

Test-retest reliability and validity of a custom-designed computerized neuropsychological cognitive test battery in young healthy adults.

Jinal P Vora¹, Rini Varghese¹, Sara L Weisenbach^{2,3}, Tanvi Bhatt¹

¹Department of Physical Therapy, University of Illinois, Chicago, IL, USA

²Department of Psychiatry, University of Utah, Salt Lake City, UT, USA

³Research and Development Program, VA Salt Lake City, Salt Lake City, UT, USA

Abstract

Objective: Dual-task methodologies are utilized to probe attentional resource sharing between motor and cognitive systems. Computerized neuropsychological testing is an advanced approach for cognitive assessment and its application in dual task testing is evolving. This study aimed to establish the test-retest reliability and concurrent validity of a custom-designed, computerized, cognitive test battery.

Methods: Fifteen healthy young adults were tested for the following domains (and tasks): 1) visuomotor function (Spot and Click, SC), 2) phonemic memory (Category Naming, Cat N) and verbal fluency (Word List Generation, WLG), 3) response inhibition (Color Naming, CN), 4) discriminant decision-making (Unveil the Star, US), 5) visual working memory (Triangle and Letter Tracking, TT and LT), 6) problem solving (Peg Game, PG) and 7) information processing speed (Letter-Number, LN). The reaction time, accuracy, time of completion, total number of responses and total number of errors were used as the outcome variables.

Results: The intraclass correlation coefficient (ICC) was used to determine reliability for all outcome variables and concurrent validity was established with respect to the Delis Kaplan Executive Function System™ (D-KEFS™). Reliability ranged from good to excellent for all seven tasks (ICC>0.65). The Cat.N, WLG and CN showed good correlation and PG task showed moderate correlation with tests of the D-KEFS.

Conclusion: Findings indicate that these computerized cognitive tests were both valid and reproducible and therefore can be easily implemented by clinicians for assessing cognition and incorporated for dual-task testing and training.

Keywords: Psychometrics; Dual-task; Neuropsychological testing; Working memory; Visuomotor function; Decision making; Verbal fluency.

Accepted August 20, 2016

Introduction

Cognitive decline is prevalent in both healthy aging [1] and in individuals with neuropathological diagnoses [2] and is among the strongest determinants of real world functioning and quality of life in affected individuals [3]. Although cognitive decline is not preventable, it can be slowed with timely and accurate testing and appropriate training. Feasible methods of neuropsychological assessment are thus crucial in enhancing the quality of care available for these individuals.

However, cognitive decline is often not an isolated problem and is known to interfere with everyday

motor activities such as walking, driving, manipulating objects, etc. [4,5]. This is referred to as cognitive-motor interference, and occurs when a motor and cognitive task is performed concurrently, resulting in poor performance on one or both tasks. These effects are attributed to competing demands of the two systems for either limited attentional resources or limited processing capacity [6,7]. Dual task assessments, using a cognitive-motor paradigm, can be helpful in capturing changes in the normal interaction of these two systems as a result of age, disease or rehabilitation, but require accurate and sensitive assessment tools that can be employed even while performing motor activities.

Traditionally, cognitive function has been evaluated using paper-based tests, which are lengthy, time consuming, often require special training and are prone to manual error [8-10]. Furthermore, by nature, these measures are also not feasible for use in cognitive-motor dual task paradigm, limiting their overall usability in rehabilitation settings.

Computerized cognitive testing (CCT) is a new and developing approach that may offer a potential alternative to some of these conventional methods of testing. CCT has many advantages over conventional methods; they are more consistent, in terms of administration and scoring, afford sensitive measurement, in the order of milliseconds, allow precise stimulus control, are relatively cost efficient, can be visually appealing and can be used to create and maintain digital records [11]. Numerous computerized batteries have been developed and reported in the literature; however, these can be fairly expensive, limiting their use.

In this study we develop an affordable alternative for computerized cognitive testing that may also have other potential applications. The cognitive battery was designed using the DirectRT Empirisoft [12] and consisted of tasks that measured domains directly relevant to dual tasking, such as executive functions, working memory, fluency and attention span. The DirectRT software used in this study is cost effective, user friendly and allows customizable test designs, making it very feasible to develop and administer a variety of cognitive tests without specialized training. Nonetheless, before this tool can be used clinically, it is important to determine its psychometric properties. The purpose of this study, therefore, was to establish the test-retest reliability and concurrent validity of a custom designed computerized neurocognitive battery in healthy young adults using the commercially available Direct RT™ Empirisoft.

Methods

Participants

Fifteen healthy young adults (23.87 years ± 1.35, 16.80 ± 1.47 years of education) were recruited for the study via informational flyers posted across the University campus. Participants were included if they had no self-reported physiological, neurological and psychiatric conditions. Approval from the University Institutional Review Board was obtained prior to start of the study. All the participants signed an informed consent form before participating in the study.

Participants completed three paper-and-pencil measures (Verbal Fluency, Color-Word Interference and Tower of Hanoi Test) drawn from the Delis Kaplan Executive Function System™ (D-KEFS™) [13], which is a commonly used neuropsychological battery that assesses different aspects of executive function with a good to moderate internal consistency. Participants also completed a computerized neurocognitive battery test administered using DirectRT™, Empirisoft. The outcome variables for each these batteries are presented in Tables 1 and 2.

Protocol

All testing was completed in a silent room in order to avoid external disturbances or distractions. The D-KEFS tests were administered by a trained examiner who sat across from the participant and read aloud the standard instructions from the manual. Depending on the test, the time of completion or responses generated were noted by the examiner as outcome measures. For the computerized testing, the screen was positioned to be in front of the participant and noise-cancelling headsets with a microphone were used to record the responses. Each test was preceded by an instruction slide and the participant was instructed to press a key when ready to begin testing. Participants were instructed to provide quick and accurate

Table 1. *Computerized cognition tests assessed for test-retest reliability and outcome variables*

| Test Name | Cognitive function assessed | Outcome Variables |
|--|--------------------------------------|--|
| Spot and click | Visuo-motor function | Reaction time |
| Category Naming and Word List Generation | Verbal Fluency | Number of responses |
| L-N Sequencing | Cognitive Flexibility/Switching | Number of correct responses |
| Triangle Tracking and Letter Tracking | Working memory | Number of correct responses |
| Color Naming | Processing Speed/Response Inhibition | Total Time of Completion |
| Unveil the Star | Spatial Working Memory | Total time of completion Total number of errors |
| Peg Board Game | Discriminant Decision Making | Total time of completion Total number of errors |

Table 2. *Computerized cognitive test used to assess validity and outcome variable*

| D-KEFS | DirectRT | Cognitive function assessed | Outcome Variables |
|----------------|---|--------------------------------------|--|
| Verbal Fluency | Category Naming Word List Generation | Verbal Fluency | Number of responses |
| Visual Stroop | Color Naming (Visual Stroop) | Processing Speed/Response Inhibition | Total Time of Completion Accuracy % |
| Tower of Hanoi | Peg Board game | Discriminant Decision Making | Total time of completion Total number of errors |

responses to each of the tests in the battery and all tests were administered in a randomized order at two sessions separated by a 10-12 day interval. The D-KEFS items (i.e., Verbal Fluency, Color- Word Interference and Tower Test) were performed at the first session and in the same order as the computerized testing.

Cognitive Test Battery

Computerized Cognitive Testing (CCT)

Category naming and word list generation test: Semantic fluency was assessed by providing a category cue to the participant (i.e., Animals, boys' names or countries), while phonemic fluency was measured by providing a letter cue to the participant (i.e., F, A or S). Participants were given one minute to provide as many words as possible and instructed not to list any proper nouns such as name of places or people. Voice responses were recorded by the computer [14] (Figure 1A).

Color naming test: This test is an adaptation of the classic Stroop paradigm measuring inhibition and cognitive flexibility. This version consisted of two conditions, with first condition recording the amount of time it takes the participant to read the color in which the word is printed (congruent condition). Condition 2 required the participant to name the ink (color) in which the color word was printed

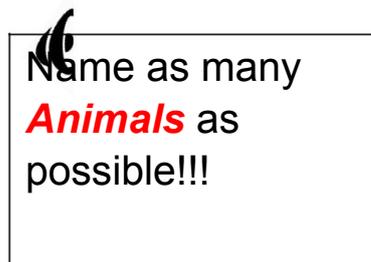
(incongruent condition). The accuracy and total time to complete the task were recorded (Figure 1B).

L-N sequencing test: This is the oral version of the paper-and-pencil Trail Making Tests A and B wherein the participant hears a pair of a letter and a number. This test is a measure of working memory but also has an added component of assessing one's cognitive flexibility as they are instructed to loudly narrate the next letter and number (e.g. if they hear "A-2" their response would be B-3, C-4, and D-5 and so on till they hear the second cue. Each cue is presented for 15 seconds and three trials are collected. Each trial started with a different cue pair. The total number of correct responses was averaged across the three trials [15].

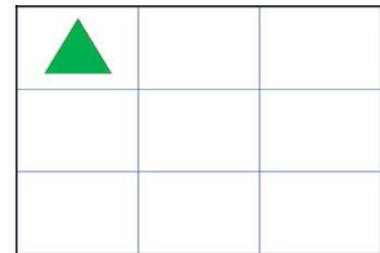
Triangle and letter tracking test: This is a measure of working memory in which the participant is presented with a sequence of stimuli, one at time, and is asked to indicate when the current stimulus matches the one from n-steps earlier in the sequence. We used 1 and 2 steps earlier for the current protocol. Two different sets of tests were used. In the first set the participant tracked a triangle that moved to different positions in a grid, and they were asked to respond when they saw the triangle in same position as in the previous trial i.e.1 trial back or 2 trials back. The second set was similar, but involved letters that



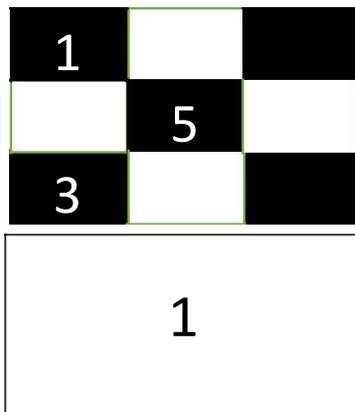
1A: Color Naming Test



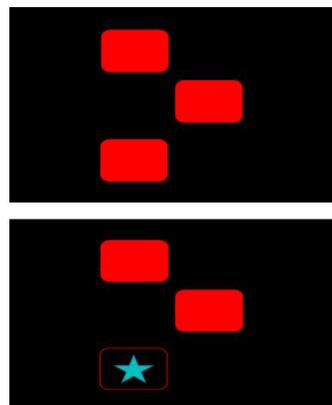
1B: Category Naming



1C: Triangle Tracking



1D: Number and Position



1E: Unveil the Star



1F: Spot and Click

Figure 1. Screen shot for different cognitive tasks used for assessment. Image 1a represents the incongruent slide for the color naming task wherein the participant has to name the ink color in which the word is printed. Image 1b and 1f represents the instructions given prior to onset of the task. Image 1-c represents the triangle in the grid which the participant tracks and responds when he observes the triangle in the same location n-trials back. Image 1d represents the number & position task (i.e., 1- upper left) response recorded by the participant. Image 1e represents unveil the star task

were presented to the participant and they were asked to respond when they saw the same letter repeating 1-trial or 2-trial back [16] (Figure 1C).

Unveil the star test: This test measures the retention and manipulation of visuo-spatial information wherein the participant was asked to search for a star in multiple boxes. Once the star in one box was found he or she continued to look for the star in other boxes but had to remember not to click in the same box where the star was found earlier. There were three levels in the games with increasing difficulty where in participants' were asked to find three stars in level 1, progressing to five stars in Level 2 and eight stars in level 3 demanding more attention and concentration in order to minimize the errors (Figure 1D).

Spot and click test: This test measures the amount of time taken by the examinee to respond after the stimulus is presented. Both Simple Reaction time (SRT) and Choice Reaction time (CRT) were assessed. The participants' were presented with a stimulus (yellow circle) after which they were asked to respond by pressing the corresponding key on the number pad representing the location of the stimulus [17] (Figure 1E).

Peg board game: This test assesses problem solving capacity along with spatial working memory. Participants were asked to move disks in order to arrange them in a predetermined position using the number pad with fewest moves possible. Pictorial representation of the computerized cognitive test battery is seen in Figure 1.

Conventional Cognitive Testing D-KEFS™

Verbal fluency test: This measure assessed semantic and phonemic fluency by providing a category cue and a letter cue to the participant (i.e., Animals, Boy's name and letters F, A or S). Participants were given one minute to provide as many words as possible; the responses were noted by the examiner. For the phonemic fluency task, participants were instructed not to list proper nouns [13].

Stroop test: This measure consists of three conditions, relying to various extents on processing speed and inhibitory control. Condition 1 and condition 2 are baseline conditions that consist of basic color naming, or reading color names printed in black ink, respectively. Condition 3 is the traditional Stroop test in which the participant is asked to name the dissonant ink color and inhibit reading the color words that is printed. Each condition consisting of 45 words. The reliability correlation stated for the color naming test ranged between moderate to high [13].

Tower of Hanoi test: This test is required to measure the participant's, memory, problem solving, planning and decision making abilities. The main objective of the test was to move the disks of varying size (i.e., small, medium, or large) across three pegs to build a designated tower in the fewest number of moves. At the start of every test the starting position of the disks was predetermined while the ending position was shown to the participant and they

were asked to match it to the target tower. The manual mentioned that the test-retest correlation for this test was within the moderate range [13].

Statistical Analysis

All the statistical analyses were performed using Statistical Package for the Social Sciences (SPSS) (version 22) for Windows. Descriptive statistics (mean \pm SD) were performed for all the cognitive test variables using paired sample t-tests and are reported in the text. Intra-class correlation coefficients (ICC) were used to determine the reliability of each of the tests in the cognitive battery. Bland Altman plot were constructed to display the level of agreement between the differences in means of the two testing sessions. One-sample t-test was used to analyze if these differences (i.e., bias) were significantly different from zero. The Pearson's product moment correlation (r) and coefficient of determination (r^2) were used to quantify the strength of the relationship between the data from sessions 1 and 2.

Results

Paired t-tests showed no significant difference between the two testing sessions in performance among the different variables for each test for example for the Spot and Click task, simple reaction time ($t=0.007$, $p=0.994$) and choice reaction time ($t=0.226$, $p=0.825$), for the total number of responses in the Category Naming Task ($t=-0.338$, $p=0.740$), Word List Generation ($t=-0.133$, $p=0.896$), Letter Number Sequencing ($t=-0.425$, $p=0.677$), the Triangle Tracking Task, i.e., 1-back ($t=-0.00$, $p=1.0$), 2-back ($t=-1.79$, $p=0.09$) and similarly for the Letter Tracking Task, i.e., 1-back ($t=1.17$, $p=0.26$) and 2-back ($t=0.42$, $p=0.69$). The total time completion for Color Naming task i.e. the congruent slide ($t=1.54$, $p=0.145$) and incongruent slide ($t=1.02$, $p=0.32$), for Unveil the Star Task ($t=1.70$, $p=0.11$) and the Peg Board Task ($t=2.25$, $p=0.04$). The total number of errors calculated for Unveil the Star Task ($t=0.44$, $p=0.67$) and the Peg Board Task ($t=1.31$, $p=0.21$) (Table 3).

Reliability

The ICC with 95% CI for each variable is presented in Table 4. The ICC's ranged from 0.65-0.92 for all the tasks. Figure 2 shows Bland-Altman plots displaying limits of agreement between the testing sessions. One sample t-test suggested that there was no significant difference from 0 in the mean difference between the two testing sessions in Spot and Click Task, choice reaction time (5.24 ± 89.91 ; $p=0.825$), the total number of responses in category naming (-0.467 ± 5.34 ; $p=0.740$), Word List Generation (-0.133 ± 3.88 ; $p=0.896$), Letter Number Sequencing (-0.200 ± 1.820 ; $p=0.677$) along with the total time completion for Unveil the Star Task (4.49 ± 10.21 ; $p=0.110$), Color Naming task: Incongruent slide (-0.467 ± 5.34 ; $p=0.740$), Peg Board Task (11.29 ± 19.44 ; $p=0.04$) and finally total number of errors in the Peg Board Task (6.19 ± 18.28 ; $p=0.211$).

Table 3. Presents mean and standard deviation on test session 1 and test session 2, Pearson's product moment correlation, *r*, and the co-efficient of determination, *r*²

| Test Names | Outcome Variables | Test session 1 | | Test session 2 | | <i>r</i> | <i>r</i> ² |
|--------------------------|------------------------------|----------------|--------|----------------|--------|----------|-----------------------|
| | | Mean | SD | Mean | SD | | |
| Spot and Click | Simple reaction time (ms) | 337.32 | 77.29 | 337.187 | 108.56 | 0.77 | 0.59 |
| | Choice reaction time (ms) | 694.267 | 150.06 | 689.02 | 135.09 | 0.81 | 0.65 |
| Category Naming | Number of responses | 22.20 | 5.82 | 22.66 | 6.75 | 0.65 | 0.42 |
| Word List Generation | Number of responses | 17.87 | 5.50 | 18 | 3.38 | 0.74 | 0.51 |
| Letter Number Sequencing | Number of responses | 6.93 | 1.83 | 7.13 | 1.72 | 0.48 | 0.23 |
| Color Naming | Congruent (total time, s) | 17.71 | 3.92 | 16.38 | 3.95 | 0.64 | 0.41 |
| | Incongruent (total time, s) | 33.608 | 7.87 | 32.57 | 6.01 | 0.88 | 0.77 |
| Unveil the Star | Total time of completion (s) | 48.32 | 12.56 | 43.82 | 11.19 | 0.63 | 0.40 |
| | Error (%) | 1475.83 | 25.75 | 1477.08 | 26.05 | 0.91 | 0.84 |
| Triangle Tracking | 2-back Accuracy (%) | 70.83 | 15.43 | 78.33 | 13.64 | 0.52 | 0.26 |
| Letter Tracking | 2-back Accuracy (%) | 74.31 | 18.04 | 72.37 | 19.82 | 0.56 | 0.32 |
| Peg Board Game | Total time of completion (s) | 91.18 | 38.12 | 79.88 | 39.62 | 0.87 | 0.76 |
| | Error (%) | 37.14 | 29.35 | 31.90 | 29.31 | 0.82 | 0.66 |

Table 4. ICC values for all the variables

| Test Names | Outcome Variables | ICC | P-Value |
|--------------------------|------------------------------|------|---------|
| Spot and Click | Simple reaction time (ms) | 0.84 | 0.00 |
| | Choice reaction time (ms) | 0.89 | 0.00 |
| Category Naming | Number of responses | 0.78 | 0.04 |
| Word List Generation | Number of responses | 0.78 | 0.04 |
| Letter Number Sequencing | Number of responses | 0.65 | 0.03 |
| Color Naming | Congruent (total time, s) | 0.78 | 0.004 |
| | Incongruent (total time, s) | 0.92 | 0.00 |
| Unveil the Star | Total time of completion (s) | 0.77 | 0.004 |
| | Error (%) | 0.95 | 0.00 |
| Triangle Tracking | 2-back Accuracy (%) | 0.71 | 0.017 |
| Letter Tracking | 2-back Accuracy (%) | 0.72 | 0.012 |
| Peg Board Game | Total time of completion (s) | 0.91 | 0.00 |
| | Error (%) | 0.89 | 0.00 |

Table 5. Correlation coefficients for the validation of the computerized cognitive test battery

| Test Names | Outcome variables | D-KEFS | | DirectRT | | <i>r</i> | <i>r</i> ² |
|--------------------------|-----------------------------|--------|-------|----------|-------|----------|-----------------------|
| | | Mean | SD | Mean | SD | | |
| Verbal Fluency (Cat N) | Number of responses | 21.625 | 4.801 | 22.18 | 5.04 | 0.762 | 0.581 |
| Verbal Fluency (WLG) | Number of responses | 16.5 | 3.59 | 18.5 | 5.04 | 0.72 | 0.518 |
| Visual Stroop Test (CN) | Congruent (total time, s) | 26.31 | 4.46 | 26.47 | 6.25 | 0.798 | 0.637 |
| | Incongruent (total time, s) | 47.44 | 8.9 | 45.24 | 7.05 | 0.731 | 0.467 |
| | Accuracy % (congruent) | 44.55 | 0.72 | 45 | 0 | * | * |
| | Accuracy % (Incongruent) | 44.44 | 0.72 | 44.88 | 0.33 | * | * |
| Tower of Hanoi Test (PG) | Total time (s) | 92.64 | 51.29 | 89.22 | 39.52 | 0.565 | 0.319 |
| | Total number of errors | 22.98 | 14.41 | 18.11 | 14.85 | 0.752 | 0.565 |

*No variance was observed between the means for accuracy % for Visual Stroop test administered via D-KEFS and Color Naming test administered via DirectRT

Validation

Correlation Coefficients are presented in Table 5. Good correlations were found for the tests measuring the domains of working memory and information processing speed, i.e., the verbal fluency test and the Visual Stroop test respectively. Moderate correlation was found for Tower of Hanoi test which measured individuals problem solving and planning abilities on comparing the computerized cognitive and conventional paper-pencil tests.

Discussion

Computerized cognitive tests may provide very reliable indications of cognitive function in longitudinal investigations [8]. However, the reliability of computerized cognitive testing administered via DirectRT, Empirisoft had not been determined. Overall results from this study indicate that CCT's administered using DirectRT™ Empirisoft provides a measurement of cognitive function that is highly reliable between sessions in young healthy adults.

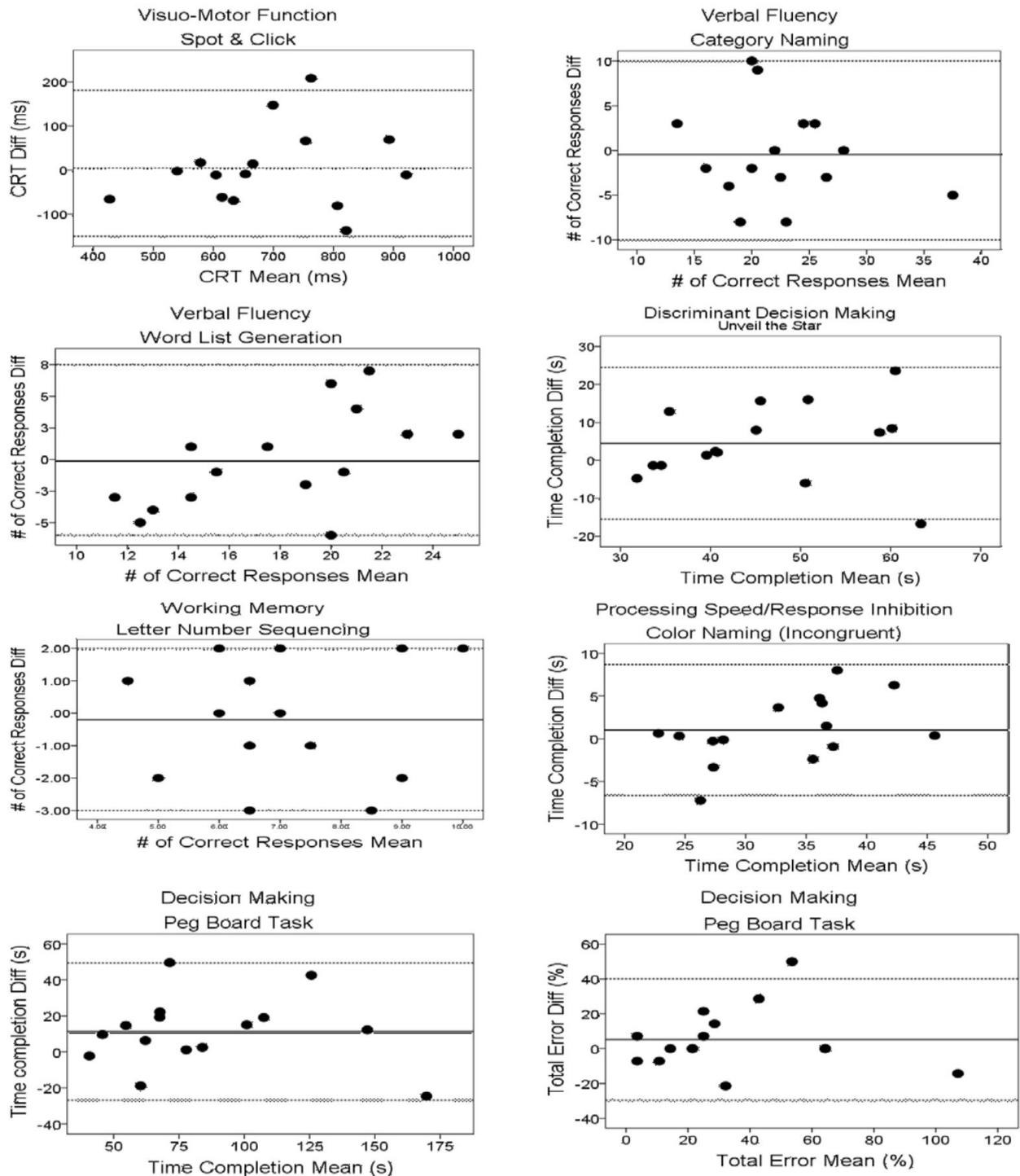


Figure 2. Bland-Altman plots displaying limits of agreement: mean (solid line) \pm 1.96 SD (dotted line) of the difference between test session 1 and test session 2 values for the variables: (a) choice reaction time (CRT); (b, c and e) number of correct responses; (d, f and g) total time completion; and (h) total error

Out of the computerized cognitive tests administered, some tests seem to have high test/retest correlation compared to others, this can be due to the different cognitive domains (i.e., visuo-motor function, working memory, executive function, discriminant decision making or verbal fluency) tapped during the testing session. The simple and choice reaction time tests had a high reliability co-efficient (i.e., ICC=0.84 and 0.89, respectively) and compare favorably with the reaction time reliability co-efficients previously reported in existing computerized cognitive assessment

tools ranging from 0.55 to 0.90 [8,11,18]. As the stimuli in both the events were suddenly presented within a fixed time interval unknown to the participant being tested, it can be inferred that the high test-retest correlation observed in this study was not biased depending on any practice effects. Similar rationale has been reported in a study by Lowe and Rabbitt [8], who mentioned that reaction time, is a non-strategic driven task thereby limiting practice effects and resulting in high test/retest correlation.

Tests measuring processing speed and response inhibition (Color Naming task), discriminant decision making (Peg Board game), spatial working memory (Unveil the Star task) and working memory (Letter Number Sequencing, Triangle and Letter Tracking task) reported a high to moderate intraclass correlation range. Total time required to complete congruent and incongruent conditions of the computerized Stroop task were highly correlated between the two sessions. Total time to complete the incongruent condition was almost twice the time taken to complete the congruent slide, and is a common finding due to the complexity imposed by the multiple stimulus-response condition [8,19].

The correlation for the three tests, i.e., Letter Number Sequencing test measuring working memory along with cognitive flexibility and switching capability and the Triangle Tracking task and Letter Tracking task that measured working memory were moderate. A possible explanation for this can be that the test of working memory tests are usually 'strategy driven tasks' and hence provide accurate results only when they are novel because performance on them can improve as soon as the participant discovers an optimal strategy, but will improve less or not at all if no strategy is found. This learning effect can also be markedly different between each participant, which can result in greater variance and hence lower the ICC. These postulations probably account for the moderate reliability for the working memory tasks compared to higher reliability seen with the reaction time task and have been previously validated [8]. Collie et al. had published findings supporting the above explanation, by suggesting that healthy people respond quickly and make fewer mistakes on tests of psychomotor functions in contrast to tests of decision making and working memory where they tend to be slow and make more errors [18]. For Unveil the Star task, in which participants are required to retain and manipulate visuo-spatial information in order to complete a complex task, a high test-retest correlation for the total number

of errors and a moderate test-retest correlation for the total time required to complete the task were found. The variance in the test/retest correlation could be because the number of errors made by each participant across the two testing sessions remained constant, but the speed of completing the task during the second session increased slightly compared to the first. The change in speed could have occurred also due to this being a strategy driven task, as mentioned earlier.

Validity was tested for only three of the seven custom-designed tasks against the conventional paper-pencil testing methods (D-KEFS). Concurrent validity, estimated by correlation of the verbal fluency and processing speed/response between the computerized testing and D-KEFS was good 0.79 and 0.72, respectively indicating that both tests correlate well with performance on conventional testing methods. No variance was observed between the responses recorded for accuracy for the Visual Stroop test administered via D-KEFS as compared to responses recorded on the Color Naming task administered via the DirectRT. The Peg board task validated against the Tower of Hanoi test, showed moderate correlation, i.e., total time completion was 0.565 and total number of errors 0.752. These results indicated that computerized cognition testing methods are reliable and despite the differences in the method of stimulus presentation (physical, e.g. peg, versus virtual) [20] and mode of administration (computer versus in person) have good concurrent validity and hence could be used to model the conventional Tower of Hanoi test [11].

The results of this study could be confounded by its small sample size. However, to address this post-hoc sample size calculation was performed for each of the outcome variables using Pearson's correlation co-efficient (r), with the power being set at 0.84 and a significance level of 0.05. The resulting sample size from this analysis was 15, horizontal black line, shown in the Figure 3 [21]. Besides

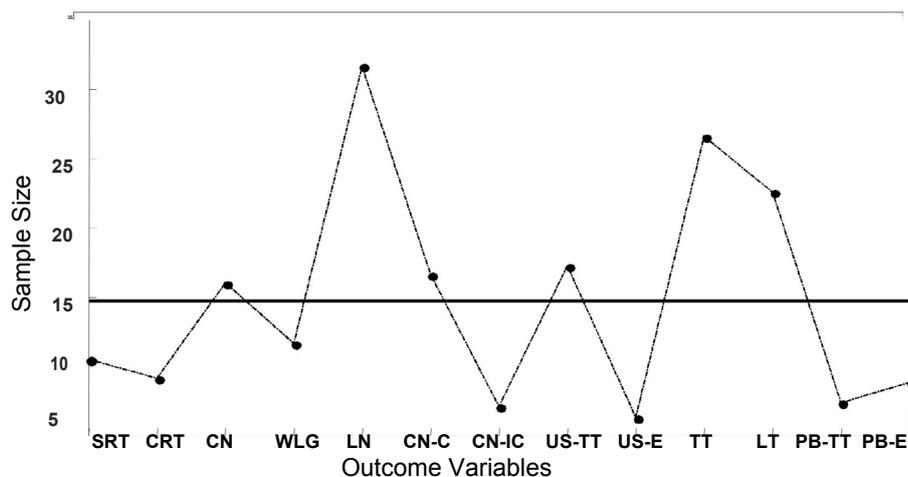


Figure 3. Post-hoc sample size analysis plot displaying number of required subjects (black solid line) for the variables of interest: Simple Reaction Time (SRT), Choice Reaction Time (CRT), Category Naming (CN), Word List Generation (WLJ), Letter Number Sequencing (LN), Color naming: Congruent slide (CN-C) and incongruent slide (CN-IC), Unveil the Star task: Total time of completion (US-TT), number of errors (US-E) and Peg Board task: Total time of completion (PB-TT), number of errors (PB-E). The plot was obtained using a custom written Matlab Algorithm [21]

this results could also be confounded due to the practice effect as tasks measuring accuracy percent, total number of errors or total time completion are strategy driven and yield best results when novel after which an individual might be able to develop a strategy which could contribute to high or moderate correlation in order to avoid this practice effect the average number of days between the two testing sessions can be increased in future studies.

To conclude, the results of the current study indicate that the custom designed computerized cognitive tests administered using DirectRT is reliable. The Category Naming task, Word List Generation task, Color Naming task and of Hanoi Peg Board task showed to be valid measures when compared to the gold standard tests administered via D-KEFS. Future studies should establish psychometric properties for this computerized test battery for other populations including older adults with and without cognitive deficits and people with neurological disorders. CCT's can further be employed to measure the effect of cognitive-motor rehabilitation interventions in neurological conditions such as stroke, traumatic brain injury, Alzheimer's or in sports medicine, i.e., concussions in contact sports by measuring their baseline and post intervention findings. Lastly, the ubiquity of computers and the ease with which this battery can be administered, it could be easily translated to clinical and community-settings and could also be feasibly used to build self-confidence and promote behavior modifications amongst different population with an multidisciplinary approach involving care managers, general practitioners, occupational, physical and speech therapists and clinical neuropsychologists [22].

Acknowledgement

The project described was supported by Grant Number P30AG022849 from the National Institute on Aging. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

Financial Support

This work was supported by a pilot grant the Midwest Roybal Centre for Health Promotion and Translation under Grant P30AG022849 awarded to Dr. Tanvi Bhatt.

References

1. Troyer AK, Rowe G, Murphy KJ, et al. Development and evaluation of a self-administered on-line test of memory and attention for middle-aged and older adults. *Frontiers in Aging Neuroscience* 2014; 6: 335.
2. McDonnell MN, Bryan J, Smith AE, et al. Assessing cognitive impairment following stroke. *Journal of Clinical and Experimental Neuropsychology* 2011; 33: 945-953.
3. Bartels C, Wegrzyn M, Wiedl A, et al. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC neuroscience* 2010; 11: 118.

4. Owsley C, Ball K, Sloane ME, et al. Visual/cognitive correlates of vehicle accidents in older drivers. *Psychology and Aging* 6: 403-415.
5. Patel P, Bhatt T. Task matters: Influence of different cognitive tasks on cognitive-motor interference during dual-task walking in chronic stroke survivors. *Topics in Stroke Rehabilitation* 2014; 21: 347-357.
6. Lundin-Olsson L, Nyberg L, Gustafson Y. Attention, frailty and falls: The effect of a manual task on basic mobility. *Journal of the American Geriatrics Society* 1998; 46: 758-761.
7. Plummer P, Eskes G, Wallace S, et al. Cognitive-motor interference during functional mobility after stroke: State of the science and implications for future research. *Archives of Physical Medicine and Rehabilitation* 2013; 94: 2565-2574
8. Lowe C, Rabbitt P. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Cambridge Neuropsychological Test Automated Battery. International Study of Post-Operative Cognitive Dysfunction. Neuropsychologia* 1998; 36: 915-923.
9. Heaton RGC, Talley J, Kay G, et al. Wisconsin Card sort test manual: Revised and expanded. Odessa, Florida Psychological Assessment Resources, Inc., USA 1993.
10. Rh BJB. Hopkins verbal learning test--revised: Professional manual. Psychological Assessment Resources 2001.
11. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS vital signs. *Archives of Clinical Neuropsychology* 2006; 21: 623-643.
12. Jarvis BG. DirectRT (Version 200x.x.x) Computer Software. Empirisoft Corporation, New York 2014.
13. Delis DC, Kramer JH, Kaplan E, et al. Reliability and validity of the Delis-Kaplan executive function system: An update. *Journal of the International Neuropsychological Society* 2004; 10: 301-303.
14. Boringa JB, Lazeron RH, Reuling IE, et al. The brief repeatable battery of neuropsychological tests: Normative values allow application in multiple sclerosis clinical practice. *Multiple Sclerosis* 2001; 7: 263-267.
15. Grigsby J, Kaye K. Alphanumeric sequencing and cognitive impairment among elderly persons. *Perceptual and motor skills* 1995; 80: 732-734.
16. Owen AM, McMillan KM, Laird AR, et al. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping* 2005; 25: 46-59.
17. Hyman R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology* 1953; 45: 188-196.
18. Collie A, Maruff P, Makdissi M, et al. CogSport: reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clinical Journal of Sport Medicine* 2003; 13: 28-32.

19. Gruber SA, Rogowska J, Holcomb P, et al. Stroop performance in normal control subjects: An fMRI study. *Neuroimage* 2002; 16: 349-360.
20. Howe AE, Arnell KM, Klein RM, et al. The ABCs of computerized naming: equivalency, reliability and predictive validity of a computerized rapid automatized naming (RAN) task. *Journal of Neuroscience Methods* 2006; 151: 30-37.
21. Shoukri M, Asyali M, Donner A. Sample size requirements for the design of reliability study: Review and new results. *Statistical Methods in Medical Research* 13: 251-271.
22. Ciccone MM, Aquilino A, Cortese F, et al. Feasibility and effectiveness of a disease and care management model in the primary health care system for patients with heart failure and diabetes (Project Leonardo). *Vasc Health Risk Manag* 2010; 6: 297-305.

Correspondence to:

Tanvi Bhatt,
PT, PhD,
Department of Physical Therapy,
1919, W Taylor St, (M/C 898), University of Illinois,
Chicago, IL 60612,
USA.
Tel: 312-355-4443
Fax: 312-996-4583
E-mail tbhatt6@uic.edu