[Athletic Training]

Test-Retest Reliability of a Computerized Concussion Test: CNS Vital Signs

Ashley C. Littleton, MA, ATC,*†‡§ Johna K. Register-Mihalik, PhD, ATC,†‡§
and Kevin M. Guskiewicz, PhD, ATC†‡§

Background: Neurocognitive testing is an important concussion evaluation tool, but for neurocognitive tests to be useful, their psychometric properties must be well established. Test-retest reliability of computerized neurocognitive tests can influence their clinical utility. The reliability for a commonly used computerized neurocognitive test, CNS Vital Signs, is not well established. The purpose of this study was to examine test-retest reliability and reliable change indices for CNS Vital Signs in a healthy, physically active college population.

Hypothesis: CNS Vital Signs yields acceptable test-retest reliability, with greater reliability between the second and third test administration compared with between the first and second administration.

Study Design: Cohort study.

Level of Evidence: Level 3.

Methods: Forty healthy, active volunteers (16 men, 24 women; mean age, 21.05 ± 2.17 years) reported to a clinical laboratory for 3 sessions, 1 week apart. At each session, participants were administered CNS Vital Signs. Outcomes included standard scores for the following CNS Vital Signs domains: verbal memory, visual memory, psychomotor speed, cognitive flexibility, complex attention, processing speed, reaction time, executive functioning, and reasoning.

Results: Participants performed significantly better on the second session and/or third session than they did on the first testing session on 6 of 9 neurocognitive domains. Pearson *r* test-retest correlations between sessions ranged from 0.11 to 0.87. Intraclass correlation coefficients ranged from 0.10 to 0.86.

Conclusion: Clinicians should consider using reliable change indices to account for practice effects, identify meaningful score changes due to pathology, and inform clinical decisions.

Clinical Relevance: This study highlights the importance of clinicians understanding the psychometric properties of computerized neurocognitive tests when using them in the management of sport-related concussion. If CNS Vital Signs is administered twice within a small time frame (such as 1 week), athletes should be expected to improve between the first and second administration.

Keywords: concussion; psychometrics; cognition; computerized neurocognitive testing; Concussion Vital Signs

Sport-related concussion has been a growing concern in the United States, with up to 3.8 million occurring each year.[13] Concussion assessment should involve a multifaceted approach, with the goal of ensuring that the individual has returned to their preinjury status before returning to play.[17] One recommended aspect of concussion assessment is neurocognitive testing[24] because deficits exist in individuals postconcussion even after symptoms have resolved.[1] Typically,

neurocognitive tests are administered preinjury (for comparison) and postinjury once an athlete is asymptomatic; however, earlier postinjury assessments may aid in determining certain aspects of management, such as returning to school or work.[17] In fact, using symptoms and neurocognitive assessments within 2 days postinjury can correctly identify 80% of athletes requiring protracted recovery.[14] This serial assessment paradigm makes the reliability and precision of concussion assessments especially important. Using a short test-retest interval, such as 1 week, is helpful in identifying when scores stabilize over time and identifying whether a second test administration is needed to stabilize scores on specific measures.[6,20,23]

There are several computerized concussion tests developed for concussion assessment. One well-developed neurocognitive platform, CNS Vital Signs (CNSVS),[8] has recently been introduced as a concussion assessment tool, Concussion Vital Signs (CNS Vital Signs LLC). While the psychometric properties of the tests that comprise CNSVS are well established,[8] the presentation of the tests is novel; therefore, reliability must be established. The reliability of CNSVS has been examined in a large age range of participants (7-90 years old), with a large range in the test-retest interval (3-282 days) and only 2 time points.[3,8] Therefore, there is a lack of reliability data for multiple testing sessions in college-aged athletes. Physical activity[15,25] and age[11,19] have been shown to influence performance on cognitive tasks and could consequently alter reliability of the measure. Reliability data for college-aged physically active individuals is needed to aid in clinical decision making, especially regarding return to play after concussion. Therefore, the purpose of this study was to examine the test-retest reliability of CNSVS across 3 sessions, each approximately 1 week apart, in a healthy, active college population. A secondary purpose was to determine reliable change indices among this same population. We hypothesized that CNSVS would yield acceptable test-retest reliability (intraclass correlation coefficient [ICC]$_{2,1}$, 0.40-0.75), with greater reliability between the second and third test administrations compared with between the first and second administrations.

METHODS

The study was conducted following the ethical guidelines set forth by the Department of Health and Human Services Office for Human Research Protection (USA) and approved by University of North Carolina–Chapel Hill Institutional Review Board prior to the study initiation (reference study #10-1831 and #10-0271). All participants signed approved consent forms prior to participation. A convenience sample of 40 healthy, active volunteers participated in this study (16 men, 24 women; mean age, 21.05 ± 2.17 years). Only participants who met the following criteria were recruited: reported consistently completing at least 30 minutes of cardiovascular and/or resistive training at least 3 times per week and no history of 3 or more concussions, concussion in the past 6 months, learning disability, attention deficit/hyperactivity disorder (ADHD), any

known neurologic disorder or psychological disorder that would affect cognition, or a primary language other than English. Forty-three participants initially reported, and 3 were excluded from the study because they did not report for the second or third session. Any scores that were deemed invalid by CNSVS criteria[2] were excluded from analysis.

Participants were tested on 3 separate sessions, each 6 to 11 days apart (mean time between sessions 1 and 2, 7.77 ± 2.67 days; between sessions 2 and 3, 6.63 ± 1.17 days). This 1-week psychometrically sound time frame was selected because athletes are often assessed serially after concussion. It also may help identify whether a second test administration is needed to stabilize scores on specific measures.[10] Prior to participation in the study, each participant completed a demographic form to ensure that all inclusion and exclusion criteria were met and completed the Graded Symptom Checklist (GSC).[9] Participants were administered CNSVS individually in a quiet controlled setting and were instructed to answer quickly and accurately, to carefully read all instructions, to sustain their attention throughout the entire test, and to notify the test administrator if they had any questions throughout the test. The test took approximately 30 minutes to complete (see Appendix 1, available at http://sph.sagepub.com/content/by/supplemental-data). Main outcome measures included standard scores for the following domains: verbal memory, visual memory, psychomotor speed, cognitive flexibility, complex attention, processing speed, reasoning, reaction time, executive functioning, and reasoning. Standard scores are based on a normative data set that matches participants by age and places all outcomes on the same scale to provide for an easier clinical understanding.

To determine whether participants reported significantly more symptoms at any of the sessions, a repeated-measures analysis of variance (ANOVA) was calculated with the dependent variable of total symptom score from the GSC and the independent variable of time (time 1, 2, and 3). To determine practice effects, a series of repeated-measures ANOVAs were calculated across all 3 time points with each of the outcome measures from CNSVS. Tukey post hoc analyses were employed when the ombibus tests were significant using the Tukey critical value. Pearson product moment correlations (r) were calculated as a general measure of the strength of the linear association between variables at each of the 3 time points. Intraclass correlation coefficients (ICC$_{2,1}$) with standard error of measurement (SEM) were calculated to determine the consistency of the participants' performances across sessions for each of the outcome measures. We used Fleiss' recommendations[7] for interpreting ICC$_{2,1}$ values: >0.75 = excellent, 0.40 to 0.75 = fair to good, and <0.40 = poor. Reliable change indices (RCIs) provide estimates of the probability that a given difference in a score would not be obtained as a result of measurement error.[12,18] RCIs were calculated using the values of times 2 and 3 in an effort to produce the most stable RCIs possible for the measures. Data were analyzed using SPSS (version 19.0; IBM SPSS Inc). Mean scores and standard deviations were calculated for each outcome measure. An a priori α level of significance was set at 0.05 for all analyses.

Table 1. Practice effects for CNS Vital Signs (standard scores)

CNS Vital Signs Domain	Session, Mean ± SD			F Value	P Value
	Time 1	Time 2	Time 3		
Verbal memory	100.97 ± 16.70	105.97 ± 16.74	99.61 ± 19.93	$F_{(2, 74)} = 2.06$	0.13
Visual memory	101.79 ± 13.72	105.00 ± 14.25	101.08 ± 14.88	$F_{(2, 74)} = 1.58$	0.21
Psychomotor speed	109.29 ± 14.93	112.47 ± 14.16	112.18 ± 14.16	$F_{(2, 74)} = 4.25$	0.02[a,b]
Cognitive flexibility	104.50 ± 18.20	111.97 ± 13.65	116.11 ± 14.19	$F_{(1.8, 72)} = 23.45$	<0.01[a,b]
Complex attention	101.76 ± 16.07	102.54 ± 15.17	103.81 ± 13.59	$F_{(1.9, 70)} = 0.87$	0.42
Processing speed	109.58 ± 13.09	116.39 ± 15.82	116.03 ± 15.55	$F_{(2, 74)} = 8.76$	<0.01[a,b]
Reaction time	101.71 ± 14.61	107.29 ± 12.69	107.54 ± 11.65	$F_{(1.6, 58.1)} = 6.40$	0.01[a,b]
Executive functioning	105.12 ± 17.46	113.05 ± 13.04	116.79 ± 13.48	$F_{(2, 74)} = 27.49$	<0.01[a,b]
Reasoning	101.68 ± 11.38	106.22 ± 11.27	107.72 ± 11.58	$F_{(2, 70)} = 4.66$	0.01[b]

[a]Main effect of time: Session 2 performance was superior to session 1 performance.
[b]Main effect of time: Session 3 performance was superior to session 1 performance.

RESULTS

There were no differences between symptom scores of the participants at any of the sessions (session 1, 3.53 ± 4.41; session 2, 3.42 ± 6.03; session 3, 4.23 ± 7.64; $F_{2, 78} = 0.33$; $P = 0.72$). Less than 3% of the scores were removed because they were invalid based on CNSVS criteria (Table 1).[2] On a number of clinical domains, participants performed significantly better on the second session and/or third session than they did on the first session (Table 1). Participants performed better at time 2 compared with time 1 and at time 3 compared with time 1 on: psychomotor speed (d_{crit} = 2.91), cognitive flexibility (d_{crit} = 4.33), processing speed (d_{crit} = 4.40), and reaction time (d_{crit} = 4.88). In addition, participants performed better at time 3 compared with time 1 on reasoning (d_{crit} = 4.86) and executive functioning (d_{crit} = 3.86). There were no significant differences in scores between the second and third sessions.

Overall, the reliability values (Table 2) were acceptable.[7] Some scores yielded poor or excellent reliability. From time 1 to 2, verbal memory yielded poor reliability ($ICC_{2,1}$ = 0.10), while psychomotor speed yielded excellent reliability ($ICC_{2,1}$ = 0.85). From time 2 to 3, cognitive flexibility ($ICC_{2,1}$ = 0.79), complex attention ($ICC_{2,1}$ = 0.79), processing speed ($ICC_{2,1}$ = 0.76), and executive function ($ICC_{2,1}$ = 0.78) exhibited excellent reliability. To account for changes over time clinically, SEM and RCIs can be used. SEMs are estimates of error, while RCIs are estimates of how much and in what direction an individual's test scores have changed and whether the changes are clinically significant. The SEMs ranged from 5.04 (psychomotor speed at time 2) to 12.08 (verbal memory at time 3). The RCIs ranged from 9.44 (psychomotor speed) to 20.22 (verbal memory).

DISCUSSION

We confirmed our hypothesis that CNSVS would yield acceptable test-retest reliability, with greater reliability between the second and third test administration compared with between the first and second test administration. Overall, our test-retest correlations (0.11-0.87) are similar to those previously reported by Gualtieri and Johnson[8] (0.31-0.88) and Cole et al[3] (0.29-0.79). We found lower correlation values for verbal memory (0.11-0.63), visual memory (0.37-0.57), and reasoning (0.25-0.57). One possible explanation for the discrepancy in findings is differences in the ages and physical activity levels of the participants and differences in the test-retest intervals. Other computerized concussion assessments have reported similar test-retest correlation values ranging from 0.19 to 0.83.[5,18,21] It is important to take the population and test-retest intervals into account when interpreting test results.

Practice effects occurred between the first and second session on some domains. However, there were no practice effects between sessions 2 and 3. This highlights the importance of taking practice effects into account when interpreting scores across serial testing sessions, especially from the initial to the second administration. This may also suggest that a second administration of the test is needed to achieve stable measurements.[10] Practice effects may influence postconcussion assessments. If patients are administered 2 tests within a week of each other, which often occurs during postinjury testing, the lack of a practice effect may actually indicate deficits.[4]

Reliable change indices can be used to account for practice effects and identify meaningful score changes due to pathology (eg, concussion).[16,22] When interpreting scores, clinicians should

Table 2. Test-retest correlations for CNS Vital Signs

CNS Vital Signs Domain	r_{12}^{a}	P_{12}^{a}	ICC^{a}	r_{23}^{b}	P_{23}^{b}	ICC^{b}	r_{13}^{c}	P_{13}^{c}	ICC^{c}	Time 2 SEM	Time 3 SEM	S_{diff}^{b}	RCI 80%	RCI 90%[b]	RCI 95%
Verbal memory	0.11	0.53	0.10	0.63	<0.01	0.59	0.29	0.08	0.41	10.14	12.08	15.77	20.22	25.94	30.91
Visual memory	0.52	<0.01	0.52	0.57	<0.01	0.55	0.37	0.02	0.38	9.38	9.80	13.57	17.49	22.32	26.60
Psychomotor speed	0.87	<0.01	0.85	0.87	<0.01	0.87	0.79	<0.01	0.86	5.04	5.36	7.36	9.44	12.11	14.43
Cognitive flexibility	0.75	<0.01	0.65	0.82	<0.01	0.79	0.82	<0.01	0.64	5.76	5.99	8.30	10.65	13.66	16.28
Complex attention	0.67	<0.01	0.67	0.79	<0.01	0.79	0.63	<0.01	0.62	6.95	6.23	9.33	11.96	15.35	18.29
Processing speed	0.65	<0.01	0.58	0.76	<0.01	0.76	0.74	<0.01	0.67	7.75	7.62	10.87	13.93	17.88	21.30
Reaction time	0.59	<0.01	0.54	0.81	<0.01	0.81	0.57	<0.01	0.51	5.63	5.10	7.60	9.74	12.50	14.90
Executive functioning	0.77	<0.01	0.66	0.80	<0.01	0.78	0.82	<0.01	0.62	5.80	6.00	8.34	10.70	13.73	16.36
Reasoning	0.57	<0.01	0.54	0.45	0.01	0.45	0.25	0.14	0.22	8.50	8.63	12.11	15.52	19.92	23.73

ICC, intraclass correlation coefficient; RCI, reliable change index; SEM, standard error of mean.
[a]Indicates time 1 to time 2.
[b]Indicates time 2 to time 3.
[c]Indicates time 1 to time 3.

take the difference between scores at 2 different time points (eg, baseline and postinjury, or postinjury day 2 and postinjury day 7) and compare them to the RCI. Any change in scores from 1 session to another that exceeds the RCI is believed to be because of some other factor, such as cognitive impairment. The 80% RCIs are the most clinically conservative because they warrant the least amount of change. From the second session to the third session 1 week later, 80% RCIs in our sample ranged from 9.44 (psychomotor speed) to 20.22 (verbal memory). The large difference in the RCIs for each of the clinical domains indicates the importance of looking at the reliability of each of the domains individually, instead of taking a standard cutoff score for all of the domains. Using a standard 95% cutoff, clinicians would expect individuals to change the same amount on verbal memory and psychomotor speed from 1 session to another; however, with RCIs, clinicians would expect individuals to change as much as 20.22 points on verbal memory but only 9.44 points on psychomotor speed. RCIs are better than standard 95% cutoffs because they provide estimates of the measurement error surrounding differences in test-retest scores, which allows for more accurate documentation of deterioration from preinjury testing and recovery during postinjury assessments.[12]

We acknowledge limitations of this study. The time in between sessions may be applicable in postinjury testing sessions but is likely not representative of the time that passes from baseline testing until initial postinjury testing. This study is an initial step in addressing the reliability of the measure. While there are a number of test batteries available, the findings of this study are only applicable to the assessment we examined, CNSVS. Furthermore, we examined healthy individuals; changes over time could be different in an injured sample. We also examined college-aged individuals, and our findings therefore do not apply to younger populations.

CONCLUSION

This study highlights the importance of clinicians understanding the reliability of computerized neurocognitive tests when using them in the evaluation and management of sport-related concussion. The most notable changes in CNSVS scores occurred between the first and second session, with no significant differences between the second and third sessions. This suggests that if the test is administered twice within a small time frame (such as within 1 week), as is often done postconcussion, athletes should be expected to improve on the test. Clinicians should consider using RCIs to account for practice effects and identify meaningful score changes due to pathology. In order to make accurate clinical decisions, clinicians should use clinical judgment and understand the reliability and precision of the tests they are using.

REFERENCES

1. Broglio SP, Macciocchi SN, Ferrara MS. Neurocognitive performance of concussed athletes when symptom free. *J Athl Train*. 2007;42:504-508.
2. CNS Vital Signs, LLC. Frequently asked questions. https://www.cnsvs.com/FAQs.html. Accessed January 9, 2014.
3. Cole WR, Arrieux JP, Schwab K, Ivins BJ, Qashu FM, Lewis SC. Test-retest reliability of four computerized neurocognitive assessment tools in an active duty military population. *Arch Clin Neuropsychol*. 2013;28:732-742.
4. Duff K, Westervelt HJ, McCaffrey RJ, Haase RF. Practice effects, test-retest stability, and dual baseline assessments with the California Verbal Learning Test in an HIV sample. *Arch Clin Neuropsychol*. 2001;16:461-476.
5. Elbin RJ, Schatz P, Covassin T. One-year test-retest reliability of the online version of ImPACT in high school athletes. *Am J Sports Med*. 2011;39:2319-2324.
6. Falleti MG, Maruff P, Collie A, Darby DG. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *J Clin Exp Neuropsychol*. 2006;28:1095-1112.
7. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York, NY: Wiley; 1986.
8. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Arch Clin Neuropsychol*. 2006;21:623-643.
9. Guskiewicz KM, Bruce SL, Cantu RC, et al. National Athletic Trainers' Association Position Statement: management of sport-related concussion. *J Athl Train*. 2004;39:280-297.
10. Hinton-Bayre AD, Geffen G, McFarland K. Mild head injury and speed of information processing: a prospective study of professional rugby league players. *J Clin Exp Neuropsychol*. 1997;19:275-289.
11. Hunt TN, Ferrara MS. Age-related differences in neuropsychological testing among high school athletes. *J Athl Train*. 2009;44:405-409.
12. Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. *Clin Neuropsychol*. 2003;17:460-467.
13. Langlois JA, Rutland-Brown W, Wald MM. The epidemiology and impact of traumatic brain injury: a brief overview. *J Head Trauma Rehabil*. 2006;21:375-378.
14. Lau BC, Collins MW, Lovell MR. Cutoff scores in neurocognitive testing and symptom clusters that predict protracted recovery from concussions in high school athletes. *Neurosurgery*. 2012;70:371-379.
15. Leckie RL, Oberlin LE, Voss MW, et al. BDNF mediates improvements in executive function following a 1-year exercise intervention. *Front Hum Neurosci*. 2014;8:985.
16. Lovell MR, Solomon GS. Neurocognitive test performance and symptom reporting in cheerleaders with concussions. *J Pediatr*. 2013;163:1192.e1-1195.e1.
17. McCrory P, Meeuwisse W, Aubry M, et al. Consensus statement on Concussion in Sport—The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *Phys Ther Sport*. 2013;14:e1-e13.
18. Register-Mihalik JK, Guskiewicz KM, Mihalik JP, Schmidt JD, Kerr ZY, McCrea MA. Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: implications for caution in clinical practice. *J Head Trauma Rehabil*. 2013;28:274-283.
19. Register-Mihalik JK, Kontos DL, Guskiewicz KM, Mihalik JP, Conder R, Shields EW. Age-related differences and reliability on computerized and paper-and-pencil neurocognitive assessment batteries. *J Athl Train*. 2012;47:297-305.
20. Resch JE, McCrea MA, Cullum CM. Computerized neurocognitive testing in the management of sport-related concussion: an update. *Neuropsychol Rev*. 2013;23:335-349.
21. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *Am J Sports Med*. 2010;38:47-53.
22. Schatz P, Sandel N. Sensitivity and specificity of the online version of ImPACT in high school and collegiate athletes. *Am J Sports Med*. 2013;41:321-326.
23. Segalowitz SJ, Mahaney P, Santesso DL, MacGregor L, Dywan J, Willer B. Retest reliability in adolescents of a computerized neuropsychological battery used to assess recovery from concussion. *NeuroRehabilitation*. 2007;22:243-251.
24. Van Kampen DA, Lovell MR, Pardini JE, Collins MW, Fu FH. The "value added" of neurocognitive testing after sports-related concussion. *Am J Sports Med*. 2006;34:1630-1635.
25. Voss MW, Chaddock L, Kim JS, et al. Aerobic fitness is associated with greater efficiency of the network underlying cognitive control in preadolescent children. *Neuroscience*. 2011;199:166-176.