Examination of the Test-Retest Reliability of a Computerized Neurocognitive Test Battery

Yusuke Nakayama, Tracey Covassin, Philip Schatz, Sally Nogle and Jeff Kovan

Am J Sports Med 2014 42: 2000 originally published online June 6, 2014
DOI: 10.1177/0363546514535901

Examination of the Test-Retest Reliability of a Computerized Neurocognitive Test Battery

Yusuke Nakayama,* PhD, ATC, CSCS, Tracey Covassin,*† PhD, ATC, Philip Schatz,‡ PhD, Sally Nogle,* PhD, ATC, and Jeff Kovan,* DO
Investigation performed at Michigan State University, East Lansing, Michigan, USA

Background: Test-retest reliability is a critical issue in the utility of computer-based neurocognitive assessment paradigms employing baseline and postconcussion tests. Researchers have reported low test-retest reliability for the Immediate Post Concussion Assessment and Cognitive Testing (ImPACT) across an interval of 45 and 50 days.

Purpose: To re-examine the test-retest reliability of the ImPACT between baseline, 45 days, and 50 days.

Study Design: Descriptive laboratory study.

Methods: Eighty-five physically active college students (51 male, 34 female) volunteered for this study. Participants completed the ImPACT as well as a 15-item memory test at baseline, 45 days, and 50 days. Intraclass correlation coefficients (ICCs) were calculated for ImPACT composite scores, and change scores were calculated using reliable change indices (RCIs) and regression-based methods (RBMs) at 80% and 95% confidence intervals (CIs).

Results: The respective ICCs for baseline to day 45, day 45 to day 50, baseline to day 50, and overall were as follows: verbal memory (0.76, 0.69, 0.65, and 0.78), visual memory (0.72, 0.66, 0.60, and 0.74), visual motor (processing) speed (0.87, 0.88, 0.85, and 0.91), and reaction time (0.67, 0.81, 0.71, and 0.80). All ICCs exceeded the threshold value of 0.60 for acceptable test-retest reliability. All cases fell well within the 80% CI for both the RCI and RBM, while 1% to 5% of cases fell outside the 95% CI for the RCI and 1% for the RBM.

Conclusion: Results suggest that the ImPACT is a reliable neurocognitive test battery at 45 and 50 days after the baseline assessment. The current findings agree with those of other reliability studies that have reported acceptable ICCs across 30-day to 1-year testing intervals, and they support the utility of the ImPACT for the multidisciplinary approach to concussion management.

Clinical Relevance: This study suggests that the computerized neurocognitive test battery, ImPACT, is a reliable test for postconcussion serial assessments. However, when managing concussed athletes, the ImPACT should not be used as a stand-alone measure.

Keywords: concussion; neurocognitive testing; test-retest reliability

The Centers for Disease Control and Prevention estimates that 1.6 to 3.8 million mild traumatic brain injuries occur in sports and recreational activities among athletes aged 15 to 24 years in the United States (US) annually.[18] Current concussion consensus statements recommend that the management of sport-related concussions be based on a multifaceted approach that includes symptom inventories, balance assessments, and neurocognitive testing.[23] The use of computerized neurocognitive assessments employing preparticipation baseline tests followed by a series of postconcussion tests has become a widely adopted element within the multidisciplinary approach to concussion evaluation and management.[15,23] The advantages of using a computerized neurocognitive test over traditional paper-and-pencil tests are the ease of test administration, increased alternate forms and the ability of the test to detect slight differences in the athletes' performance

†Address correspondence to Tracey Covassin, PhD, ATC, 308 W Circle Drive, Department of Kinesiology, Michigan State University, East Lansing, MI 48824, USA (e-mail: Covassin@msu.edu).
*Department of Kinesiology, Michigan State University, East Lansing, Michigan, USA.
‡Department of Psychology, Saint Joseph's University, Philadelphia, Pennsylvania, USA.

One or more of the authors has declared the following potential conflict of interest or source of funding: P.S. has received funding to study the effect of concussions in high school and college athletes from the International Brain Research Foundation, the Sports Concussion Center of New Jersey, and ImPACT Applications Inc. However, ImPACT Applications Inc had no role in the conceptualization of the current study, the collection or analysis of data, the writing of the article, or the decision to submit for publication.

The American Journal of Sports Medicine, Vol. 42, No. 8
DOI: 10.1177/0363546514535901

(ie, reaction time), and decreased staffing requirements as compared with the individualized administration of paper-based measures.[32] However, there remains a concern over the psychometric properties of these computerized tests[21,24] as well as the interpretation of test data by untrained personnel.[11]

The Immediate Post Concussion Assessment and Cognitive Testing (ImPACT) is a widely used neurocognitive test that evaluates concentration, attention, memory, visual motor speed, and reaction time. High sensitivity (eg, 81.9%,[30] 91.4%,[31] 91.7%[6]) and specificity (eg, 69.1%,[30] 89.4%[6]) for concussions using the ImPACT have been reported, as well as good construct validity[19] and convergent and divergent validity[20] with standardized neuropsychological tests, in samples composed of high school athletes,[30] both high school and college athletes,[31] and college athletes.[6,19]

As for the reliability of the ImPACT, there have been mixed results reported. Some studies have documented adequate reliability of the ImPACT across a wide time span, ranging from 1 month to 2 years. Recently, Cole and colleagues[9] published the test-retest reliability of active military personnel (mean age, 34 years; range, 19-59 years) across a 30-day test period; the intraclass correlation coefficients (ICCs) for the ImPACT ranged from 0.50 to 0.83. Similarly, across a 30-day interval, Schatz and Ferris[28] documented ICC values ranging from 0.59 to 0.88 for ImPACT composite scores in a sample of college student volunteers (mean age not reported; range, 18-22 years). Across a 1-year time span, Elbin et al[12] documented ICC values ranging from 0.62 to 0.85 for ImPACT composite scores in a sample of high school athletes (mean age, 14.8 $\pm$ 0.9 years). At a span of 2 years, Schatz[27] reported ICCs ranging from 0.46 to 0.74 for ImPACT composite scores in a sample of college athletes (mean age, 18.8 $\pm$ 0.6 years).

However, Broglio and colleagues[5] reported low ICCs for all ImPACT composite scores through 3 test administrations—baseline, 45 days after baseline, and 50 days after baseline—in a sample of college student volunteers (mean age, 21.4 $\pm$ 2.8 years). The respective ICCs for each composite score at baseline to day 45 and day 45 to day 50 were as follows: verbal memory (0.23 and 0.40), visual memory (0.32 and 0.39), motor processing speed (0.38 and 0.61), and reaction time (0.39 and 0.51). These low ICC values may have been attributable to methodological problems caused by administering 3 different computerized neurocognitive assessments in succession (ImPACT, Headminder's Concussion Resolution Index, and Concussion Sentinel).[33] Given that the administration of 3 computerized neurocognitive tests in a row may result in mental fatigue and a lack of motivation in participants, Resch and colleagues[25] replicated the Broglio et al[5] study using only the ImPACT. In their study, groups of Irish (mean age, 22.4 $\pm$ 1.9 years) and US student volunteers (mean age, 20.9 $\pm$ 1.7 years) completed the ImPACT as well as an effort test, and no participants were excluded because of suboptimal effort. Subsequent ICCs for days 1 to 45 ranged from 0.26 to 0.84 in the Irish sample and from 0.45 to 0.76 in the US sample, and values for days 45 to 50 ranged from 0.41 to 0.88 in the Irish sample and from 0.40 to 0.71 in the US sample.

The purpose of the current study was to re-examine the test-retest reliability of the ImPACT neurocognitive test battery between baseline, 45 days after baseline, and 50 days after baseline in a sample of physically active college students. We hypothesized that there would be acceptable test-retest reliability (ICC $\geq$ 0.60) in verbal and visual memory composite scores, motor processing speed composite scores, and reaction time composite scores over 3 test administrations (baseline, 45 days, and 50 days).

MATERIALS AND METHODS

Research Design

A repeated-measures design was used to evaluate the test-retest reliability of the ImPACT neurocognitive test battery. The independent variable was the time of test administration (baseline, 45 days after baseline, and 50 days after baseline). We selected these test periods to specifically replicate the time frame of both the Broglio et al[5] and Resch et al[25] studies. The dependent variables were the composite scores of the ImPACT, including verbal memory, visual memory, visual motor (processing) speed, and reaction time. In addition, we used failure to reproduce 12 items from a 15-item memory test used by Rey[26] as an indicator of effort from test takers (replacing the Memory and Concentration Test, used by Broglio et al,[5] and the Green Word Memory Test, used by Resch et al[25]). Thus, dependent variables also included the number of items that test takers reproduced from the Rey[26] 15-item memory test.

Participants

The participants were recruited via word of mouth, advertisements throughout the campus, and announcements in kinesiology and athletic training classes. Thus, the participants were 96 college students who met all 3 inclusionary criteria that satisfy the following categories of basic recommendations for physical activity based on the American College of Sports Medicine (ACSM)[13] guidelines:

1. Cardiorespiratory exercise: at least 150 minutes of moderate-intensity exercise (4.8-7.1 metabolic equivalents of task [METs]) or 75 minutes of vigorous-intensity exercise (7.2-10.1 METs) per week. One continuous session and multiple shorter sessions of at least 10 minutes per session are both acceptable to accumulate the desired amount of exercise.
2. Resistance exercise: 2 to 4 sets of 8 to 12 repetitions on a major muscle group such as the chest, shoulders, back, hips, legs, trunk, and arms 2 or 3 days per week using a variety of exercises and equipment.
3. Neuromotor exercise: 2 or 3 days per week of exercises that involve motor skills (balance, agility, coordination, and gait).

Participants were excluded from the study if they

- were diagnosed with a learning disability, color blindness, attention deficit disorder, psychological disorder, brain surgery, or major neurological condition (ie, demyelinating disease, acute disseminated encephalomyelitis);
- were diagnosed with a concussion within 6 months before or experienced a concussion during the study;
- had a self-reported history of intracranial injuries (eg, subdural hematoma) as determined by a positive computed tomography or magnetic resonance imaging scan; or
- had a primary language other than English.

Instrumentation

Immediate Post Concussion Assessment Cognitive Testing (ImPACT). The computerized neurocognitive assessment program used was the online version of the ImPACT (ImPACT Applications Inc, Pittsburgh, Pennsylvania, USA). The ImPACT program consists of 3 main sections: (1) a demographics section, (2) a postconcussion symptom scale (PCSS) section, and (3) neurocognitive test modules. In the demographics section, participants were asked to input demographic and descriptive information such as years of experience playing sports, history of alcohol and drug use, learning disabilities, attention deficit hyperactive disorder, major neurological disorder, and concussion history. In the PCSS section, participants were asked to self-report a total of 22 concussion-related symptoms based on how they are feeling at the moment using a 7-point Likert scale (0 = not experiencing this symptom and 6 = severely experiencing this symptom). The third category consists of 6 neurocognitive modules that evaluate attention, reaction time, memory, and concentration. Participants were instructed to complete the given tasks as quickly and accurately as possible with their best effort.

15-Item Memory Test. Before administration of the ImPACT, the Rey[26] 15-item memory test was administered to evaluate test takers' efforts. The test consists of 15 symbols on a piece of paper in 5 rows of 3 categorically related symptoms. The symbols are easily identified letters, numbers, or figures. Presented as a memory task (eg, in which 15 items need to be memorized during a 10-second exposure), the instructions are repeated with redundancy, and the stimuli contain easily recognizable patterns of numbers and letters (eg, A, B, C, . . . a, b, c), such that only a few pieces of information need to be remembered. As a simple test of immediate memory and attention, normal performance would be expected in all but the most severely impaired patients. Participants were exposed to the symbols for 10 seconds and then given a 2-minute window to reproduce them, without concern about their order. A score of ≤ 9 was considered below the acceptable effort level.

Procedures

This study received approval from the institutional review board at Michigan State University before recruiting participants, and all participants provided written consent. All participants reported to the same reserved computer laboratory for a total of 3 visits. Only ≤ 5 participants were tested at the same time in the computer laboratory to prevent distraction from large group size testing. All tests were administered by the same certified athletic trainer at the same time of day and place each test session. There were 45 days between the first and second session and 5 days between the second and third session (eg, 50 days between the first and third session). All participants completed the Rey[26] 15-item memory test and the ImPACT neurocognitive test battery at each test session. Each session took approximately 30 minutes to complete.

Data Analysis

Descriptive statistics were used for all tests to determine means $\pm$ standard deviations. SPSS version 19.0 (SPSS Inc, Chicago, Illinois, USA) was used to obtain ICCs for each composite score. Specifically, a 2-way random-effects analysis of variance ICC for absolute agreement was calculated to estimate the test-retest reliability of the ImPACT for baseline to day 45 and day 50 assessments and for day 45 to day 50 assessments. The ICCs range from 0 to 1, with an ICC closer to 1.00 indicating stronger reliability. The ICC model "two-way mixed" type "consistency" was calculated using "average measures."[34] There are a variety of recommendations for ICC interpretation. Several researchers suggest an ICC of 0.60 to be minimally acceptable.[1,3,8] Moreover, Anastasi[1] suggested that the test-retest reliability for clinical decision making over a 45-day period has a minimum ICC value of 0.60. Therefore, in the present study, an ICC of 0.60 was used as the minimum acceptable ICC value. Test takers who reproduced fewer than 9 items on the Rey[26] memory test were considered to be providing less than the optimal effort,[4,14] which indicated a lack of effort. Reliable change indices (RCIs)[17] were calculated to assess whether a change between repeated assessments was reliable and meaningful. The RCI provides an estimate of the probability that a given difference in the score would not be obtained as a result of a measurement error[16]; a modified RCI formula,[7] which includes an adjustment for practice effects, was utilized (see Barr[2] for a more detailed discussion). Finally, regression-based methods (RBMs) were applied to the data. With an RBM, the scores from the first assessment are placed into a regression analysis, using the score at time 2 as the dependent variable, with the resulting equation providing an adjustment for the effect of the initial performance level as well as controlling for any regression to the mean.[22] With this technique, regression equations can be built to predict a participant's level of performance on a neuropsychological instrument at retest from the initial testing.[10] The percentage of cases falling within 80% and 95% confidence intervals (CIs) was documented. Statistical significance was set a priori at $P = .05$.

RESULTS

Demographic Data

There were a total of 96 participants in this study; however, data from 11 participants were excluded from further

TABLE 1
ICCs for ImPACT Composite Scores Between Baseline and Day 45, Day 45 and Day 50, Baseline and Day 50, and Overall[a]

	ImPACT Composite Score			
Interval	Verbal Memory	Visual Memory	Motor Processing Speed	Reaction Time
Baseline to day 45	0.76 (0.63-0.85)	0.72 (0.59-0.82)	0.87 (0.78-0.92)	0.67 (0.52-0.79)
Day 45 to day 50	0.69 (0.53-0.80)	0.66 (0.46-0.78)	0.88 (0.81-0.92)	0.81 (0.73-0.89)
Baseline to day 50	0.65 (0.45-0.78)	0.60 (0.38-0.73)	0.85 (0.72-0.91)	0.71 (0.57-0.81)
Overall	0.78 (0.69-0.85)	0.74 (0.64-0.83)	0.91 (0.86-0.94)	0.80 (0.73-0.87)

[a]Values in parentheses are 95% confidence intervals. ICC, intraclass correlation coefficient; ImPACT, Immediate Post Concussion Assessment and Cognitive Testing.

TABLE 2
ImPACT Composite Scores at Baseline, 45 Days After Baseline, and 50 Days After Baseline[a]

	ImPACT Composite Score			
	Verbal Memory	Visual Memory	Motor Processing Speed	Reaction Time
Baseline	85.55 ± 10.37	76.36 ± 11.91	41.26 ± 6.07	0.58 ± 0.08
45 days after baseline	88.22 ± 10.21	78.00 ± 13.20	42.61 ± 6.22	0.57 ± 0.07
50 days after baseline	89.20 ± 8.59	73.21 ± 12.80	43.00 ± 5.99	0.58 ± 0.07

[a]Values are expressed as mean ± standard deviation. ImPACT, Immediate Post Concussion Assessment and Cognitive Testing.

analysis for various reasons including invalid baseline assessments (as "flagged" by the ImPACT) (n = 1; <1%), a value >30 on impulse control scores (n = 2; 1%), a value >0.80 on reaction time (n = 3; 2%), not maintaining the required physical activity level (n = 2; 1%), and a technical problem during testing (n = 3; 2%). Therefore, the resultant sample was composed of 85 participants (51 male, 34 female). The majority of participants were white (75%), followed by African American (6%), Hispanic or Latino (3%), and Asian American (2%), and 2% of participants reported their ethnicity as "other" (the remaining individuals did not report their race). A history of concussions was reported by 18 participants, with a mean of 1.89 ± 1.08 previous concussions that ranged from 1 to 4 previous concussions; however, no participants sustained a concussion during the testing period or in the 6 months before participating in this study.

Rey Memory Test

All participants achieved a perfect score (15 points) on the Rey[26] 15-item memory test, across all 3 test sessions, with the exception of 1 participant, who scored 12 points on the second administration. In this regard, all participants provided optimal effort.

ICCs for ImPACT Composite Scores at Baseline, 45 Days, and 50 Days

The ICC values for all ImPACT composite scores were ≥0.60, which indicates acceptable test-retest reliability. Among the composite scores, visual motor (processing) speed reported the highest overall ICC value with 0.91, and visual memory reported the lowest overall ICC value with 0.74 (Table 1). Specifically, the highest ICC value (0.88) was reported for visual motor (processing) speed between day 45 and day 50, and the lowest ICC value (0.60) was reported for visual memory between baseline and day 50. With the exception of verbal and visual memory, the ICC values from day 45 to day 50 were higher than the ICC values from baseline to day 45 and from baseline to day 50. The means ± standard deviations for each ImPACT composite score can be found in Table 2.

RCI and RBM

The RCI and RBM were calculated using ImPACT composite scores from day 1 and day 45. All scores fell within the 80% CI using both the RCI and RBM. When change scores were restricted to the 95% CI, 1% to 5% of cases fell outside the CI for the RCI, but only 1% of cases fell outside the CI for the RBM (Table 3).

DISCUSSION

The primary purpose of the current study was to re-examine the test-retest reliability of the ImPACT neurocognitive test battery between baseline, 45 days after baseline, and 50 days after baseline in physically active college students while assessing the participants' self-reported physical and mental conditions on the day of testing. The ICC values for each composite score from the current study met or

TABLE 3
Rates of Impairment Using RCI Versus RBM[a]

	ImPACT Composite Score			
	Verbal Memory	Visual Memory	Motor Processing Speed	Reaction Time
RCI				
80% CI				
Improvement	10	9	13	7
Decline	1	2	2	8
Total	11	11	15	15
95% CI				
Improvement	8	3	8	2
Decline	1	2	2	4
Total	9	5	10	6
RBM				
80% CI				
Improvement	3	3	7	5
Decline	7	7	7	7
Total	10	10	14	12
95% CI				
Improvement	1	1	0	5
Decline	5	5	2	1
Total	6	6	2	6

[a]Values represent the percentage of participants scoring beyond cut-off values at 80% (1.65) and 95% (1.96) CIs. ImPACT, Immediate Post Concussion Assessment and Cognitive Testing; RBM, regression-based method; RCI, reliable change index.[7]

exceeded the 0.60 level, which is the indication of acceptable test-retest reliability,[3] and were higher than the ICC values obtained by Broglio et al[5] and Resch and colleagues.[25]

While the current findings are consistent with those of other studies,[12,28] they are inconsistent with those of Brogio et al[5] and Resch and colleagues,[25] who documented much lower test-retest reliability coefficients. There are several factors that may contribute to the differences between these 3 studies. First, there is a notable difference in the primary reason for data exclusion between the Broglio et al[5] study and the current study. While invalid ImPACT baseline assessments accounted for 64% of the data exclusion (29/45 exclusions) in the study of Broglio et al,[5] invalid baseline assessments (ie, flagged by the ImPACT as invalid) accounted for less than 3% (1/39) of the data exclusion in the current study. When we considered that outliers, who were participants with values >30 on impulse control and >0.80 on reaction time, were considered as reasons for data exclusion (n = 5), test takers' performance-related reasons (ie, invalid baseline assessment, >30 on impulse control, >0.80 on reaction time) accounted for <5% (6/127) of the total cases excluded in the current study. It is important to note that participants in the Broglio et al[5] and Resch et al[25] studies completed "desktop" versions (versions 4.5 and 6.7, respectively) of the ImPACT, which were programmed using FoxBase Pro. Participants in the current study completed the online version of the ImPACT, which was programmed in Flash. Research has shown considerably fewer invalid baseline assessments using the online version (4.1% of college athletes) compared with the desktop version (10.2% of college athletes).[29]

However, the relatively high number of invalid baseline tests by Broglio et al[5] calls into question the quality of the test taker's effort. While both the current study and that of Resch and colleagues[25] utilized only the ImPACT and a lone validity test, Broglio and colleagues[5] administered 3 complete batteries along with a validity measure. It is possible to infer that their participants were not able to provide their best performance because of fatigue, stress, and interference effects caused by 3 consecutive administrations of neurocognitive tests. However, even after excluding "invalid" cases, the resultant sample may reflect high levels of variability due to these confounders.

Inspection of RCI and RBM data reveals that no case fell outside expected values using 80% CIs, and 1% to 5% of cases fell outside 95% CIs using the RCI. However, only 1% of cases fell outside the 95% CI using the RBM. These data are similar to RCI and RBM data at 30-day,[28] 1-year,[12] and 2-year[27] intervals in which all cases fell within 80% CIs and only a small percentage of cases fell beyond 95% CIs.

While the utilization of computerized neurocognitive assessments employing preparticipation baseline tests followed by a series of postconcussion tests has become a cornerstone element of the multidisciplinary approach to concussion evaluation and management,[23] there have been ongoing discussion and debate[21,33] on the test-retest reliability of computerized neurocognitive tests among clinicians and researchers. The current findings support the test-retest reliability of one of the most widely used computerized neurocognitive test batteries, ImPACT, by providing acceptable ICC values over 45 and 50 days. Moreover, the current findings are in agreement with those of other test-retest studies,[12,27] which suggest that the ImPACT is a reliable tool within the multidisciplinary approach to concussion management. However, clinicians must be aware that computerized neurocognitive assessments are not intended to be a stand-alone diagnostic measure and should be administered in a quiet environment with very few athletes in the room.[23] In addition, the current results suggest that college "volunteers" completing the ImPACT in research experiments (ie, as part of human participant pools) may be inherently less interested or motivated than athletically active students or varsity athletes.

When clinicians utilize the ImPACT for concussion management including return-to-play decisions by comparing baseline and postconcussion data, they should feel more confident in knowing that the ImPACT has been found to have good test-retest reliability. In the past, clinicians had difficulty accepting that the changes that athletes exhibited on the ImPACT after a concussion were truly caused by effects of a concussion rather than an inherent variability in the ImPACT.

As with all studies, this study is not without its limitations. With respect to the participants studied in the current study, the most limiting factor is that the participants were physically active college students and not athletes. The ImPACT is utilized with high school,

college, and professional athletes rather than the typical physically active college students. Even though the participants were physically active based on the ACSM recommendations, caution should be taken when attempting to generalize the findings to competitive athletes. Another limitation of the current study involves the validity of self-reported data including the participants' mental/physical conditions on testing days and their physical activity levels.

This research utilized a similar methodology to that of Resch and colleagues,[25] but used a different validity measure, and was similar to that of Broglio et al,[5] but used only a single neurocognitive measure. Thus, the ImPACT has acceptable test-retest reliability when it is administered at baseline, 45 days after baseline, and 50 days after baseline, which supports the utilization of the ImPACT in concussion management as part of a multidisciplinary approach. Future research should be conducted using the same research design with an athletic population and different age groups to increase external validity.

REFERENCES

1. Anastasi A, ed. *Psychological Testing*. 6th ed. New York: Macmillan; 1998.
2. Barr WB. Neuropsychological testing for assessment of treatment effects: methodologic issues. *CNS Spectr*. 2002;7(4):300-302, 304-306.
3. Baumgartner T, Chung H. Confidence limits for intraclass reliability coefficients. *Meas Phys Educ Exerc Sci*. 2001;5(3):179-188.
4. Bernard LC, Fowler W. Assessing the validity of memory complaints: performance of brain-damaged and normal individuals on Rey's task to detect malingering. *J Clin Psychol*. 1990;46(4):432-436.
5. Broglio S, Ferrara M, Macciocchi S, Baumgartner T, Elliott R. Test-retest reliability of computerized concussion assessment programs. *J Athl Train*. 2007;42(4):509-514.
6. Broglio S, Macciocchi S, Ferrara M. Sensitivity of the concussion assessment battery. *Neurosurgery*. 2007;60(6):1050-1058.
7. Chelune GJ, Naugle RI, Lüders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology*. 1993;7:41-52.
8. Chinn S. Repeatability and method comparison. *Thorax*. 1991;46:454-456.
9. 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(7):732-742.
10. Crawford JR, Garthwaite PH. Comparing patients' predicted test scores from a regression equation with their obtained scores: a significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology*. 2006;20(3):259-271.
11. Echemendia RJ, Herring S, Bailes J. Who should conduct and interpret the neuropsychological assessment in sports-related concussion? *Br J Sports Med*. 2009;43 Suppl 1:i32-i35.
12. Elbin R, 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(11):2319-2324.
13. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011;43(7):1334-1359.
14. Goldberg JO, Miller HR. Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *J Clin Psychol*. 1986;42(5):792-795.
15. Guskiewicz KM, Bruce SL, Cantu R, et al. National Athletic Trainers' Association position statement: management of sports-related concussion. *J Athl Train*. 2004;39(3):280-297.
16. Iverson GL, Sawyer DC, McCracken LM, Kozora E. Assessing depression in systemic lupus erythematosus: determining reliable change. *Lupus*. 2001;10(4):266-271.
17. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59(1):12-19.
18. Langlois J, Rutland-Brown W, Wald M. The epidemiology and impact of traumatic brain injury: a brief overview. *J Head Trauma Rehabil*. 2006;21(5):375-378.
19. Maerlender A, Flashman L, Kessler A, et al. Discriminant construct validity of ImPACT™: a companion study. *Clin Neuropsychol*. 2013;27(2):290-299.
20. Maerlender A, Flashman L, Kessler A, et al. Examination of the construct validity of ImPACT computerized test, traditional, and experimental neuropsychological measures. *Clin Neuropsychol*. 2010;24(8):1309-1325.
21. Mayers LB, Redick TS. Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: psychometric issues. *J Clin Exp Neuropsychol*. 2012;34(3):235-242.
22. McCrea M, Barr WB, Guskiewicz K, et al. Standard regression-based methods for measuring recovery after sport-related concussion. *J Int Neuropsychol Soc*. 2005;11(1):58-69.
23. McCrory P, Meeuwisse W, Aubry M, et al. Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport, Zurich, November 2012. *J Athl Train*. 2013;48(4):554-575.
24. Randolph C, McCrea M, Barr W. Is neuropsychological testing useful in the management of sport-related concussion? *J Athl Train*. 2005;40(3):139-152.
25. Resch J, Driscoll A, McCaffrey N, et al. ImPact test-retest reliability: reliably unreliable? *J Athl Train*. 2013;48(4):506-511.
26. Rey A. L'examen psychologie dans las cas d'encephalopathie traumatique. *Arch Psychol*. 1941;23(112):286-340.
27. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *Am J Sports Med*. 2010;38(1):47-53.
28. Schatz P, Ferris CS. One-month test–retest reliability of the ImPACT test battery. *Arch Clin Neuropsychol*. 2013;28(5):499-504.
29. Schatz P, Moser RS, Solomon GS, Ott SD, Karpf R. Prevalence of invalid computerized baseline neurocognitive test results in high school and collegiate athletes. *J Athl Train*. 2012;47(3):289-296.
30. Schatz P, Pardini J, Lovell M, Collins M, Podell K. Sensitivity and specificity of the ImPACT test battery for concussion in athletes. *Arch Clin Neuropsychol*. 2006;21:91-99.
31. 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(2):321-326.
32. Schatz P, Zillmer EA. Computer-based assessment of sports-related concussion. *Appl Neuropsychol*. 2003;10(1):42-47.
33. Schatz S, Kontos A, Elbin R III. Response to Mayers and Redick. ''Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: psychometric issues.'' *J Clin Exp Neuropsychol*. 2012;34(4):428-434.
34. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240.