ORIGINAL RESEARCH

Reliability and validity of the revised Gibson Test of Cognitive Skills, a computer-based test battery for assessing cognition across the lifespan

Amy Lawson Moore
Terissa M Miller

Gibson Institute of Cognitive Research, Colorado Springs, CO, USA

Purpose: The purpose of the current study is to evaluate the validity and reliability of the revised Gibson Test of Cognitive Skills, a computer-based battery of tests measuring short-term memory, long-term memory, processing speed, logic and reasoning, visual processing, as well as auditory processing and word attack skills.

Methods: This study included 2,737 participants aged 5–85 years. A series of studies was conducted to examine the validity and reliability using the test performance of the entire norming group and several subgroups. The evaluation of the technical properties of the test battery included content validation by subject matter experts, item analysis and coefficient alpha, test–retest reliability, split-half reliability, and analysis of concurrent validity with the Woodcock Johnson III Tests of Cognitive Abilities and Tests of Achievement.

Results: Results indicated strong sources of evidence of validity and reliability for the test, including internal consistency reliability coefficients ranging from 0.87 to 0.98, test–retest reliability coefficients ranging from 0.69 to 0.91, split-half reliability coefficients ranging from 0.87 to 0.91, and concurrent validity coefficients ranging from 0.53 to 0.93.

Conclusion: The Gibson Test of Cognitive Skills-2 is a reliable and valid tool for assessing cognition in the general population across the lifespan.

Keywords: testing, cognitive skills, memory, processing speed, visual processing, auditory processing

Introduction

From the ease of administration and reduction in scoring errors to cost savings and engaging user interfaces, computer-based testing has increased in popularity over recent years for various compelling reasons. Although a recent push has been to utilize digital neurocognitive screening measures to assess postconcussion and age-related cognitive decline,[1,2] the primary uses of existing computer-based measures appear to be focused on either clinical diagnostics or academic skill evaluation. An interesting gap in the literature on computer-based assessment, however, is the dearth of studies evaluating individual cognitive skills for the purpose of researching therapeutic cognitive interventions or cognitive skills' remediation. Given the association of cognitive skills and academic achievement,[3] sports performance,[4] career success,[5] and psychopathology-related deficits,[6] it is reasonable to suggest the need for an affordable, easy to administer test that identifies deficits in cognitive skills in order to recommend an intervention for addressing these deficits.

Traditional neuropsychological testing is lengthy and costly and requires advanced training in assessment to score and utilize the results. In clinical diagnostics, cognitive

Correspondence: Amy Lawson Moore
Gibson Institute of Cognitive Research, 5085 List Drive, Suite 220, Colorado Springs, CO 80919, USA
Tel +1 719 219 0940
Fax +1 719 522 0434
Email amoore@gibsonresearch.org

skills are assessed as a part of traditional intelligence or neuropsychological testing, typically to rule out intellectual disability as a primary diagnosis. However, there are additional critical uses for cognitive testing as well. As clinicians or researchers, if we adopt the perspective that cognitive skills underlie learning and behavior, we necessarily must seek an efficient, affordable, yet psychometrically sound method of evaluating cognitive skills in order to suggest existing interventions or to examine new methods of remediating cognitive deficits.

Despite the growing availability of commercially available cognitive tests, there are notable gaps in the field. For example, digital cognitive tests typically do not include measures of auditory processing skills. School-based screening of these early reading skills dominates the digital achievement testing marketplace but is not traditionally found in digital cognitive tests. Not only do auditory processing skills serve as the foundation for reading ability and language development in childhood,[7] they also impact receptive and written language functioning in which these deficits are associated with higher rates of unemployment and lower income[8] and influence the trajectory of lifespan decline in auditory perceptual abilities frequently misattributed to age-related hearing loss.[9] Auditory processing is a key component of the Cattell–Horn–Carroll (CHC) model of intelligence,[10] which serves as the theoretical grounding for most major intelligence tests. As such, it would be a valuable measure on digital cognitive test batteries. Furthermore, a cross-battery approach to assessment aligns with contemporary testing practice that approximates a measurement of multiple constructs succinctly and more efficiently than through the use of individual cognitive and achievement batteries.[11] Another notable shortcoming of traditional cognitive tests is the use of numerical stimuli to assess memory in children with learning struggles, given the reliance on numerical

processing to perform standard digit span memory tasks.[12] Instead, digital cognitive tests should offer non-numerical stimuli in assessing children to ensure that a true measure of memory is captured.

The current study addresses these gaps in the literature on digital cognitive testing. The Gibson Test of Cognitive Skills – Version 2[13] is a computer-based screening tool that evaluates performance on tasks that measure 1) short-term memory, 2) long-term memory, 3) processing speed, 4) auditory processing, 5) visual processing, 6) logic and reasoning, and 7) word attack skills. The 45-minute assessment includes nine different tasks organized as puzzles and games. The development of the new Gibson Test of Cognitive Skills fills a gap in the existing testing market by offering a digital battery that measures seven broad constructs outlined by CHC theory including three narrow abilities in auditory processing and a basic reading skill, word attack. The inclusion of auditory processing subtests is a unique and critical contribution to the digital assessment market. To the best of our knowledge, it is the only digital cognitive test that measures auditory processing and basic reading skills in addition to five other broad cognitive constructs.[14] The Gibson Test also addresses the inadequacy of using numerical stimuli in assessing memory in children by using a variety of visual and auditory stimuli to drill down on short-term memory span and delayed retrieval of meaningful associations. With the exception of the US military's ANAM test,[15] the Gibson Test boasts the largest normative database among major commercially available digital cognitive tests and is the largest normative database among those tests that include children. Table 1 illustrates the comparison of the Gibson Test and other commercially available digital cognitive tests in the measurement of cognitive constructs and norming sample size.

Prior versions of the Gibson Test of Cognitive Skills have been used in several research studies[22–24] and by

Table 1 Comparison of Gibson Test and other major digital cognitive tests

Digital cognitive test	STM	LTM	VP	PS	LR	AP	WA	Norming sample	Norming group ages (years)
Gibson Test of Cognitive Skills-V2[13]	X	X	X	X	X	X	X	2,737	5–85
NeuroTrax[16]	X	X	X	X	X			1,569	8–120
MicroCog[17]	X		X	X	X			810	18–89
ImPACT[18]	X			X				931	13 to college
CNS Vital Signs[19]	X	X	X	X				1,069	7–90
CANS-MCI[20]	X			X				310	51–93
ANAM[15]	X	X	X	X	X			107,801	17–65
CANTAB[21]	X		X	X	X			2,000	4–90

Abbreviations: ANAM, Automated Neuropsychological Assessment Metrics; AP, auditory processing; CANTAB, Cambridge Neuropsychological Test Automated Battery; CANS-MCI, Computer-Administered Neuropsychological Screen; LR, logic and reasoning; LTM, long-term memory; PS, processing speed; STM, short-term memory; VP, visual processing; WA, word attack.

clinicians since 2002. It was initially developed as a progress monitoring tool for cognitive training and visual therapy interventions. Although the evidence of validity supporting the original version of the test is strong,[25,26] recognition that a lengthier test would increase the reliability of cognitive construct measurement served as the primary impetus to initiate a revision. The secondary impetus was the need to add a long-term memory measure.

Methods

A series of studies was conducted to examine the validity and reliability of the Gibson Test of Cognitive Skills (Version 2) using the test performance of the entire norming group and several subgroups. The evaluation of the technical properties of the test battery included content validation by subject matter experts, item analysis and coefficient alpha, test–retest reliability, split-half reliability, and analysis of concurrent validity with the Woodcock Johnson III (WJ III) Tests of Cognitive Abilities and Tests of Achievement. The development and validation process for the revised Gibson Test of Cognitive Skills aligned with the *Standards for Educational and Psychological Testing*.[27] Ethics approval to conduct the study was granted by the Institutional Review Board (IRB) of the Gibson Institute of Cognitive Research prior to recruiting participants. During development, subject matter experts in educational psychology, experimental psychology, special education, school counseling, neuropsychology, and neuroscience were consulted to ensure that the content of each test adequately represented the skill it aimed to measure. A formal content validation review by three experts was conducted prior to

field testing. Data collection began following the content review.

Measures

The Gibson Test of Cognitive Skills (Version 2) battery contains nine tests leading to the measurement of seven broad cognitive constructs. The battery is designed to be administered in its entirety to ensure proper timing for the long-term memory assessment. Each test in the battery is designed to measure at least one aspect of seven broad constructs explicated in the Cattell–Horn–Carroll model of intelligence:[28] fluid reasoning (Gf), short-term memory (Gsm), long-term storage and retrieval (Glr), processing speed (Gs), visual processing (Gv), auditory processing (Ga), and reading and writing (Grw).

Long-term memory test

The test for long-term memory is presented in two parts. It measures the CHC construct of meaningful memory, under the broad CHC construct of long-term storage and retrieval (Glm). The test is given in two parts. At the beginning of the test battery, the examinee sees a collection of visual scenes and short auditory scenarios. After studying the prompts, the examinee responds to questions about them. After the examinee finishes the remaining battery of tests, the original long-term memory test questions are revisited but without the visual and auditory prompts. The test is scored for accuracy and for consistency between answers given during the initial prompted task and the final nonprompted task. There are 24 questions on this test for a total of 48 possible points. An example of a visual prompt is shown in Figure 1.

Figure 1 Example of a long-term memory test visual prompt.

Visual processing test

The visual processing test measures visualization, or the ability to mentally manipulate objects. Visualization is a narrow skill under the broad construct of visual processing (Gv). The examinee is shown a complete puzzle on one side of the screen and individual pieces on the other side of the screen. As each part of the puzzle is highlighted, the examinee must select the corresponding piece that best matches a highlighted part of the puzzle. There are 14 puzzles on the test for a total of 92 possible points. An example of one visual processing puzzle is shown in Figure 2.

Logic and reasoning test

The logic and reasoning test measures inductive reasoning, or induction, which is the ability to infer underlying rules from a given problem. This ability falls under the broad CHC construct of fluid reasoning (Gf). The test uses a matrix reasoning task where the examinee is given an array of images from which to determine the rule that dictates the missing image. There are 29 matrices for a possible total of 29 points (Figure 3).

Processing speed test

The processing speed (Gs) test measures perceptual speed, or the ability to quickly and accurately search for and compare visual images or patterns presented simultaneously. The examinee is shown an array of images and must identify a matching pair in each array (Figure 4) in the time allotted. There are 55 items for a total of 55 possible points.

Short-term memory test

The short-term memory test measures visual memory span, a component of the broad construct of short-term memory (Gsm). In CHC theory, memory span is the ability to "encode information, maintain it in primary memory, and immediately reproduce the information in the same sequence in which it was represented".[28] The examinee studies a pattern of shapes on a grid and then reproduces the pattern from memory when the visual prompt is removed (Figure 5). The patterns become more difficult as the test progresses. There are 21 patterns for a total of 63 possible points.

Auditory processing test

The auditory processing test measures the following three features of the broad CHC construct auditory processing (Ga): phonetic coding analysis, phonetic coding synthesis, and sound awareness. A sound blending task measures phonetic coding – synthesis, or the ability to blend smaller units of speech into a larger one. The examinee listens to the individual sounds in a nonsense word and then must blend the sounds to identify the completed word. For example, the narrator says, "/n/-/e/-/f/". The examinee then sees and selects from four choices on the screen (Figure 6).

There are 15 sound blending items on the test. A sound segmenting task measures phonetic coding analysis, or the ability to segment larger units of speech into smaller ones. The examinee listens to a nonsense word and then must separate the individual sounds. There are 15 sound segmenting items on this subtest. Finally, a sound dropping
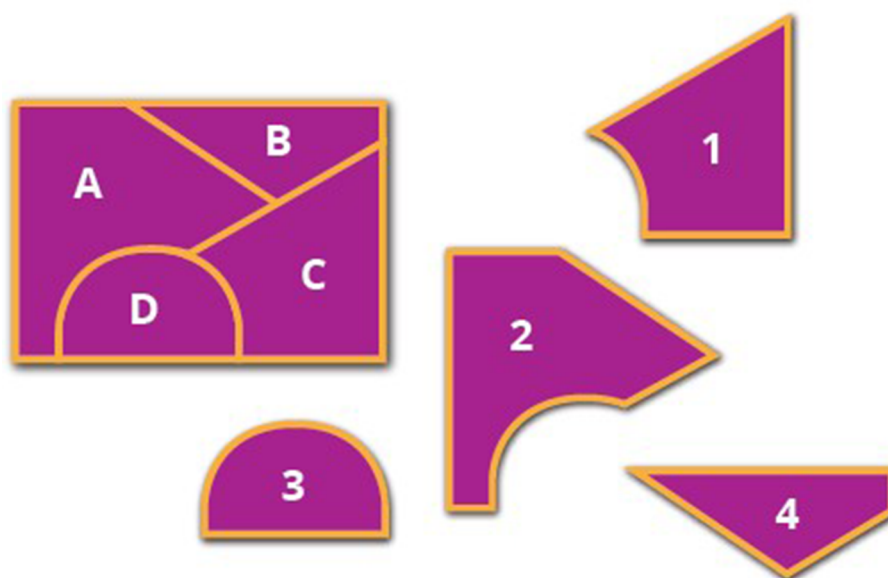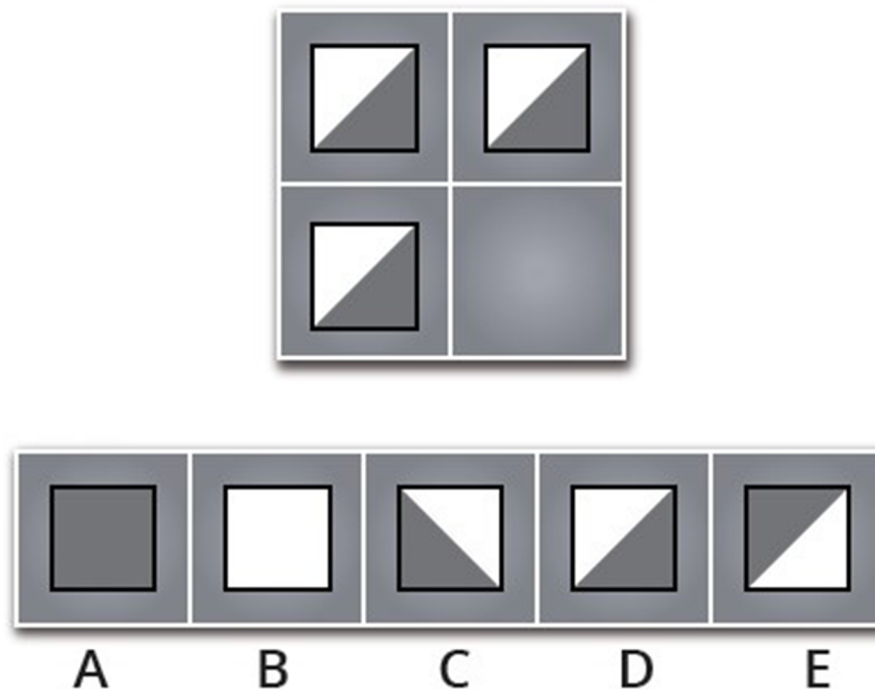

Figure 2 Example of a visual processing test item.

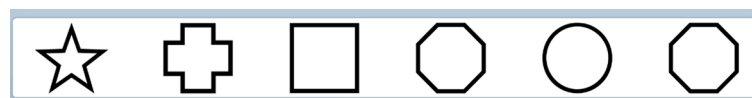**Figure 3** Example of a logic and reasoning test item.

Figure 4 Example of a processing speed test item.

task measures sound awareness. The examinee listens to a nonsense word and is told to delete part of the word to form a new word. The examinee must mentally drop the sounds and identify the new word. There are 15 sound dropping items on this subtest. The complete auditory processing test comprised 45 items for a total of 72 possible points.

Word attack test

The word attack test measures reading decoding ability, or the skill of reading phonetically irregular words or nonsense words. The measure falls under the broad CHC construct of reading and writing (Grw). The examinee listens to the narrator say four nonsense words aloud. Then, the examinee selects from a set of four options of how the nonsense word should be spelled. For example, the narrator says, say "upt". The test taker then sees and selects from four choices on the screen (Figure 7). There are 25 nonsense words for a total of 55 possible points.

Sample and procedures

The sample (n=2,737) consisted of 1,920 children aged 5–17 years (*M*=11.4, standard deviation [SD] =2.7) and 817

adults aged 18–85 years (*M*=41.7, SD =15.4) in 45 states. The child sample was 50.1% female and 49.9% male. The adult sample was 76.4% female and 23.6% male. Overall, the ethnicity of the sample was 68% Caucasian, 13% African-American, 11% Hispanic, and 3% Asian with the remaining 5% of mixed or other race. Detailed demographics are available in the technical manual.[29] Norming sites were selected based on representation from the following four primary geographic regions of the USA and Canada: Northeast, South, Midwest, and West. Tests were administered in three types of settings over 9 months between 2014 and 2015. In the first phase, test results were collected from clients in seven learning centers around the country. With written parental consent, clients (n=42) were administered the Gibson Test of Cognitive Skills and the WJ III Tests of Cognitive Abilities and Tests of Achievement to assess concurrent validity. The WJ III was selected because it is a test battery that is also grounded in the CHC model of intelligence and included multiple auditory processing and word attack subtests by which we could effectively compare the Gibson Test. It is also a widely accepted comprehensive cognitive test battery, which would strengthen
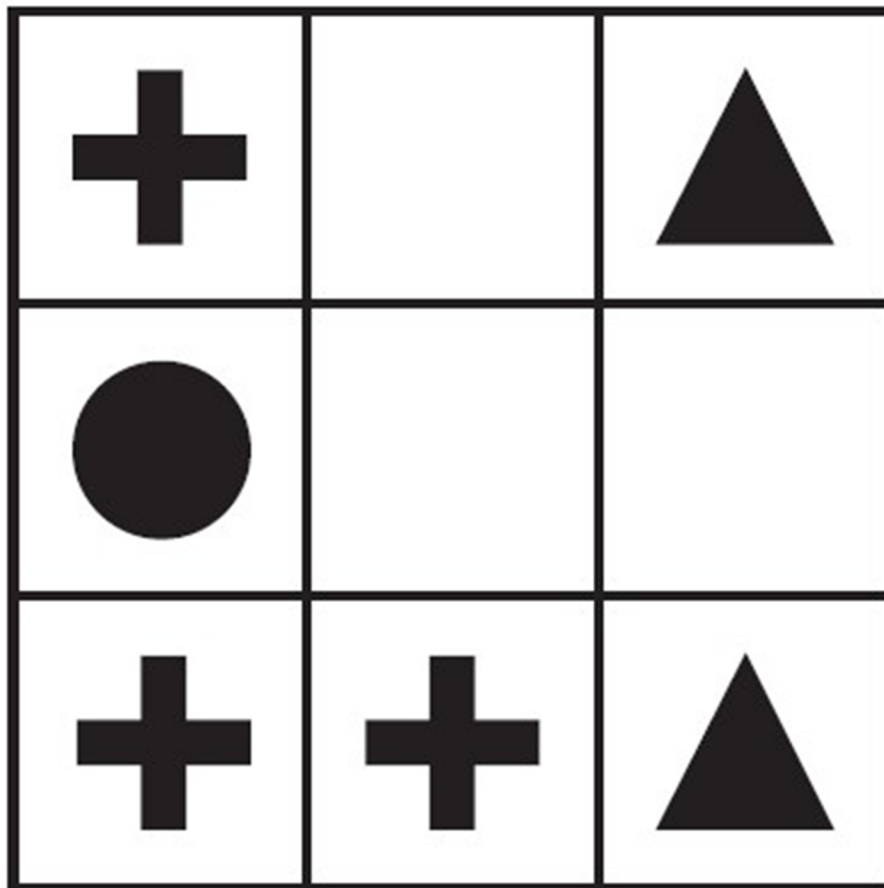
Figure 5 Example of a short-term memory test item.

Figure 6 Example of an auditory processing test item.

Figure 7 Example of a word attack test item.

the trust factor in concurrent validation. The sample ranged in age from 8 to 59 years (*M*=19.8, SD =11.1) and was 52% female and 48% male. Ethnicity of the sample was 74% Caucasian, 10% African-American, 5% Asian, 2% Hispanic, and the remaining 10% mixed or other race.

After obtaining written informed consent, an additional sample of clients (n=50) between the ages of 6 and 58 years (*M*=20.1, SD =11.6) was administered the Gibson Test of Cognitive Skills once and then again after 2 weeks to participate in test–retest reliability analysis. The sample was 55% female and 45% male, with 76% Caucasian, 8% African-American, 4% Asian, 4% Hispanic, and the remaining 8% mixed or other race.

In the second phase of the study, the test was administered to students and staff members in 23 different elementary and high schools. Schools were invited to participate via email from the researchers. Parents of participants in schools were given a letter with a comprehensive description of the norming study with an opt-out form to be returned to the school if they did not want their child to participate. Students could decline participation and quit taking the test at any point. The schools provided all de-identified demographic information to the researchers. Finally, adults and children responded via social media to complete the test from a home computer. Participants and parents who responded via social media provided digital informed consent by clicking through the test after reading the consent document. A demographic information survey was completed by participants along with the test. Participants could quit taking the test at any time, if desired. None of the participants were compensated for participating. The results of this phase of the study were used to calculate internal consistency reliability for each test, split-half reliability for each test, and inter-test correlations and to create the normative database of scores.

Data analysis
Data were analyzed using Statistical Package for the Social Sciences (SPSS) for Windows Version 22.0 and jMetrik software programs. First, we ran descriptive statistics to determine mean scores and all demographics. We conducted item analyses to determine internal consistency reliability with a coefficient alpha for each test. We ran Pearson's correlations to determine split-half reliability, test–retest reliability, and concurrent validity with other criterion measures. Finally, we examined differences by gender and education level.

Results
Descriptive statistics are presented in Table 2 to illustrate the mean scores, SD, and 95% confidence intervals for each

test by age interval. Table 2 illustrates the intercorrelations among all of the tests. Auditory processing and word attack show stronger intercorrelations than with other measures because they measure similar constructs. Visual processing is correlated with logic and reasoning and short-term memory presumably because they are tasks that require the manipulation or identification of visual images. Long-term memory is more correlated with short-term memory than with any other task. These intercorrelations among the tests provide general evidence of convergent and discriminant internal structure validity.

Internal consistency reliability
Item analyses for each test revealed strong indices of internal consistency reliability, or how well the test items correlate with each other. Overall coefficient alphas range from 0.87 to 0.98. Overall coefficient alphas for each test as well as coefficient alphas based on age intervals are all robust (Table 3), indicating strong internal consistency reliability of the Gibson Test of Cognitive Skills.

Split-half reliability
Reliability of the Gibson Test of Cognitive Skills battery was also evaluated by correlating the scores on two halves of each test. Split-half reliability was calculated for each test by correlating the sum of the even numbered items with the sum of the odd numbered items. Then, a Spearman–Brown formula was applied to the Pearson's correlation for each subtest to correct for length effects. Because split-half correlation is not an appropriate analysis for a speeded test, the alternative calculation for the processing speed test was based on the formula: $r_{11}=1-(\text{SEM}^2/\text{SD}^2)$. Overall and subgroup split-half reliability coefficients are robust, ranging from 0.89 to 0.97 (Table 3), indicating strong evidence of reliability of the Gibson Test of Cognitive Skills (Table 4).

Test–retest reliability (delayed administration)
Reliability of each test in the battery was evaluated by correlating the scores on two different administrations of the test to the same sample of test takers 2 weeks apart. The overall test–retest reliability coefficients ranged from 0.69 to 0.91 (Table 5). The results indicate strong evidence of reliability. All overall coefficients were significant at *P*<0.001; and all subgroup coefficients were significant at *P*<0.001 except for long-term memory in adults, which was significant at *P*<0.004.

Table 2 Mean and SD (95% CI) for Gibson Test scales by age group

| Test | Statistics | Age (years) | | | | | | Overall |
		6–8	9–12	13–18	19–30	31–54	55+	
Long-term memory	n	392	943	545	204	379	156	2,619
	M	15.9	21.5	26.1	30.2	25.3	22.5	23.2
	SD	10.3	11.5	11.8	11.2	10.4	12.2	12.2
	95% CI	±1.0	±0.73	±1.5	±1.5	±1.0	±1.9	±0.46
Short-term memory	n	352	811	297	128	348	145	2,081
	M	27.4	37.9	43.7	48.7	45.1	38.7	38.9
	SD	11.2	9.2	9.3	11.4	9.2	9.3	11.5
	95% CI	±1.2	±0.63	±1.0	±1.9	±.96	±1.5	±0.49
Visual processing	n	373	835	308	155	400	166	2,237
	M	17.5	29.4	37.8	54.2	42.7	33.9	33.0
	SD	12.6	14.9	18.0	19.9	20.5	18.5	19.4
	95% CI	±1.3	±1.0	±2.0	±3.1	±2.0	±2.8	±0.80
Auditory processing	n	382	840	314	159	408	162	2,265
	M	27.3	38.3	47.5	57.1	54.9	48.9	42.8
	SD	18.8	20.0	19.2	18.1	18.3	18.5	21.5
	95% CI	±1.9	±1.3	±1.9	±2.8	±1.8	±2.8	±0.88
Logic and reasoning	n	365	822	301	129	354	151	2,122
	M	9.3	13.2	15.3	18.8	17.2	14.8	14
	SD	3.9	3.9	3.9	3.4	3.5	3.5	4.7
	95% CI	±0.40	±0.26	±0.44	±0.58	±0.50	±0.96	±0.27
Processing speed	n	362	819	301	123	353	155	2,115
	M	14.3	30.4	35.1	39.4	36.6	33.5	32.2
	SD	1.9	5.0	5.5	4.9	4.8	6.1	6.3
	95% CI	±0.19	±0.34	±0.62	±0.86	±0.50	±0.96	±0.26
Word attack	n	346	806	295	125	349	145	2,066
	M	24.6	35.3	41.6	46.9	46.2	44.3	37.6
	SD	14.7	13.6	10.9	7.8	8.8	9.2	14.2
	95% CI	±1.5	±0.48	±0.63	±0.69	±0.47	±0.76	±0.61

Abbreviations: CI, confidence interval; n, number in sample; M, mean score. SD, standard deviation.

Table 3 Coefficient alpha for each test in the Gibson Test of Cognitive Skills battery

| Test | Age (years) | | | | | | Overall |
	6–8	9–12	13–18	19–30	31–54	55+	
Long-term memory	0.91	0.92	0.93	0.92	0.91	0.93	0.93
Short-term memory	0.87	0.82	0.83	0.90	0.82	0.83	0.88
Visual processing	0.96	0.96	0.97	0.98	0.98	0.97	0.98
Auditory processing	0.95	0.95	0.95	0.96	0.95	0.95	0.96
Logic and reasoning	0.85	0.83	0.81	0.74	0.77	0.79	0.87
Processing speed	0.88	0.81	0.87	0.87	0.87	0.91	0.88
Word attack	0.93	0.92	0.89	0.83	0.86	0.85	0.93

Concurrent validity

Validity was assessed by running a Pearson's product-moment correlation to examine if each test on the Gibson Test was correlated with other tests measuring similar constructs to determine concurrent validity with other criterion measures. Correlation coefficients were attenuated based on reliability coefficients of the individual criterion tests and corrected for possible range effects using the formula $r_{xy}/\sqrt{(r_{xx} \times r_{yy})}$,

where r_{xy} is the concurrent correlation coefficient, r_{xx} is the test–retest coefficient of each WJ III subtest, and r_{yy} is the test–retest coefficient of each Gibson Test subtest. The resulting correlations range from 0.53 to 0.93, indicating moderate-to-strong relationships between the Gibson Test and other standardized criterion tests (Table 6). All correlations are significant at an alpha of $P<0.001$, indicating strong evidence of concurrent validity.

Table 4 Split-half reliability of the Gibson Test of Cognitive Skills

Test	Age (years)						Overall
	6–8	9–12	13–18	19–30	31–54	55+	
Long-term memory	0.95	0.94	0.95	0.93	0.94	0.95	0.95
Short-term memory	0.90	0.84	0.86	0.92	0.84	0.83	0.90
Visual processing	0.97	0.98	0.98	0.98	0.99	0.99	0.99
Auditory processing	0.97	0.97	0.96	0.97	0.96	0.96	0.97
Logic and reasoning	0.90	0.86	0.86	0.80	0.81	0.86	0.90
Processing speed[a]	0.88	0.81	0.87	0.88	0.88	0.91	0.89
Word attack	0.94	0.94	0.90	0.89	0.90	0.85	0.94

Note: [a]Split-half correlation is not an appropriate analysis for a speeded test; the alternative calculation was based on the formula: $r_{11}=1-(SEM^2/SD^2)$.
Abbreviations: r, reliability; SD, standard deviation; SEM, standard error of measurement.

Table 5 Test–retest reliability of the revised Gibson Test of Cognitive Skills

Test	Child (n=29)	Adult (n=21)	Overall (n=50)
Long-term memory	0.53	0.67	0.69
Short-term memory	0.76	0.75	0.82
Visual processing	0.89	0.74	0.90
Auditory processing	0.88	0.77	0.91
Logic and reasoning	0.84	0.66	0.82
Processing speed	0.83	0.76	0.73
Word attack	0.89	0.68	0.90

Table 6 Correlations between the Gibson Test and the Woodcock Johnson tests

Gibson Test	Woodcock Johnson III	r_{uc}	r_c
Short-term memory	Numbers reversed	0.71	0.84
Logic and reasoning	Concept formation	0.71	0.77
Processing speed	Visual matching	0.50	0.60
Visual processing	Spatial relations	0.70	0.82
Long-term memory	Visual auditory learning	0.43	0.53
Word attack	Word attack	0.82	0.93
Auditory processing	Spelling of sounds	0.75	0.90
	Sound awareness	0.70	0.82

Notes: r_{uc}, uncorrected correlation calculated with Pearson's product momentum of z scores; r_c, corrected correlation using $r_{xy}/\sqrt{(r_{xx} \times r_{yy})}$.

Post hoc analyses of demographic differences

Because the male-to-female ratio of participants in the adult sample was disproportionate to the population, we examined differences by gender in every adult age range on each subtest through linear regression analyses. After Bonferroni correction for 35 comparisons, gender proved to be a significant predictor of score differences in only five of the comparisons. In the 18–29 years age group, gender was a significant predictor of score differences on the test of auditory processing, $P<0.001$, $R^2=0.02$, $B=-5.8$. That is, females outperformed males by 5.8 points on auditory

processing in the 20–29 years age group. In the 40–49 years age group, gender was a significant predictor of score differences on the test of auditory processing, $P<0.001$, $R^2=0.16$, $B=-17.8$; on the test of visual processing, $P<0.001$, $R^2=0.12$, $B=-17.3$; and on the test of word attack, $P<0.001$, $R^2=0.07$, $B=-5.7$; that is, females outperformed males by 17.8 points on auditory processing, by 17.3 points on visual processing, and by 5.7 points on word attack in the 40–49 years age group. In the >60 years age group, gender was a significant predictor of score differences on the test of processing speed, $P<0.001$, $R^2=0.27$, $B=-5.1$; that is, females outperformed males by 5.1 points on processing speed in the >60 years age group. Gender was not a statistically significant predictor of differences between males and females on any subtest in the 30–39 years or 50–59 years age groups.

In addition, the proportion of adults with higher education degrees in the sample was higher than in the population. Therefore, we also examined education level as a predictor of differences in scores on each subtest. Although effect sizes were small, education level was a significant predictor of score differences on all subtests except for processing speed. On working memory, for every increase in educational level, there was a 1.2-point increase in score ($P=0.001$, $R^2=0.01$, $B=1.2$). On visual processing, for every increase in educational level, there was a 3-point increase in score ($P<0.001$, $R^2=0.02$, $B=3.0$). On logic and reasoning, for every increase in educational level, there was a 0.9-point increase in score ($P<0.001$, $R^2=0.01$, $B=1.2$). On word attack, for every increase in educational level, there was a 1.6-point increase in score ($P<0.001$, $R^2=0.03$, $B=1.6$). On auditory processing, for every increase in educational level, there was a 4.4-point increase in score ($P<0.001$, $R^2=0.07$, $B=4.4$). Finally, on long-term memory, for every increase in educational level, there was a 1.6-point increase in score ($P=0.001$, $R^2=0.02$, $B=1.6$).

Discussion and conclusion

The current study evaluated the psychometric properties of the revised Gibson Test of Cognitive Skills. The results indicate that the Gibson Test is a valid and reliable measure of assessing cognitive skills in the general population. It can be used for the assessment of individual cognitive skills to obtain a baseline of functioning in individuals across the lifespan. In comparison with existing computer-based tests of cognitive skills, the overall test–retest reliabilities (0.69–0.91) of the Gibson Test battery are impressive. For example, the test–retest reliabilities range from 0.17 to 0.86 for the Cambridge Neuropsychological Test Automated Battery (CANTAB),[21] from 0.38 to 0.77 for the Computer-Administered Neuropsychological Screen (CANS-MCI),[20] and from 0.31 to 0.86 for CNS Vital Signs.[19] In addition to strong split-half reliability metrics, evidence for the internal consistency reliability of the Gibson Test of Cognitive Skills is also strong, with coefficient alphas ranging from 0.87 to 0.98. The convergent validity of the tests provides a key source of evidence that the test is a valid measure of cognitive skills represented by constructs identified in the CHC model of intelligence.

The ease with which the test can be administered coupled with automated scoring and reporting is a key strength of the battery. The implications for use are encouraging for a variety of fields including psychology, neuroscience, and cognition research as well as all aspects of education. The battery covers a wide range of cognitive skills that are of interest across multiple disciplines. It is indeed exciting to have an automated, cross-battery assessment that includes not only long-term and short-term memories, processing speed, fluid reasoning, and visual processing but also three aspects of auditory processing along with basic reading skills. The norming group traverses the lifespan, making the test suitable for use with all ages. This, too, is a key strength of the current study. With growing emphasis on age-related cognitive decline, the test may serve as a useful adjunct to brain care and intervention with an aging population. Equally useful with children, the test can continue to serve as a valuable screening tool in the evaluation of cognitive deficits that might contribute to learning struggles and, therefore, help inform intervention decisions among clinicians and educators.

There are a couple of limitations and ideas for future research worth noting. First, the study did not include a clinical sample by which to compare performance with the nonclinical norming group. Future research should evaluate the discriminant and predictive validity with clinical populations. The test has potential for clinical use, and such metrics would be a critical addition to the findings from the current study. Next, the sample for the test–retest reliability analysis was moderate (n=50). A larger study on test–retest reliability would serve to strengthen these first psychometric findings. In addition, the adult portion of the norming group had a higher percentage of females than males. This outcome was due to chance in the sampling and recruitment response. Any normative updates should consider additional recruitment methods to increase equal sampling distributions by sex. However, to adjust for this in the current normative database, we weighted the scores of the adult sample to match the demographic distribution of gender and education level of the most recent US census. Weighting minimizes the potential for bias in the sample due to disproportionate stratification and unit nonresponse. Although the current study noted a few score differences by gender in the adult sample, it is important to note that only 2% of the individual items showed differential item functioning for gender during the test development stage.[29] Regardless, the current study provided multiple sources of evidence of the validity and reliability of the Gibson Test of Cognitive Skills (Version 2) for use in the general population for assessing cognition across the lifespan. The Gibson Test of Cognitive Skills (Version 2) has been translated into 20 languages and is commercially available worldwide (www. GibsonTest.com).[13]

Disclosure

The authors report employment by the research institute associated with the assessment instrument used in the current study. However, they have no financial stake in the outcome of the study. The authors report no other conflicts of interest in this work.

References

1. Kuhn AW, Solomon GS. Supervision and computerized neurocognitive baseline test performance in high school athletes: an initial investigation. *J Athl Train*. 2014;49(6):800–805.
2. Brinkman SD, Reese RJ, Norsworthy LA, et al. Validation of a self-administered computerized system to detect cognitive impairment in older adults. *J Appl Gerontol*. 2014;33(8):942–962.
3. Geary DC, Hoard MK, Byrd-Craven J, Nugent L, Numtee C. Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev*. 2007;78(4):1343–1359.
4. Williams AM, Ford PR, Eccles DW, Ward P. Perceptual-cognitive expertise in sport and its acquisition: implications for applied cognitive psychology. *Appl Cogn Psychol*. 2011;25:432–442.
5. Bertua C, Anderson N, Salgado JF. The predictive validity of cognitive ability tests: a UK meta-analysis. *J Occup Organ Psychol*. 2005; 78:387–409.
6. Tsourtos G, Thompson JC, Stough C. Evidence of an early information processing speed deficit in unipolar major depression. *Psy Med*. 2002;32(2):259–265.

7. Moats L, Tolman C. *The Challenge of Learning to Read*. 2nd ed. Dallas: Sopris West; 2009.

8. Ruben RJ. Redefining the survival of the fittest: communication disorders in the 21st century. *Int J Pediatr Otorhinolaryngol*. 1999;49:S37–S38.

9. Parbery-Clark A, Strait D, Anderson S, et al. Musical experience and the aging auditory system: implications of cognitive abilities and hear speech in noise. *PLoS One*. 2011;6(5):e18082.

10. McGrew KS. CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*. 2009;37(1):1–10.

11. Flanagan DO, Alfonso VC, Ortiz SO. The cross-battery assessment approach: an overview, historical perspective, and current directions. In: Flanagan DP, Harrison PL, editors. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. 3rd ed. New York: Guilford Press; 2012:459–483.

12. Gathercole S, Alloway TP. Practitioner review: short term and working memory impairments in neurodevelopmental disorders: diagnosis and remedial support. *J Child Psychol Psychiatry*. 2006;47(1):4–15.

13. Gibsontest.com [homepage on the Internet]. Colorado: Gibson Test of Cognitive Skills (Version 2). Available from: http://www.gibsontest.com/web/. Accessed October 4, 2017.

14. Alfonso VC, Flanagan DP, Radwan S. The impact of Cattell–Horn–Carroll theory on test development and interpretation of cognitive and academic abilities. In: Flanagan DP, Harrison PL, editors. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. 2nd ed. New York: Guilford Press; 2005:185–202.

15. Vincent AS, Roebuck-Spencer T, Gilliland K, Schlegel R. Automated neuropsychological assessment metrics (v4) traumatic brain injury battery: military normative data. *Mil Med*. 2012;177(3):256–269.

16. Doniger GM. *NeuroTrax: Guide to Normative Data*; 2014. Available from: https://portal.neurotrax.com/docs/norms_guide.pdf. Accessed October 4, 2017.

17. Elwood RW. MicroCog: assessment of cognitive functioning. *Neuropsychol Rev*. 2001;11(2):89–100.

18. Iverson GL. *Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) Normative Data*; 2003. Available from: https://www.impacttest.com/ArticlesPage_images/Articles_Docs/7ImPACTNormativeDataversion%202.pdf. Accessed October 4, 2017.

19. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Arch Clin Neuropsychol*. 2006;21(7):623–643.

20. Tornatore JB, Hill E, Laboff JA, McGann ME. Self-administered screening for mild cognitive impairment: initial validation of a computerized test battery. *J Neuropsychiatry Clin Neurosci*. 2005;17(1):98–105.

21. Strauss E, Sherman E, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 3rd ed. New York: Oxford University Press; CANTAB; 2006.

22. Moxley-Paquette E, Burkholder G. Testing a structural equation model of language-based cognitive fitness. *Procedia Soc Behav Sci*. 2014;112:64–76.

23. Moxley-Paquette E, Burkholder G. Testing a structural equation model of language-based cognitive fitness with longitudinal data. *Proc Soc Behav Sci*. 2015;171:596–605.

24. Hill OW, Serpell Z, Faison MO. The efficacy of the LearningRx cognitive training program: modality and transfer effects. *J Exp Educ*. 2016;84(3):600–620.

25. Moore AL [webpage on the Internet]. *Technical Manual: Gibson Test of Cognitive Skills (Paper Version)*. Colorado Springs, CO: Gibson Institute of Cognitive Research; 2014. Available from: http://www.gibsontest.com/web/technical-manual-paper/. Accessed October 14, 2017.

26. Moxley-Paquette E. *Testing a Structural Equation Model of Language-Based Cognitive Fitness* [dissertation]. Minnesota: Walden University; 2013.

27. American Education Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 2014.

28. Schneider WJ, McGrew KS. The Cattell–Horn–Carroll model of intelligence. In: Flanagan DP, Harrison PL, editors. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. 3rd ed. New York: Guilford Press; 2012:99–144.

29. Moore AL, Miller TM [webpage on the Internet]. *Technical Manual: Gibson Test of Cognitive Skills (Version 2 Digital Test)*. Colorado Springs: Gibson Institute of Cognitive Research; 2016. Available from: http://www.gibsontest.com/web/technical-manual-digital/. Accessed October 14, 2017.