory results.

Van [26] developed a set of visual WM test procedures for children aged 6–12 years that is similar to a visual version of a running memory task called "Lion Game." In each trial, eight lions of different colors were displayed consecutively at different locations in a 4 x 4 matrix. The children were asked to remember the last location where a lion of a certain color had appeared and use the mouse to click on that location after the sequence had ended. The task consisted of five levels, and WM load was scored based on the number of colors (i.e., locations) that the children had to remember and update. The most notable features of this task were the self-reliant administration design and web browser—based format. However, the selection element of the task was too simple, and thus the testing effect was limited in the following manners: First, the task was not a test of pure visual span; with the mission ongoing, each child had to continually update his or her memory to recall the preceding locations, which served as indicators of the updating function in the central executive function. Second, a mouse was required, rendering it difficult to accurately record reaction times, which are a crucial indicator for evaluating WM.

Oswald [27] developed a shortened computerized measure of WM capacity by representatively sampling items from operation span, reading span, and symmetry span tasks. Although these tasks were similar to those employed in other studies, notably, the test time was shortened.

In summary, the main features of the WM test battery are as follows: (1) Shorter time; implementation of WM measures is generally time-consuming, and examinees and researchers are often required to manage the tension between limited testing time and the need to reliably measure numerous constructs. A short and effective measure of WM capacity could be a major practical benefit in the future [27]. Previous studies have also proven that shortened test tasks can effectively test WM ability [28]. (2) A computerized or web-based format, thereby facilitating the collection of high quantities of data. The most significant research obstacle is

that the quality of collected data is unsatisfactory. Without the supervision of research assistants, online task participants may cheat.

We proposed a test with a simple and clear paradigm to test all WM components within a short period of approximately 30 minutes. Because our test was based on a local computer and conducted entirely through keyboard operation, no network latency or response time bias could be caused by operational problems. This test program can not only record the accuracy of each task, but also accurately record the reaction time of each trial to evaluate WM capacity more accurately. Several general limitations of WM tests are defined in the following section.

1.2. Limits of the current literature

The limitations of extant research on WM tests are described as follows.

1. The literature lacks complete sets of comprehensive tests. Tests of WM components are either relatively homogeneous or involve test tasks that do not clearly reflect the functions of specific components. Langer designed the “tower of fame” task and argued that in addition to the relational integration component of WM [29], successful completion of this task may also be related to the storage function. The only task used in the WM test by Siegel was the auditory span test, which does not reflect overall WM ability [30]. León-Domínguez recently used an improved version of the n-back task to measure the central executive function [31], but this approach does not measure the inhibitory function or switch function. The more widely recognized WM test sets currently in use are the Automated Working Memory Assessment (AWMA) [32] and Working Memory Index in the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) [33]. Moreover, some scales pertaining to WM such as the Working Memory Rating Scale [34], Behavior Rating Inventory of Executive Function [35], and Conners’ Teacher Rating Scale [36] are not cognitive tests, but are rather questionnaires designed mainly for students. According to teachers’ responses to questionnaires, researchers can determine whether students’ WMs are defective. However, the results of this approach are influenced by the teachers’ subjective feelings, and thus do not reflect the students’ WM potential [37]. The establishment of the tests in the present study was based on the Baddeley’s three-component model of WM [1, 2], which has been thoroughly tested and is widely accepted as a reliable model among researchers. In Baddeley’s model, the phonological loop, visuospatial sketch pad, and central executive (including inhibition, switching, and updating functions) are measured. A participant’s final score, which is derived from five subtest scores, provides a clear and comprehensive indication of his or her overall WM capacity.
2. Evaluation criteria are ambiguous. Most WM tests adhere to the original score level, thereby creating the problem that using the task accuracy as a direct indicator impairs comparability between tests. In the present study, all of the test scores were standardized to enable comparisons between various studies. Each participant’s test score could be converted into a derived score after completion of the measurement to illustrate the degree of his or her capability.
3. A lack of large-scale group tests and score norms render further comparison of the performance of WM tests impossible. With the exception of the AWMA and WISC-IV, which contain scoring norms, a regular norm for WM tests has not been established. One aim of the present study was to establish an appropriate test norm for young adults with a view to constructing an embryonic national model for future reference.

4. Few computerized WM tests have been developed; consequently, tasks can be applied for individuals but not for groups, yielding low test efficiency. Berg used a visual matrix task, the Corsi block-tapping test, semantic classification, and an auditory digit sorting task to examine the relationship between WM and mathematical ability [38]. The tasks were difficult to program and operate on a computer screen, which influenced the test results. Moreover, (verbal) digit sort, (verbal) digit-span backward, the WISC-IV Working Memory Index, AWMA, and other practical tasks are prone to similar problems. All the experimental tasks in the present study were computerized to ensure that the experimenter effect would be avoided and reaction times would be recorded accurately.

1.3. Study objective

The aim of the present study was to establish a set of effective WM capacity evaluation tools and scoring norms suitable for adults.

The aforementioned studies suggest that cognitive ability, health status, and daily life are closely related to WM ability. Moreover, a survey of WM ability seems necessary. Regarding cognitive ability, a battery of WM tests could provide a more comprehensive scale for measuring information processing ability, which would be a viable predictive intelligence index for personnel assessment and selection. Regarding disease screening, a decline in WM is an early warning sign of mental disorders such as schizophrenia [39], hyperactivity, senile dementia [14, 15, 40], and insomnia [41]. Furthermore, WM capacity test results can be used to guide further cognitive intervention for individuals with certain weak WM components, thereby ensuring that such individuals receive well-directed training on those components. Theoretically, studying the composition and functions of WM as well as the development or decline of cognitive ability is crucial in forming a set of indicators to describe the various abilities of WM.

The test presented in this study can be applied not only in cognitive ability assessments—which can be used in school entrance examinations and personnel selection and assessment procedures—but also for the rapid diagnosis of psychological disorders. The scores for the proposed set of WM tests were standardized to avoid confusion due to different test tasks and test indices, thereby rendering the scores comparable for humans.

Methods

2.1. Participants

In this study, 115 college students (mean age: 22 ± 3.47 years; 62 females) were enrolled in Beijing and Nanjing. They all reported strong unassisted or corrected vision and standard color perception, and reported no history of psychiatric or neurological diseases, neuropsychological problems, or substance abuse. Each subject signed an informed consent form and received a reward after completing all the tests. The study was approved by the Beijing Normal University Ethics Committee.

2.2. Measurements

2.2.1. WM measurements. The test programs were conducted using computers; therefore, the duration of each trial and interval between consecutive trials were designed precisely according to a program statement. The rules of each test were simple, methods of data collection and analysis were convenient, and the test was suitable for group testing. The entire set of tests comprised the following five subtests that correspond to specific WM components: (1) The visuospatial sketchpad function was tested using a visual delayed match-to-sample task,

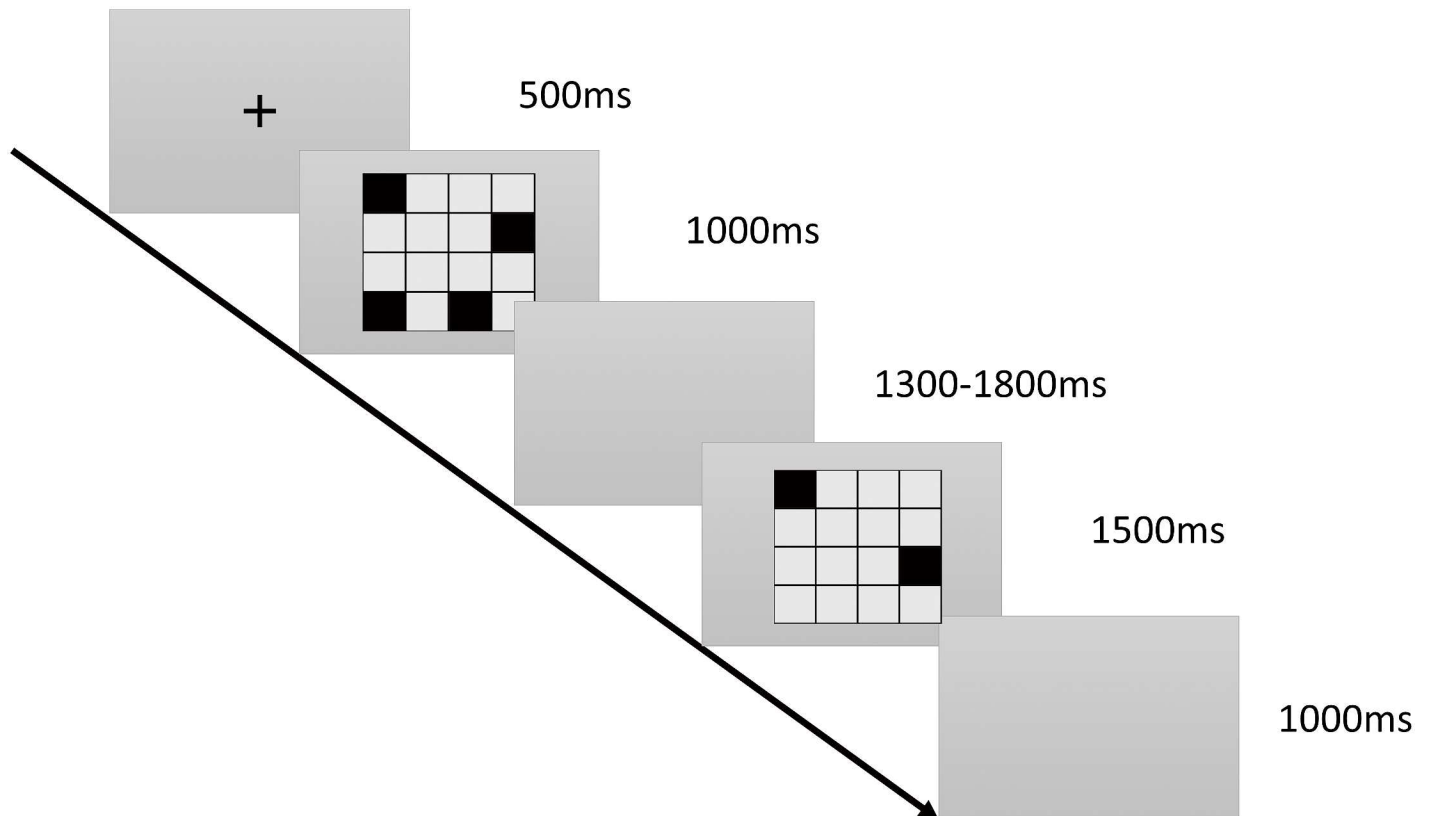


Fig 1. Visuospatial sketchpad subtest. Time course of the stimulus presentation in the visual delayed match-to-sample task. A black “+” symbol is displayed at the center of the screen for 500 ms as a reminder to the participants to pay attention. In the 4×4 grid, 4–9 black squares are displayed in a random order for 1,000 ms each. The participants must remember the locations of the black squares. Subsequently, the black squares disappear, and the screen remains blank for approximately 1,500 ms before the grid appears again with only two black squares. Participants must determine whether these two black squares are located in the same positions as any of the previously displayed 4–9 black squares.

<https://doi.org/10.1371/journal.pone.0175047.g001>

(2) the phonological loop was tested using a letter-delayed match-to-sample task, (3) the switching function was measured through a number-switching determination task, (4) the updating function was measured using a running WM task, and (5) the inhibition function was measured using the numerical Stroop task. These tasks were selected because (1) the preparation, rules, and completion of the tasks were simple, rendering them easy for experimenters and participants to understand; (2) the tasks were considered “pure” for measuring the corresponding components of WM [42], had been used extensively, and had high content validity; and (3) the tasks were appropriate for computer software, and thus were conducive for conducting various assessments and collecting data.

2.2.1.1. Visuospatial sketchpad subtest: In this task, a 4×4 grid is displayed at the center of the screen. Each participant must memorize the locations of 4–9 black squares. Subsequently, the screen is cleared and another two black squares appear on the screen. The participant must determine whether the latter two squares are located in the same position as any of the previously displayed black squares. The more squares there are to be memorized, the more difficult the task is. Participants who complete the task in shorter times and with higher accuracy are considered to have stronger visuospatial sketchpad functions, and thus higher visual WM capacities. The details of the subtest are showed in Fig 1.

2.2.1.2. Phonological loop subtest: In this task, a circle formed by 4–12 uppercase letters is presented at the center of the screen. Each participant must memorize the uppercase letters.

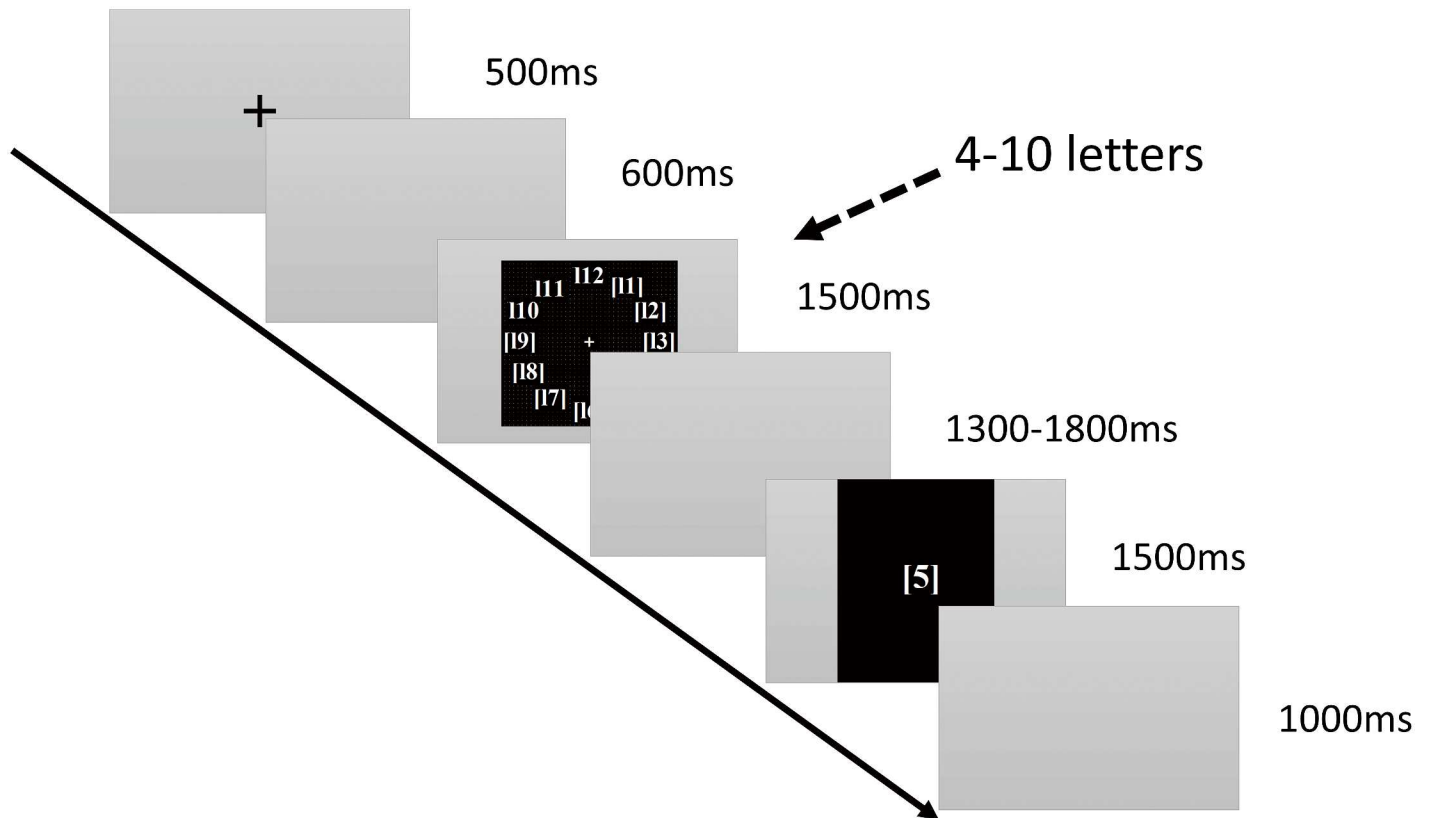


Fig 2. Phonological loop subtest. Time course of the stimulus presentation in the letter delayed match-to-sample task. A black “+” symbol is displayed at the center of the screen for 500 ms as a reminder to the participants to pay attention. Subsequently, a blank screen is displayed for approximately 600 ms before a letter loop appears. In a clock-like loop, 4–12 upper case letters are displayed in a random order for 1,500 ms each. After the letter loop disappears, the screen remains blank for 1,300–1,800 ms before a lowercase letter appears at the center of the screen for a further 1,500 ms. Participants must determine whether this letter was among the letters in the clock-like loop (irrespective of case). Subsequently, another blank screen is displayed for 1,000 ms, after which the task ends.

<https://doi.org/10.1371/journal.pone.0175047.g002>

Subsequently, another lowercase letter is displayed in the screen. The participant must determine whether the same letter was among the previously displayed uppercase letters (irrespective of case). The more uppercase letters there are to be memorized, the more difficult the task is. Participants who complete the task in shorter times and with higher accuracy are considered to have stronger phonological loop function, and thus, higher verbal WM capacities. The details of the subtest are showed in Fig 2.

2.2.1.3. Central executive system: Updating function subtest: In the letter delayed version of the running WM task, a series of letters is sequentially displayed at the center of the screen. Each series contains a different number of letters. Participants must memorize the last three letters that appear. Subsequently, another three letters are displayed on the screen simultaneously. Participants must determine whether the three letters match the three letters they have memorized. The more letters in a series, the more difficult the task is. Participants who complete this task in shorter times and with higher accuracy are considered to have stronger updating functions. The details of the subtest are showed in Fig 3.

2.2.1.4. Central executive system: Inhibition function subtest: The inhibition function can be measured through a number-size Stroop task [4], where two differently sized Arabic numerals are displayed on the screen simultaneously. Each participant must determine which number has the higher value while ignoring their sizes. According to the visual reaction

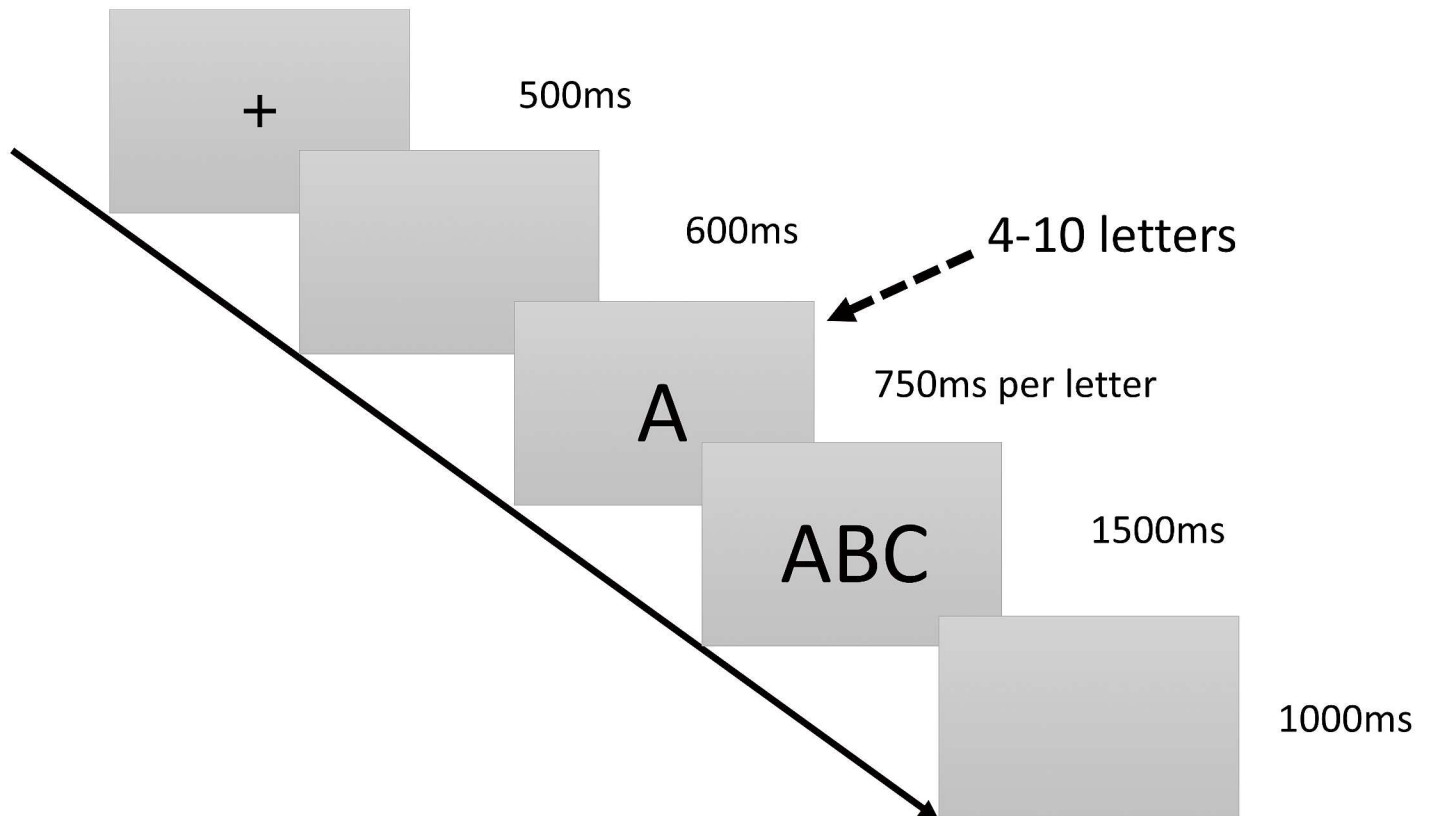


Fig 3. Updating function subtest. Time course of the stimulus presentation in the running WM task. A black “+” symbol is displayed at the center of the screen for 500 ms as a reminder to the participants to pay attention. Subsequently, a blank screen is displayed for approximately 600 ms before a letter series appears. The letter series contains 4–12 uppercase letters that appear sequentially, with each letter being displayed for 750 ms. After the letter series disappears, three of the letters reappear at the center of the screen for approximately 1500 ms. Participants must determine whether these three letters were the last three letters of the previously displayed series. Subsequently, another blank screen is displayed for 1,000 ms, after which the task ends.

<https://doi.org/10.1371/journal.pone.0175047.g003>

characteristics of individuals, when determining two-digit number values, one should inhibit prepotent responses. In this task, the congruent condition is that the value of the larger sized digit is greater. The incongruent condition is that the value of the smaller sized digit is greater, forcing participants to restrain their prepotent responses to make accurate judgments. The neutral condition is that both digits are congruent in size but incongruent in value. A participant’s reaction time difference between the incongruent and neutral conditions was considered as an index of the inhibition function. Participants who complete this task in shorter times and with higher accuracy are considered to have stronger inhibition functions. The details of the subtest are showed in Fig 4.

2.2.1.5. Central executive system: Switching function subtest: This subtest involves two single tasks to determine digit value and parity, and one dual task to determine the values of red digits and parities of blue digits. When the mean reaction time of the dual task differs from that of the single task, the difference is regarded as an index of switching ability. Participants who complete this task in shorter times and with higher accuracy are considered to have stronger switching functions. The details of the subtest are showed in Fig 5.

2.2.2. Fluid intelligence measurement. All participants completed a computerized version of Raven’s Advanced Progressive Matrices (APM) test. The test scores were recorded as reflections of the participants’ fluid intelligence.

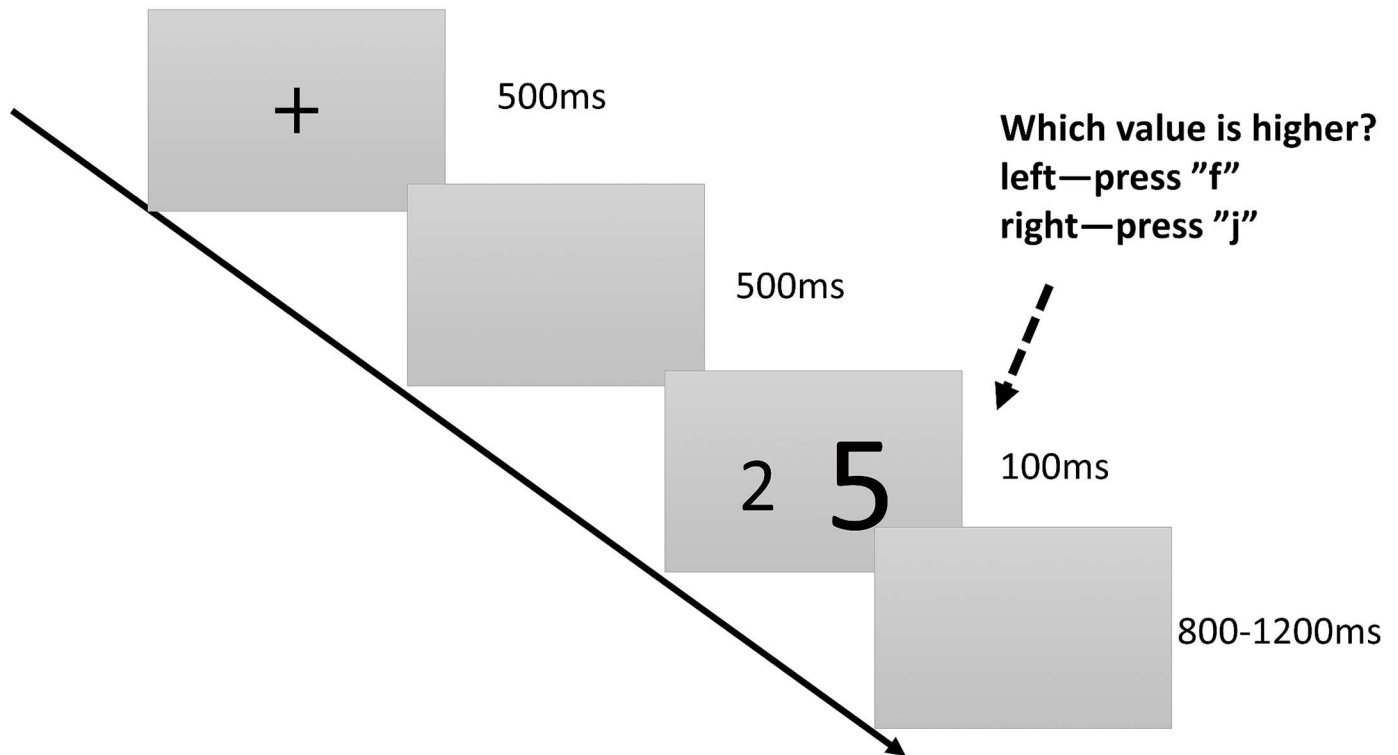


Fig 4. Inhibition function subtest. Time course of the stimulus presentation in the number-size Stroop task. A black “+” symbol is displayed at the center of the screen for 500 ms as a reminder to the participants to pay attention. Subsequently, a blank screen is displayed for approximately 500 ms before two numbers appear for 100 ms each. Participants must determine which number has the higher value while ignoring the digit sizes. After another blank screen is displayed for 800–1,200 ms, the task ends.

<https://doi.org/10.1371/journal.pone.0175047.g004>

2.3. Processing

Because the test set in the present study was based on a computer program, the process was strictly standardized, thereby almost eliminating operational errors, and enabling the reaction times to be recorded more precisely. All of the tests were conducted in standard behavioral laboratories or computer classrooms. The computer screen display resolution was set to 1366 × 768 and the brightness was adjusted to 100%. To avoid the order effect influencing the results, we randomized the order of the five subtests in the WM test battery. All participants undertook the APM test after completing the computerized WM test battery.

Results

3.1. Descriptive statistics

SPSS 22.0 was used to analyze the data. [Table 1](#) provides descriptive statistics for the whole sample.

[Table 2](#) provides correlations among observed variables for the whole sample, showing high correlation among the subtests of the battery assessment.

3.2. Discrimination analysis

The 27% of the participants with highest scores and 27% with the lowest scores were selected as high-score and low-score groups ($n = 31$ in each). An independent sample *t* test was conducted between the two groups, revealing significant differences between the tested groups,

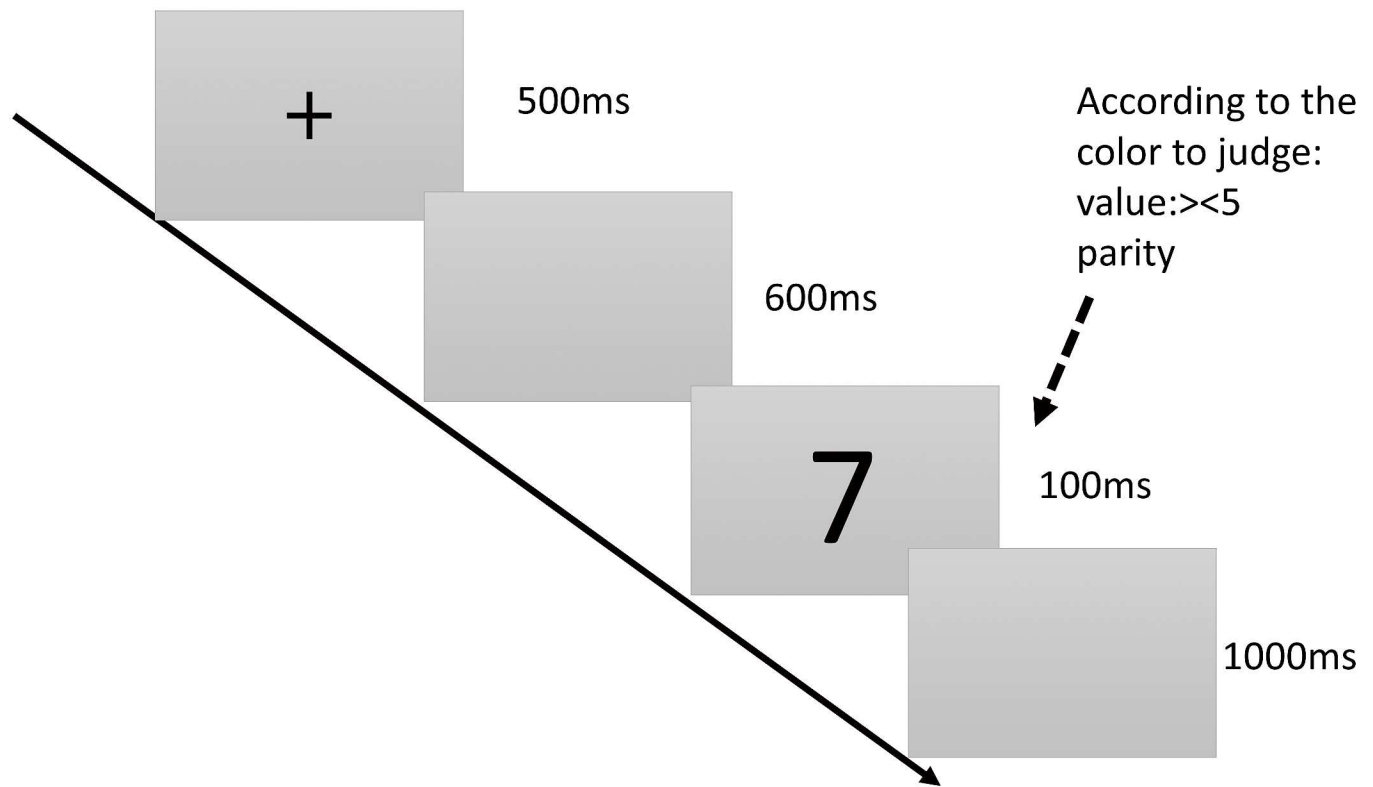


Fig 5. Switching function subtest. Time course of the stimulus presentation in the number-switching task. A black “+” is displayed at the center of the screen for 500 ms as a reminder to the participants to pay attention. Subsequently, a blank screen is displayed for approximately 600 ms before a number appears at the center of the screen. Each number appears for 100 ms. The participants must determine whether the values of the red numbers are higher than 5 and whether the blue numbers are odd or even. Subsequently, another blank screen is displayed for 1,000 ms, after which the task ends.

<https://doi.org/10.1371/journal.pone.0175047.g005>

thereby establishing discrimination in the computerized test battery. The results are shown in Tables 3 and 4.

3.3. Reliability analysis

Across the whole sample, the split-half reliability for accuracy was 0.673, and that for reaction time was 0.614; both were calculated using the Spearman—Brown prediction formula. The

Table 1. Descriptive statistics among observed variables for the whole sample (n = 115).

Task	Group	M ± SD	Skewness	Kurtosis
Visual delayed match-to-sample task (visuospatial sketchpad)	ACC	0.76 ± 0.11	-0.73534	0.669303
	RT	1139.54 ± 148.37	0.179744	-0.25576
Letter-delayed match-to-sample task (phonological loop)	ACC	0.83 ± 0.06	-0.55522	0.070389
	RT	947.61 ± 155.38	0.562865	0.614193
Running WM task (updating function)	ACC	0.93 ± 0.06	-1.78034	4.394886
	RT	78.22 ± 29.17	0.444423	0.346707
Number-size Stroop task (inhibition function)	ACC	0.82 ± 0.14	-2.62499	9.254913
	RT	1035.16 ± 143.64	-0.18702	-0.11504
Number-switching determination task (switching function)	ACC	0.91 ± 0.06	-0.6326	-0.47601
	RT	361.94 ± 142.46	2.69902	16.36699

Note: ACC = accuracy, RT = reaction time.

<https://doi.org/10.1371/journal.pone.0175047.t001>

Table 2. Correlations among observed variables for the whole sample (n = 115).

	1	2	3	4	5	6	7	8	9	10
1 VACC	1									
2 VRT	.186*	1								
3 PACC	.432**	.070	1							
4 PRT	-.004	.587**	-.070	1						
5 IACC	.009	.314**	-.046	.363**	1					
6 IRT	-.115	.003	-.089	.008	.030	1				
7 UACC	.381**	.038	.207*	-.127	.108	-.097	1			
8 URT	-.087	.546**	-.188*	.566**	.236*	.132	-.174	1		
9 SACC	.192*	.218*	.078	.122	.404**	.072	.168	.082	1	
10 SRT	-.065	.066	-.218*	.048	.134	-.072	.009	.165	.251**	1

Note: n = 115; VACC = accuracy of visual delayed match-to-sample task; VRT = reaction time of visual delayed match-to-sample task; PACC = accuracy of letter-delayed match-to-sample task; PRT = reaction time of letter-delayed match-to-sample task; IACC = accuracy of Number-size Stroop task; IRT = reaction time of Number-size Stroop task; UACC = accuracy of running WM task; URT = reaction time of running WM task; SACC = accuracy of number-switching determination task; SRT = reaction time of number-switching determination task;

*P < 0.05,
**P < 0.01.

<https://doi.org/10.1371/journal.pone.0175047.t002>

Cronbach’s alpha of accuracy and reaction time across the whole sample were 0.529 and 0.625, respectively, thereby establishing the internal reliability of the scale.

3.4. Validity analysis

3.4.1. Construct validity. Exploratory factor analysis (EFA) was conducted with varimax rotation and the principal component extraction method, the results of which are shown in Table 5. Only components with eigenvalues of >1 were selected. The scores of the 10 test indicators (five accuracies and five reaction times) revealed the following four factors: storage system reaction time (Factor 1), storage system accuracy (Factor 2), central executive system accuracy (Factor 3), and central executive system reaction time (Factor 4), which cumulatively explained 68% of the variance. Each index loading in its respective dimension ranged from 0.5 to 1.

Table 3. Accuracy differences between high- and low-score groups (%).

Task	Group	M ± SD	t
Visual delayed match-to-sample task (visuospatial sketchpad)	High-score	88.74 ± 2.79	-18.238**
	Low-score	62.31 ± 7.57	
Letter delayed match-to-sample task (phonological loop)	High-score	89.63 ± 2.27	-19.084**
	Low-score	75.20 ± 3.54	
Running WM task (updating function)	High-score	93.98 ± 2.46	-8.807**
	Low-score	65.62 ± 17.76	
Number-size Stroop task (inhibition function)	High-score	98.74 ± 0.92	-12.055**
	Low-score	85.30 ± 6.52	
Number-switching determination task (switching function)	High-score	97.26 ± 1.23	-20.388**
	Low-score	82.38 ± 2.5	

Note: *P < 0.05,
**P < 0.01.

<https://doi.org/10.1371/journal.pone.0175047.t003>

Table 4. Reaction time differences between high- and low-score groups (ms).

Task	Group	M ± SD	t
Visual delayed match-to-sample task (visuospatial sketchpad)	High-score	961 ± 66	20.323**
	Low-score	1330 ± 75	
Letter delayed match-to-sample task (phonological loop)	High-score	770 ± 62	17.259**
	Low-score	1144 ± 104	
Running WM task (updating function)	High-score	861 ± 83	19.108**
	Low-score	1211 ± 59	
Number-size Stroop task (inhibition function)	High-score	44 ± 11	17.943**
	Low-score	115 ± 19	
Number-switching determination task (switching function)	High-score	226 ± 49	9.988**
	Low-score	527 ± 160	

Note: *P < 0.05,
**P < 0.01.

<https://doi.org/10.1371/journal.pone.0175047.t004>

Confirmatory factor analysis (CFA) was employed to evaluate the relative fit of the four-factor WM model by using Mplus 7.0. Table 6 presents the indices of the fitting results between the test sets of the four-factor model and Baddeley’s three-component model of WM. The path model for the four-factor model is shown in Fig 6.

A commonly used index for the goodness of fit of a model is the χ^2 value, which compares the degree to which the predicted covariance in a model differs from the observed covariance. Adequate fitness is usually determined by low, nonsignificant χ^2 values. Because this index is sensitive to the sample size, very large samples (as in the present study) cause even the best-fitting models to frequently yield significant χ^2 values [43]. Therefore, additional global fit indices that are more sensitive to the model specification were used to indicate the fitness of the model. Fit indices such as the comparative fit index (CFI) [44], Bollen’s incremental fit index (IFI) [45], and the root mean square error of approximation (RMSEA) provide further measures of fit by comparing the hypothesized model against a null model in which the relations between the latent variables are not specified and are consequently set at 0. Fit indices with values equal to or higher than 0.90 demonstrate a good fit. In addition, an RMSEA of <0.09 demonstrates a good fit. Table 6 shows that the fit indices met the basic criteria requirements, indicating that the test in the present study exhibited acceptable construct validity.

Table 5. EFA results for the participants’ accuracy and reaction times in the five subtests.

	Factor 1	Factor 2	Factor 3	Factor 4
Phonological loop RT (PRT)	0.86			
Visuospatial sketchpad RT (VRT)	0.83			
Updating RT (URT)	0.80			
Visuospatial sketchpad ACC (VACC)		0.8		
Phonological loop ACC (PACC)		0.77		
Updating ACC (UACC)		0.58		
Switching ACC (SACC)			0.79	
Inhibition ACC (IACC)			0.62	
Switching RT (SRT)			-0.64	-.399
Inhibition RT (IRT)				0.91

Note: ACC = accuracy, RT = reaction time.

<https://doi.org/10.1371/journal.pone.0175047.t005>

Table 6. Fitness of the CFA model.

	χ^2	df	P	RMSEA	CFI	TLI
Model	38.545	27	0.0696	0.061	0.941	0.902

Note: RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker—Lewis Index.

<https://doi.org/10.1371/journal.pone.0175047.t006>

3.4.2. Convergent validity. Regarding the use of the APM test scores as criteria, Table 7 shows high correlations between the storage function test scores and APM scores, indicating that the assessment tools’ calibration validity met the requirements of the basic criteria.

3.5. Common method variance

In this study, we used the Harman one-factor test method to test for common method variance. The results are shown in Table 8.

Table 8 shows that Factor 1 accounted for approximately 25%, which is less than half of the 67% of the cumulative variance accounted for by Factors 1–4. The CFA results verifying the one-factor model are shown in Table 9.

The difference between the two models is significant, with the original model providing a significantly closer fit to the data than the one-factor model. Therefore, common method bias should not affect the interpretation of the results, indicating that the tests have adequate validity and the data is relatively suitable for analysis.

3.6. Norm establishment and fractional conversion

The accuracy index was the integral index of the visuospatial sketchpad, phonological loop, and updating function tests. However, in the actual tests, the reaction times for these three

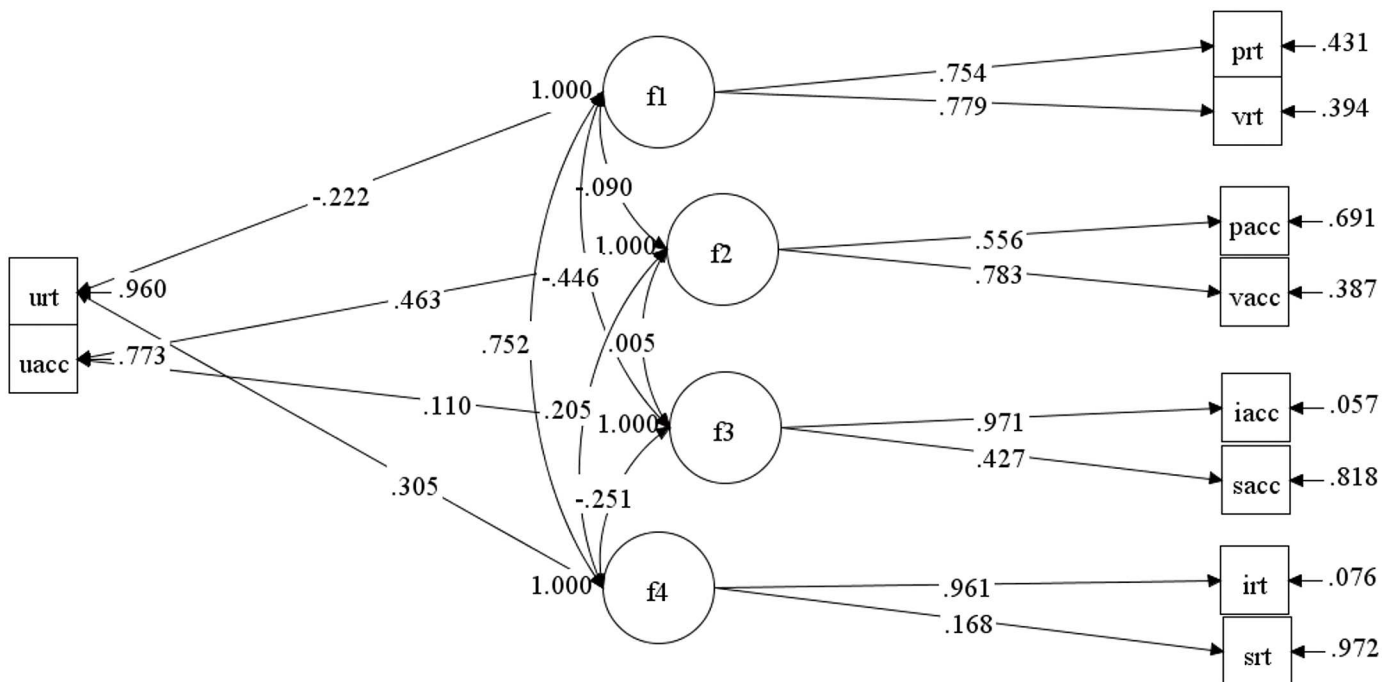


Fig 6. Path model and standardized coefficients of confirmatory factor analysis.

<https://doi.org/10.1371/journal.pone.0175047.g006>

Table 7. Pearson's correlation coefficients for WM measures and Raven's APM accuracy.

Task	<i>r</i>
Visual delayed match-to-sample task (ACC)	0.424**
Visual delayed match-to-sample task (RT)	-0.063
Letter delayed match-to-sample task (ACC)	0.255**
Letter delayed match-to-sample task (RT)	-0.178**
Running WM task (ACC)	0.092
Running WM task (RT)	-0.039
Number-size Stroop task (ACC)	-0.029
Number-size Stroop task (RT)	-0.167
Number-switching determination task (ACC)	0.095
Number-switching determination task (RT)	-0.014

Note: **P* < 0.05,
 ***P* < 0.01,
 ACC = accuracy, RT = reaction time.

<https://doi.org/10.1371/journal.pone.0175047.t007>

tasks were recorded as a secondary index for reference. The inhibition and switching function indices, which were calculated through the subtraction rule, were the absolute values of the reaction time difference during the specific cognitive process. These scores were standardized using their opposite numbers, resulting in proportional changes in the models of reaction time and accuracy. In other words, the greater the fractional conversion value, the stronger the capacity of the project is. However, for these two tasks, the accuracy in the actual tests was recorded as a supplementary reference. In the WM tests, each subtest score was converted into a *z*-score to obtain the participants' overall WM scores. In addition, we determined the corresponding percentile ranks of the overall scores in the established scoring norm. However,

Table 8. Total variance explained.

Component	Initial eigenvalues	Total % of variance	Cumulative %
1	2.512	25.125	25.125
2	1.905	19.046	44.171
3	1.29	12.902	57.073
4	1.041	10.406	67.48
5	0.802	8.021	75.501
6	0.738	7.385	82.885
7	0.508	5.077	87.962
8	0.463	4.632	92.595
9	0.38	3.801	96.396
10	0.36	3.604	100

<https://doi.org/10.1371/journal.pone.0175047.t008>

Table 9. Comparison of fitness indices between the one-factor model and original model.

	χ^2	<i>df</i>	<i>P</i>	RMSEA	CFI	TLI
Original	38.545	27	0.0696	0.061	0.941	0.902
One-factor	186.623	35	<0.001	0.194	0.228	0.007

Note: $\Delta\chi^2 = 148.078 > \Delta\chi^2_{2=0.05}(8) = 15.51$.

<https://doi.org/10.1371/journal.pone.0175047.t009>

participants with identical total scores may have had different subtest scores, indicating individual differences that highlight the participants' strengths and weaknesses in WM.

Discussion

In each test, significant differences were observed between the accuracy and reaction times of the high- and low-score groups, indicating that the test results reveal differences in reaction quality and speed with high discriminatory power.

Because accuracy and reaction time are two distinct index systems with relatively low internal consistency, the Cronbach's α values for the participants' accuracy and reaction times in the five subtests were analyzed. The results were within the acceptable range of 0.673 and 0.614, possibly because of the excessively centralized controlling of the tests' difficulty levels. As shown in [Table 1](#), the score distribution for most tasks is relatively narrow, and the accuracy of the high scores for the inhibition and switching function tasks exhibited a weak ceiling effect.

4.1. Questions regarding the purity of the updating function task

In accordance with the EFA results, the indices were classified into the following four factors: storage function accuracy (Factor 1), storage function reaction time (Factor 2), central executive function accuracy (Factor 3), and central executive function reaction time (Factor 4). However, the switching function task reaction time being classified as Factor 3 might be a result of the EFA's emphasis on the data. According to the factor loading matrix, a loading of approximately 0.4 was found for the switching value in Factor 4, indicating that double loading occurred when EFA was conducted for the reaction time switching function index. This may have been due to the greater dependency of the data or the assumption of the WM component model itself, which requires improvement. In addition, the reliability of the EFA results should be discussed in combination with the CFA results. Nevertheless, the accuracy and reaction time indices of the inhibition function tasks were divided into the appropriate storage systems, possibly because of the data model of the updating function tasks being similar to those of the visuospatial sketchpad and phonological loop tasks. Although Morris and Jones demonstrated that the performance of a running WM task is determined by the central executive function rather than the storage function, the updating function is more closely related to the storage function than the other two central executive functions are [46]. Morris and Jones argued that the updating task is essentially an updating operation involving the assignment of ordinal tags that are stored in the phonological loop of the storage function and must be constantly updated and repeated. By contrast, in the Stroop tasks, only the inhibition of the familiar operation reaction was conducted, of which the targets were mostly reliant on long-term memory. For example, the values and sizes of the numbers in the number-size task were derived from the participants' existing cognition of numbers. Furthermore, the operation targets of the switching task were not stored in the WM storage function. Therefore, the overlap of the updating task necessary for the treatment of the message in the storage functions may actually be unavoidable, which is in agreement with the findings of Ecker and Lewandowsky [47]. The updating function is divided into the following three components: retrieval, transformation, and substitution. Retrieval refers to the retrieval process of updating existing messages, transformation refers to transforming an old retrieved message into a new message, and substitution refers to replacing an old message with a new one. The target of the retrieval process is the message stored in the storage function. Hence, the interpretation of the double loading in the updating function task results is reasonable.

Table 10. Fit conditions of the various models.

	χ^2	df	P	RMSEA	CFI	TLI
M (original model)	38.545	27	0.0696	0.061	0.941	0.902
M1	56.68	29	0.002	0.091	0.859	0.781
M2	36.605	26	0.08	0.06	0.946	0.907

<https://doi.org/10.1371/journal.pone.0175047.t010>

CFA was conducted by combining our results with those of the WM tests and EFA obtained by Baddeley[3] and Miyake[4]. The path model is shown in Fig 6, where f1, f2, f3, and f4 denote Factors 1, 2, 3, and 4, respectively. Based on the EFA results, the double loading between the accuracy and reaction time factors in the central executive function was formed when the switching reaction time was formed, yielding Model M2 (Table 10). Although M2 achieved a good fit, approximating the original model (Model M; $\Delta\chi^2 = 1.94 < \Delta\chi^2_{\alpha=0.05}(1) = 3.84$) by allocating the reaction time index to the accuracy index is meaningless. Hence, under the same fitting conditions, Model M should be used. In addition, Model M1 was obtained by limiting the loading for the updating function to the central executive function based on Baddeley’s three-component model of WM. Table 10 indicates that Model M1 did not achieve a good fit and was inferior to Model M ($\Delta\chi^2 = 18.135 > \Delta\chi^2_{\alpha=0.05}(2) = 5.99$), as demonstrated by the EFA results.

Overall, the test results of this study accord with those of the expected model, with the well-fitting CFA model and adequate construct validity meeting surveying requirements.

4.2. Relationship between WM and fluid intelligence in adults

Raven’s APM test was conducted to measure the fluid intelligence of all the participants. Fluid intelligence was used to verify the criterion validity of the WM battery tests. The results show that only the accuracy of the visuospatial sketchpad function and phonological loop function in the storage function were related to the fluid intelligence test, and that no indices in the central executive function correlated significantly with fluid intelligence. According to numerous related studies with child participants, the development of WM in children is the direct basis of the development of their fluid intelligence. Such a close relationship is likely due to the central executive function[12, 48–50]. Nevertheless, the results of the present study differ considerably from those of previous studies, likely because the present study enrolled adult participants. In a study on calculation strategy tests for children, teenagers, and young adults that involved tracking functional magnetic resonance imaging, Qin showed that older people tend to rely more on the information storage area of the brain (hippocampus) when solving problems, and less on the information processing area of the brain (prefrontal lobe) [51]. After reaching adulthood, an individual’s cognitive strategies and problem-solving mechanisms no longer focus on the quality of information processing, but on the invocation of information stored in a more mature brain. Although the central executive function in adults might not be outstanding, they possess sufficient storage capacity or memory capacity to solve problems after simply processing information. Based on the results obtained from participants who completed the testing tasks in this study, fluid intelligence performance is related to storage function performance. Therefore, the storage function of WM in adults is more closely related to fluid intelligence, possibly because of adults’ greater dependence on cognitive strategies for problem-solving. Similar findings were reported by Swanson in a study involving children with and without dyslexia selected for cognitive strategy training [52]. The results showed that although the participants’ WM capacities improved, their cognitive abilities after training remained limited by their WM capacities. The aforementioned studies have proven that the

more advanced an individual's problem-solving strategies, the greater the role of the storage function of WM in problem-solving and reasoning is.

The reason that no reaction time indicators were related to the APM test scores was that the fluid intelligence test index measured accuracy. According to the previous discussion, the extent of the variation exhibited by the accuracy and reaction time indices precluded their being related.

The test in the present study had the following deficiencies, and could be improved as follows: The number of participants could be increased to expand the national norm, and the duration of the test was too long, causing the participants to become bored and tired. Therefore, the battery tests should be divided into two or three sections with a break between each section to prevent the participants' mental states from influencing the results.

Conclusion

Following Baddeley's three-component model of WM, the present study employed a visual delayed match-to-sample task, letter-delayed match-to-sample task, running WM task, number-size Stroop task, and number-switching determination task to develop a computerized WM test battery for adults. The test battery can measure the following five WM components reflecting various aspects of WM: the visuospatial sketchpad, phonological loop, and central executive function (i.e., inhibition function, switching function, and updating function). The tests had moderate difficulty levels, and the results show that the tests had high discriminatory power, reliability, and validity. The test battery was conducted using computers to ensure that the test process was more standardized and enabling reaction times to be recorded more precisely, thereby guaranteeing the feasibility of large-scale or web-based tests in the future. A standard scoring norm was established to ensure comparability among the test results. The results revealed disadvantages of Baddeley's WM model, thereby laying a foundation for further research on WM capacity development.

Supporting information

S1 File. All the data.

(RAR)

Acknowledgments

This research was funded by the Key Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu (2015JDXM001) and the Joint-Construction Project of the Beijing City Board of Education (403101). We express our gratitude for the support from these projects.

Author Contributions

Conceptualization: LM XC RZ.

Data curation: LM.

Formal analysis: LM.

Funding acquisition: RZ.

Investigation: LM RZ.

Methodology: XC.

Project administration: RZ.

Resources: RZ.

Software: RZ LC.

Supervision: RZ.

Validation: RZ LM.

Visualization: LM LC.

Writing – original draft: LM.

Writing – review & editing: LC RZ.

References

1. Baddeley A. Working memory[J]. *Current Biology*. 2010, 20(4): R136–R140. <https://doi.org/10.1016/j.cub.2009.12.014> PMID: 20178752
2. Baddeley A D, Hitch G. Working memory[J]. *Psychology of learning and motivation*. 1974, 8: 47–89.
3. Baddeley A. The fractionation of working memory[J]. *Proceedings of the national academy of sciences of the united states of America*. 1996, 93(24): 13468–13472. PMID: 8942958
4. Miyake A, Friedman N P, Emerson M J, Witzki A. H., Howerter A., & Wager T. D. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis [J]. *Cognitive Psychology*. 2000, 41(1): 49–100. <https://doi.org/10.1006/cogp.1999.0734> PMID: 10945922
5. Demetriou A, Spanoudis G, Shayer M, Van der Ven S., Brydges C. R., Kroesbergen E. et al. Relations between speed, working memory, and intelligence from preschool to adulthood: Structural equation modeling of 14 studies[J]. *Intelligence*. 2014, 46: 107–121.
6. Conway A R A, Kane M J, Engle R W. Working memory capacity and its relation to general intelligence [J]. *Trends in Cognitive Sciences*. 2003, 7(12): 547–552. PMID: 14643371
7. Fry A F, Hale S. Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade[J]. *Psychological Science*. 1996, 7(4): 237–241.
8. Harrison T L, Shipstead Z, Engle R W. Why is working memory capacity related to matrix reasoning tasks? [J]. *Memory & Cognition*. 2015, 43(3): 389–396.
9. Süß H M, Oberauer K, Wittmann W W, Wilhelm O., & Schulze R. Working-memory capacity explains reasoning ability—and a little bit more. [J]. *Intelligence*. 2002.
10. Kyllonen P C, Christal R E. Reasoning Ability Is (Little More Than) Working-Memory Capacity[J]. *Intelligence*. 1990, 14(4): 389–433.
11. Jaeggi S M, Buschkuhl M, Jonides J, & Perrig W. J. Improving fluid intelligence with training on working memory. [J]. *Proceedings of the National Academy of Sciences*. 2008.
12. Zhao X, Wang Y, Liu D, & Zhou R. Effect of updating training on fluid intelligence in children[J]. *Chinese Science Bulletin*. 2011, 56(21): 2202–2205.
13. Rudebeck S R, Bor D, Ormond A, O'Reilly J. X., & Lee A. C. A Potential Spatial Working Memory Training Task to Improve Both Episodic Memory and Fluid Intelligence[J]. *Plos One*. 2012.
14. Wang Z, Zhou R, Shah P. Spaced cognitive training promotes training transfer[J]. *Frontiers in Human Neuroscience*. 2014.
15. Monsell S E, Mock C, Hassenstab J, Roe C. M., Cairns N. J., Morris J. C., et al. Neuropsychological changes in asymptomatic persons with Alzheimer disease neuropathology. [J]. *Neurology*. 2014.
16. Pillai J A, Bonner-Jackson A, Walker E, Mourany L., & Cummings J. L. Higher Working Memory Predicts Slower Functional Decline in Autopsy-Confirmed Alzheimer's Disease[J]. *Dementia and Geriatric Cognitive Disorders*. 2014, 38(3–4): 224–233. <https://doi.org/10.1159/000362715> PMID: 24903099
17. Schweitzer J B, Faber T L, Grafton S T, Tune L. E., Hoffman J. M., & Kilts C. D. Alterations in the functional anatomy of working memory in adult attention deficit hyperactivity disorder. [J]. *American Journal of Psychiatry*. 2000.
18. Hartman M, Steketee M C, Silva S, Lanning K., & Andersson C. Wisconsin Card Sorting Test performance in schizophrenia: the role of working memory[J]. *Schizophrenia Research*. 2003, 63(3): 201–217. PMID: 12957700

19. Fink A, Neubauer A. Individual differences in time estimation related to cognitive ability, speed of information processing and working memory[J]. *Intelligence*. 2005, 33(1): 5–26.
20. Carretti B, Borella E, Cornoldi C, & De Beni R.. Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis[J]. *Learning and Individual Differences*. 2009, 19(2): 246–251.
21. Frieze M, Bargas-Avila J, Hofmann W, & Wiers R. W.. Here's Looking at You, Bud: Alcohol-Related Memory Structures Predict Eye Movements for Social Drinkers with Low Executive Control[J]. *Social Psychological and Personality Science*. 2010, 1(2): 143–151.
22. Hofmann W, Gschwendner T, Frieze M, Wiers R. W., & Schmitt M.. Working memory capacity and self-regulatory behavior: Toward an individual differences perspective on behavior determination by automatic versus controlled processes[J]. *Journal of personality and social psychology*. 2008, 95(4): 962–977. <https://doi.org/10.1037/a0012705> PMID: 18808271
23. Gathercole S E, Pickering S J. Assessment of working memory in six- and seven-year-old children[J]. *Journal of educational psychology*. 2000, 92(2): 377–390.
24. Hicks K L, Foster J L, Engle R W. Measuring Working Memory Capacity on the Web with the Online Working Memory Lab (the OWL) [J]. *Journal of Applied Research in Memory and Cognition*. 2016, 5(4): 478–489.
25. Redick T S, Broadway J M, Meier M E, Kuriakose P. S., Unsworth N., Kane M. J., et al. Measuring Working Memory Capacity with Automated Complex Span Tasks[J]. *European Journal of Psychological Assessment*. 2012, 28(3): 164–171.
26. van der Donk M, Hiemstra-Beemink A, Tjeenk-Kalff A, Van Der Leij A., & Lindauer R.. Cognitive training for children with ADHD: a randomized controlled trial of cogmed working memory training and 'paying attention in class' [J]. *Frontiers in Psychology*. 2015, 6.
27. Oswald F L, Mcabee S T, Redick T S, & Hambrick D. Z.. The development of a short domain-general measure of working memory capacity[J]. *Behavior Research Methods*. 2015, 47(4): 1343–1355. <https://doi.org/10.3758/s13428-014-0543-2> PMID: 25479734
28. Foster J L, Shipstead Z, Harrison T L, Hicks K. L., Redick T. S., & Engle R. W. Shortened complex span tasks can reliably measure working memory capacity[J]. *Memory & Cognition*. 2015, 43(2): 226–236.
29. Langer N, von Bastian C C, Wirz H, Oberauer K., & Jäncke L. The effects of working memory training on functional brain network efficiency[J]. *Cortex*. 2013, 49(9): 2424–2438. <https://doi.org/10.1016/j.cortex.2013.01.008> PMID: 23489778
30. Siegel L S. Working-memory and reading—a life-span perspective[J]. *International journal of behavioral development*. 1994(SI): 109–124.
31. León-Domínguez U, Martín-Rodríguez J F, León-Carrión J. Executive n-back tasks for the neuropsychological assessment of working memory[J]. *Behavioural Brain Research*. 2015, 292: 167–173. <https://doi.org/10.1016/j.bbr.2015.06.002> PMID: 26068585
32. Alloway T P, Temple K J. A comparison of working memory skills and learning in children with developmental coordination disorder and moderate learning difficulties[J]. *Applied Cognitive Psychology*. 2007, 21(4): 473–487.
33. Wechsler D. Wechsler intelligence scale for children—Fourth Edition (WISC-IV) [J]. San Antonio, TX: The Psychological Corporation. 2003.
34. Alloway T P, Gathercole S E. Working Memory Rating Scale (WMRS): Manual[M]. Pearson Education, 2008.
35. Gioia G A, Isquith P K, Guy S C, & Kenworthy L. Test review behavior rating inventory of executive function[J]. *Child Neuropsychology*. 2000, 6(3): 235–238. <https://doi.org/10.1076/chin.6.3.235.3152>
36. Conners C K. Conners' rating scales revised[M]. Multi-Health Systems, Incorporated, 2001.
37. Alloway T P, Gathercole S E, Kirkwood H, & Elliott J. The working memory rating scale: A classroom-based behavioral assessment of working memory[J]. *Learning and Individual Differences*. 2009, 19(2): 242–245.
38. Berg D H. Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading[J]. *Journal of Experimental Child Psychology*. 2008, 99(4): 288–308. <https://doi.org/10.1016/j.jecp.2007.12.002> PMID: 18242631
39. Glahn D C, Therman S, Manninen M, Huttunen M., Kaprio J., Lönnqvist J., et al. Spatial working memory as an endophenotype for schizophrenia[J]. *Biological Psychiatry*. 2003, 53(7): 624–626. PMID: 12679242
40. Macdonald M C, Almor A, Henderson V W, Kempler D., & Andersen E. S. Assessing Working Memory and Language Comprehension in Alzheimer's Disease[J]. *Brain and Language*. 2001, 78(1): 17–42. <https://doi.org/10.1006/brln.2000.2436> PMID: 11412013

41. Thomas A G, Monahan K C, Lukowski A F, & Cauffman E. Sleep Problems Across Development: A Pathway to Adolescent Risk Taking Through Working Memory[J]. *Journal of Youth and Adolescence*. 2015, 44(2): 447–464. <https://doi.org/10.1007/s10964-014-0179-7> PMID: 25213135
42. Van D, Brédart S, Beerten A. Age-related differences in updating working memory. [J]. *British Journal of Psychology*. 1994.
43. Kline R B, Santor D A. Principles & Practice of Structural Equation Modelling[J]. *Canadian Psychology*. 1999, 40(4): 381.
44. Bentler P M. Comparative fit indexes in structural models[J]. *Psychological bulletin*. 1990, 107(2): 238. PMID: 2320703
45. Bollen Kenneth A. Structural equations with latent variables[J]. New York Wiley. 1989.
46. Morris N, Jones D M. Memory updating in working memory: The role of the central executive[J]. *British journal of psychology*. 1990, 81: 111.
47. Ecker U K H, Lewandowsky S, Oberauer K, & Chee A. E. The components of working memory updating: An experimental decomposition and individual differences. [J]. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2010, 36(1): 170–189. <https://doi.org/10.1037/a0017891> PMID: 20053053
48. Chen T, Li D. The roles of working memory updating and processing speed in mediating age-related differences in fluid intelligence[J]. *Aging Neuropsychology and Cognition*. 2007, 14(6): 631–646.
49. Friedman N P, Miyake A, Corley R P, Young S. E., DeFries J. C., & Hewitt J. K. Not All Executive Functions Are Related to Intelligence[J]. *Psychological science*. 2006, 17(2): 172–179. <https://doi.org/10.1111/j.1467-9280.2006.01681.x> PMID: 16466426
50. Ackerman P L, Beier M E, Boyle M O. Working Memory and Intelligence: The Same or Different Constructs? [J]. *Psychological Bulletin*. 2005, 131(1): 30–60. <https://doi.org/10.1037/0033-2909.131.1.30> PMID: 15631550
51. Qin S, Cho S, Chen T, Rosenberg-Lee M., Geary D. C., & Menon V. Hippocampal-neocortical functional reorganization underlies children's cognitive development[J]. *Nature Neuroscience*. 2014.
52. Swanson H L, Kehler P, Jerman O. Working Memory, Strategy Knowledge, and Strategy Instruction in Children with Reading Disabilities[J]. *Journal of Learning Disabilities*. 2010, 43(1): 24–47. <https://doi.org/10.1177/0022219409338743> PMID: 19749089