One-Year Test-Retest Reliability of the Online Version of ImPACT in High School Athletes

R.J. Elbin,*† PhD, Philip Schatz,‡ PhD, and Tracey Covassin,§ PhD, ATC
Investigation performed at Saint Joseph's University, Philadelphia, Pennsylvania

Background: The ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing) neurocognitive testing battery is a popular assessment tool used for concussion management. The stability of the baseline neurocognitive assessment is important for accurate comparisons between postconcussion and baseline neurocognitive performance. Psychometric properties of the recently released online version of ImPACT have yet to be established; therefore, research evaluating the reliability of this measure is warranted.

Purpose: The authors investigated the 1-year test-retest reliability of the ImPACT online version in a sample of high school athletes.

Study Design: Case series; Level of evidence, 4

Methods: A total of 369 varsity high school athletes completed 2 mandatory preseason baseline cognitive assessments approximately 1 year apart as required by their respective athletics program. No diagnosed concussion occurred between assessments.

Results: Intraclass correlation coefficients (ICCs) for ImPACT online indicated that motor processing speed (.85) was the most stable composite score, followed by reaction time (.76), visual memory (.70), and verbal memory (.62). Unbiased estimates of reliability were consistent with ICCs: motor processing speed (.85), reaction time (.76), visual memory (.71), and verbal memory (.62).

Conclusion: The online ImPACT baseline is a stable measure of neurocognitive performance across a 1-year time period for high school athletes. These reliability data for online ImPACT are higher than the 2-year ICCs previously reported from the desktop version.

Clinical Relevance: It is recommended that the ImPACT baseline assessment (both desktop and online) continue to be updated every 2 years. The online version of ImPACT appears to be a stable measure of neurocognitive performance over a 1-year period, and systematic evaluation of its stability over a 2-year period is warranted.

Keywords: concussion; neurocognitive testing; reliability; ImPACT

The management of sport-related concussion has evolved considerably over the past decade. Position statements and consensus reports have consistently recommended an objective, multifaceted approach to concussion evaluation and management.[2,12,24,25] These reports have also emphasized the utility of neurocognitive assessment as an objective measure to aid the sports medicine professional in making return-to-play decisions. More specifically, serial neurocognitive assessments that include baseline (ie, preinjury) and postinjury tests have been deemed best practice for monitoring the cognitive recovery of a concussed athlete.[20]

The value of neurocognitive assessment for tracking cognitive recovery after sports-related concussion was first utilized in the multicenter Virginia Football Study,[1,4] in which paper-and-pencil neurocognitive tests were administered to assess changes in cognitive function from preseason baseline to postconcussion. This within-groups research design is now widely adopted, as it allows athletes to serve as their own controls, which has made managing concussion more individualized and specific to each concussed athlete. This model has been evaluated and supported by many researchers as an effective method of assessing cognitive impairment following a concussion.[14,15,32] However, the serial administrations of neurocognitive tests following concussion has highlighted the need to control for practice effects, especially when using tests that were not developed for these purposes.

*Address correspondence to R.J. Elbin, PhD, Department of Kinesiology, Leisure and Sport Science, East Tennessee State University, Box 70654, Johnson City, TN 37614 (e-mail: elbinr@etsu.edu).

†East Tennessee State University Department of Kinesiology, Leisure and Sport Science, Johnson City, Tennessee.

‡Saint Joseph's University Department of Psychology, Philadelphia, Pennsylvania.

§Michigan State University Department of Kinesiology, East Lansing, Michigan.

The authors declared that they have no conflicts of interest in the authorship and publication of this contribution.

The American Journal of Sports Medicine, Vol. 39, No. 11
DOI: 10.1177/0363546511417173

Furthermore, traditional paper-and-pencil neurocognitive tests are quite costly and time-consuming for sports medicine practitioners testing entire teams or athletic programs. To this end, computerized versions of neurocognitive assessment have been made available, which has decreased practice effects, increased test-retest reliability, and improved scoring and administration of these test batteries.[9]

The Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) battery is one of the most widely used computerized neurocognitive test batteries for concussion management. This tool minimizes practice effects through the use of multiple versions and has been evaluated for test-retest reliability. Specifically, Iverson et al[16] found reliability coefficients for ImPACT composite scores ranging from .67 to .85 over a 7-day test-retest timespan. Broglio et al[5] examined test-retest reliability of ImPACT over the course of 45- and 50-day retest intervals, and reported intraclass correlation coefficients (ICCs) ranging from .23 to .38 and .39 to .61, respectively. These low ICC values may have been attributable to methodologic problems caused by administering 3 different computerized neurocognitive assessments in 1 test session. In contrast to these findings, Miller et al[26] found ImPACT to be stable over a 4-month time period in football players who were exposed to repetitive contact over the course of 1 season.

There are no consistent or evidence-based guidelines regarding how often baseline neurocognitive tests should be updated or readministered. As a result, there are conflicting and confusing recommendations in the literature: some researchers have suggested that baseline assessments should be readministered after a concussion and at the start of a new academic year,[8,33] and others recommend an updated baseline every 2 to 3 years in younger athletes (13-17 years of age) to account for cognitive maturation.[9,33] Clearly, research evaluating the long-term reliability of baseline ImPACT assessments has been warranted. Schatz[29] evaluated the long-term test-retest reliability of ImPACT in a sample of 95 collegiate athletes who completed the ImPACT baseline assessment approximately 2 years apart without a concussion between the assessment periods. The ICCs for the composite scores were as follows: verbal memory, .46; visual memory, .65; motor processing speed, .74; reaction time, .68; and total symptom score, .43. Regression-based methods and reliable change indices revealed that less than 7% of the athletes showed a clinically significant change on 1 composite score and only 5% to 10% of athletes demonstrated a clinically significant change in symptom reporting. These results suggest that baseline ImPACT data remain stable over a 2-year time period in collegiate athletes, so annual re-evaluations may not be necessary.

The aforementioned reliability studies were conducted using the ImPACT personal computer–based desktop version that was first released in 2000. More recently, ImPACT launched an online-only version in 2008 to address inefficiencies in program installation, data storage, and data encryption/security. This online version is very similar to the desktop version of ImPACT, but the 2 systems are not identical.[19] The online version is programmed in Flash, and runs through a Web browser, whereas the desktop version was programmed in FoxBase Pro, and runs as a native program in Windows. From a recordkeeping perspective, the online version stores all data in a remote, password-protected database. In contrast, the desktop version stores data on a local computer or network drive. From a user-input perspective, most notably, the desktop version of ImPACT utilizes mouse button input for the Xs and Os choice reaction time task. This subscale contributes to the motor processing speed and impulse control composite scores, often resulting in "invalid" baseline data because of left-right confusion.[27] The new online version employs keyboard input on the Xs and Os task, with the intention of decreasing left-right mouse confusion.

Currently there are no data published on the stability of the online ImPACT neurocognitive test battery. Given the change from desktop to online, the stability of the online baseline version of ImPACT should be assessed and compared with previous reliability data of the desktop version. Therefore, the purpose of this study was to examine the long-term test-retest reliability of baseline assessments collected using the online ImPACT test battery, as a comparison with similar data collected using the desktop version.[29]

METHODS

Participants

Participants were varsity high school athletes, ages 13 to 18 years (mean, 14.8 years), participating in a variety of sports (see Table 1). Anonymous, deidentified data were obtained from the ImPACT test developers for the purpose of psychometric evaluation, and Institutional Review Board approval was obtained for secondary data analysis. All athletes completed 2 mandatory preseason baseline cognitive assessments as required by their athletics program. Assessments were administered in high school computer laboratories in groups of up to 20 to 25 athletes, and were supervised by either an athletic trainer or a member of the school's medical staff.

A total of 484 athletes completed preseason baseline assessments approximately 1.2 years apart (range, 0.5-2.35 years). One hundred fifteen athletes were excluded from the main analysis: 50 (10.3%) were not high school athletes, 15 (3.1%) completed baselines less than 6 months apart, and 12 (2.5%) were non–native English speakers. An additional 38 (7.9%) were excluded on the basis of invalid baselines; this rate is consistent with other high school samples using ImPACT (desktop version, 8.7%[31]; online version, 6.3%[30]) and below the rate for high school football players using paper-based tests (12%[13]). The resultant sample was composed of 369 athletes. Sex distribution was nearly equal, with 168 males (46%) and 201 females (54%) participating in the study. History of concussion was reported by 6.2% of the athletes, diagnosis of attention deficit disorder was reported by 4.1% of athletes, and diagnosis of learning disorder was reported by 2.2% of athletes.

TABLE 1
Demographics

Characteristic	n (%)
Sex	
Male	168 (45.5)
Female	201 (54.5)
Total	369 (100)
Sport	
Soccer	72 (19.5)
Football	68 (18.4)
Volleyball	42 (11.4)
Basketball	35 (9.5)
Lacrosse	33 (8.9)
Field hockey	24 (6.5)
Softball	19 (5.1)
Cheerleading	14 (3.8)
Track/cross country	12 (3.3)
Wrestling	11 (3)
Baseball	11 (3)
Tennis	6 (1.6)
Other/unknown	5 (1.4)
Swimming	5 (1.4)
Track and field	4 (1.1)
Skiing	3 (0.8)
Golf	2 (0.5)
Ice hockey	2 (0.5)
Rugby	1 (0.3)
Age, y[a]	14.8 ± 0.9
Time between baselines, y[a]	1.2 ± 0.4

[a]Data presented as mean ± standard deviation.

Materials

The ImPACT online version was used for the study. The ImPACT instrument is a computer-based program used to assess neurocognitive function and concussion symptoms. It consists of 6 tests that evaluate attention, working memory, and processing speed, yielding composite scores on the areas of verbal memory, visual memory, motor processing speed, reaction time, and impulse control (see Iverson et al[14-16] for more detail on the subscales of ImPACT).

Design and Analyses

Participants completing 2 baseline assessments composed a within-subjects sample, allowing for the comparison between time 1 and time 2. Dependent measures included the verbal memory, design memory, motor processing speed, and reaction time composite scores from ImPACT, and the total symptom scores.

Pearson product moment correlations (r) were calculated as a general measure of the strength of linear association between variables at times 1 and 2. A weakness of the Pearson r as a measure of test-retest reliability is when coefficients are high and group means are similar, but there is considerable individual variation in scores from time 1 to time 2.[28] As such, ICCs, considered a better measure of association than the Pearson r,[34] were also calculated as the primary indicator of test-retest reliability. The ICC can distinguish those sets of scores that are merely ranked in the same order from test to retest from those that are not only ranked in the same order but are in low, moderate, or complete agreement.[7] The ICC model "2-way mixed" type "consistency" was used; ICC analyses also yield an unbiased estimate of reliability, which reflects the consistency of the baseline assessments.[11]

Reliable change indices (RCIs)[18] 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 score would not be obtained as a result of measurement error.[17] A modified RCI formula,[6] which includes an adjustment for practice effects, was also calculated (see Barr[3] for a more detailed discussion). Finally, regression-based methods (RBMs) were applied to the data. In RBMs, 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 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] Following Bonferroni correction, alpha level for all analyses was set at $P < .01$.

RESULTS

Pearson r correlations between baseline assessments ranged from .40 to .74. Intraclass correlation coefficients reflected higher reliability than Pearson r, across all measures. Motor processing speed scores showed the most stability (mean ICC = .85; .82 to .88 lower and upper 95% confidence intervals [CIs]), followed by reaction time (.76; 95% CI, .71-.85), visual memory (.70; 95% CI, .64-.76), verbal memory (.62; 95% CI, .53-.70), and total symptom scores (.57; 95% CI, .47-.65). Unbiased estimates of reliability were consistent with ICCs: motor processing speed, .85; reaction time, .76; visual memory, .71; verbal memory, .62; and total symptom scores, .57. Mean ImPACT composite and symptom scores showed significant improvement between the 2 assessments on visual memory, motor processing speed, and reaction time indices (Table 2).

Regression-based measures using 80% and 95% CIs assume that 80% and 95% of cases will fall within these ranges. Results showed that follow-up baseline scores showed considerable stability (Table 3). All scores from follow-up baseline assessments fell within an 80% CI. Specifically, 88% to 91% of follow-up baseline composite scores and 86% of follow-up symptoms scale scores were within this 80% CI. Nearly all scores from follow-up baseline assessments fell within a 95% CI; only scores from visual memory (1.0%), reaction time (0.2%), and symptom scores (2.4%) fell outside of the cut-off.

Reliable change indices were calculated for all composite and total symptom scores. The RCIs are presented at 80% and 95% CIs (Table 4). There was little variation between the "traditional" method[18] (lower line) and Chelune's "corrected" method[6] (top line). Rates of impairment

TABLE 2
Test-Retest Reliability[a]

	Time 1[b]	Time 2[b]	r	ICC	95% CI, Lower	95% CI, Upper	UER	t^c	Sig.	d
Verbal memory	85.6 ± 9.1	86.4 ± 9.1	.45	.619	.532	.689	.621	−1.6	.11	.08
Visual memory	72.0 ± 12.7	75.5 ± 14.0	.55	.703	.636	.758	.705	−5.3	.001	.28
Motor speed	37.5 ± 6.7	39.8 ± 6.8	.74	.851	.818	.879	.852	−9.1	.001	.47
Reaction time	.59 ± .08	.56 ± .07	.62	.761	.707	.851	.762	7.2	.001	1.80
Symptom scale	4.7 ± 8.5	4.4 ± 8.1	.40	.569	.471	.649	.571	0.6	.55	.40

[a]ICC, intraclass correlation coefficient; CI, confidence interval; UER, unbiased estimate of reliability; Sig, significance (*P*).
[b]Shown as mean ± standard deviation.
[c]*df* = 368; Bonferroni-corrected alpha, $P < .01$.

TABLE 3
Regression-Based Methods[a]

Variable	Time 1[b]	Time 2[b]	α	β	Sxy	80% CI[c]	95% CI[c]
Verbal memory	85.6 ± 9.1	86.4 ± 9.1	47.97	.449	8.114	91.3%	96.2%
Visual memory	72.0 ± 12.7	75.5 ± 14.0	32.05	.603	11.783	88.1%	94.0%
Motor speed	37.5 ± 6.7	39.8 ± 6.8	11.28	.760	4.575	87.8%	95.7%
Reaction time	.59 ± .08	.56 ± .07	0.224	.578	0.056	88.1%	94.8%
Symptom scale	4.7 ± 8.5	4.4 ± 8.1	2.601	.381	7.466	85.6%	92.6%

[a]α, intercept; β, slope; Sxy, standard error of estimate (Crawford and Garthwaite[10]).
[b]Shown as mean ± standard deviation.
[c]CI, confidence interval; numbers represent the percent of participants with change scores within cut-off (80% CI = 1.65; 90% CI = 1.96).

using RBMs and RCIs are presented in Table 5. After a time interval of approximately 1 year, only a small percentage of participants' scores fell outside the range of normality, as denoted by these techniques. The RBM served as a more conservative measure, with fewer follow-up baseline scores falling in the "impaired" range, as compared with RCIs. Of note, nearly equal numbers of participants showed improvement and declines over the time period, across all measures.

DISCUSSION

The primary purpose of the current study was to examine the long-term test-retest reliability of baseline assessments collected using the online version of the ImPACT test battery. The results from the current study collectively suggest that online ImPACT baseline composite scores are quite stable across a 1-year time period for high school athletes. Recent consensus and position statements have suggested the readministration of neurocognitive baseline assessments to account for potential changes in neurocognitive performance during the athletes' career or continued cognitive development and maturation in adolescent athletes (ie, high school athletes).[21,23] The findings of the current study reveal that the stability data (ie, ICCs) from the online version surpass those measured over a 2-year interval using the desktop ImPACT test battery in a sample of collegiate athletes.[29] However, it is important to note that the current data represent stability over a 1-year period. In addition, nearly half as many athletes demonstrated "reliable" change across repeated baselines over a 1-year period (ie, RCI data, improving and declining) in the current study, as compared with data reported by Iverson et al[16] across a 7-day time period (using the desktop version in a mixed sample of high school and university athletes). It is important to note that the current data obtained over an interval of approximately 1 year (online) compare favorably to data obtained over 1 week (desktop), speaking to the stability of the online version. The current study supports the test-retest reliability of the new online version of the ImPACT test battery and provides additional data on the long-term stability of this neurocognitive measure.

Although baseline data from collegiate and high school athletes reflect stability across a 1- and 2-year time period, using both the desktop and online version of ImPACT, these scores are not without degradation or change. Therefore, it is recommended that high school and collegiate athletes receive updated baselines every 2 years. Given the lack of data over a 2-year period, when resources permit, high school athletes may benefit from updated baselines every year.

This study is not without limitations. First, while the sample is comparably larger than any previous test-retest study using ImPACT, the present sample is composed of only high school athletes. As such, these results cannot be uniformly generalized to collegiate or professional athletes. Second, this study examined test-retest reliability over a 1-year interval. While one would expect increased stability over a shorter interval, future research would need to illuminate the stability of this instrument over shorter or longer time periods. Finally, the sample

TABLE 4
Reliable Change Indices (RCIs)[a]

Variable	r	SE[1]	SE[2]	Sdiff[b]	RCI[c]	80% CI[d]	95% CI[d]	Δ80%[e]	Δ95%[e]
Verbal memory	.30	6.73	6.74	9.53	8.45	12.19	15.62	16	19
				9.52	8.46	12.18	15.61	16	19
Visual memory	.55	8.56	9.48	12.77	27.50	16.35	20.95	22	25
				12.11	29.00	15.50	19.86	20	24
Motor speed	.74	3.38	3.48	4.84	46.19	6.20	7.94	6.0	9.5
				4.78	47.79	6.12	7.84	5.9	9.3
Reaction time	.62	.047	.044	0.065	37.60	0.083	0.106	.11	.13
				0.067	36.46	0.085	0.109	.11	.13
Symptom scale	.40	6.60	6.31	9.13	3.11	11.69	14.98	16	18
				9.33	3.05	11.95	15.31	16	19

[a] r, Pearson correlation between time 1 and time 2 scores; SE, standard error of measure at time 1 (SE[1]) and time 2 (SE[2]) (SD*√[1–r_{xy}]); SD, standard deviation.
[b] Sdiff, standard error of difference scores based on Chelune et al,[6] upper row, √([SEM1^2]+[SEM2^2]); and Jacobson and Truax,[18] lower row, √(2*[SEM1^2]).
[c] Reliable change index based on Chelune et al,[6] upper row; and Jacobson and Truax,[18] lower row.
[d] CI, confidence interval; numbers represent reliable change scores at 80% (1.65) and 95% (1.96) CIs.
[e] Absolute point change required for reliable change at 80% and 95% range.

TABLE 5
Rates of Impairment Using Reliable Change Indices (RCIs) Versus Regression-Based Model (RBM)[a]

	RCI[b]						RBM					
	80% CI			95% CI			80% CI			95% CI		
Variable	Impr	Decl	Tot	Impr	Decl	Tot	Impr	Decl	Tot	Impr	Decl	Tot
Verbal memory	6%	4%	10%	3%	2%	5%	1%	7%	8%	0%	4%	4%
	6%	4%	11%	3%	2%	5%						
Visual memory	7%	3%	10%	4%	1%	5%	5%	7%	12%	1%	5%	6%
	9%	4%	13%	4%	2%	6%						
Motor speed	11%	2%	13%	7%	1%	8%	6%	6%	12%	2%	2%	4%
	11%	1%	12%	7%	1%	8%						
Reaction time	8%	2%	10%	6%	2%	8%	7%	4%	1%	3%	2%	5%
	8%	2%	10%	6%	2%	8%						
Symptom scale	4%	4%	8%	4%	4%	8%	12%	2%	14%	6%	1%	7%
	4%	4%	8%	3%	4%	7%						

[a] CI, confidence interval; numbers represent percentage of participants scoring beyond cut-off values at 80% (1.65) and 95% (1.96) CIs. Impr, improved; Decl, declined; Tot, total.
[b] RCI, based on Chelune et al,[6] upper row; and Jacobson and Truax,[18] lower row.

reflected mixed-sport high school athletes, so additional research should be conducted on specific sports such as football, ice hockey, and soccer.

In summary, the pivotal role that neurocognitive testing has in the objective management of sport-related concussion warrants consideration of both the short- and long-term stability of the baseline assessment. The findings from the current study provide additional evidence for the long-term (ie, 1-year) stability of the new online ImPACT neurocognitive test battery. However, it is still best practice for athletes to receive updated baselines every 2 years to account for any changes in performance. This practice will ensure that an athlete who sustains a concussion will received accurate and individualized management for their injury.

REFERENCES

1. Alves WM, Rimel RW, Nelson WE. University of Virginia prospective study of football-induced minor head injury: status report. *Clin Sports Med*. 1987;6(1):211-218.
2. Aubry M, Cantu R, Dvorak J, et al. Summary and agreement statement of the 1st International Symposium on Concussion in Sport, Vienna 2001. *Clin J Sports Med*. 2002;12(1):6-11.
3. Barr WB. Neuropsychological testing for assessment of treatment effects: methodologic issues. *CNS Spectr*. 2002;7(4):300-302, 304-306.
4. Barth JT, Alves W, Ryan T, et al. Mild head injury in sports: neuropsychological sequelae and recovery of function. In: Levin H, Eisenberg H, Benton A, eds. *Mild Head Injury*. New York, NY: Oxford University Press; 1989:257-275.
5. Broglio SP, Ferrara MS, Macciocchi SN, Baumgartner TA, Elliott RE. Test-retest reliability of computerized concussion assessment programs. *J Athl Train*. 2007;42:509-514.

6. 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.
7. Chicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994;6(4):284-290.
8. Collie A, Darby D, Maruff P. Computerised cognitive assessment of athletes with sports-related head injury. *Br J Sports Med*. 2001;35(5): 297-302.
9. Collie A, Maruff P, Darby D, McStephen MG. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc*. 2003;9(3):419-428.
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. Fleishman J, Benson J. Using LISREL to evaluate measurement models and scale reliability. *Educational and Psychological Measurement*. 1987;47:925-939.
12. Guskiewicz KM, Bruce SL, Cantu R, et al. Recommendations on management of sport-related concussion: summary of the National Athletic Trainers' Association position statement. *Neurosurgery*. 2004;55(4):891-895.
13. Hunt TN, Ferrara MS, Miller LS, Macciocchi S. The effect of effort on baseline neuropsychological test scores in high school football athletes. *Arch Clin Neuropsychol*. 2007;22(5):615-621.
14. Iverson GL, Brooks BL, Collins MW, Lovell MR. Tracking neuropsychological recovery following concussion in sport. *Brain Inj*. 2006;20(3):245-252.
15. Iverson GL, Franzen M, Lovell MR, Collins MW. Construct validity of computerized neuropsychological screening in athletes with concussion. *Arch Clin Neuropsychol*. 2004;19:961-962.
16. Iverson GL, Lovell MR, Collins MW. Interpreting change in ImPACT following sport concussion. *Clin Neuropsychol*. 2003;17(4):460-467.
17. Iverson GL, Sawyer DC, McCracken LM, Kozora E. Assessing depression in systemic lupus erythematosus: determining reliable change. *Lupus*. 2001;10(4):266-271.
18. 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.
19. Lovell MR. ImPACT online version: what are the similarities and differences between the desktop version and the online version? Available at: http://impacttest.com/faq#faq_71. Accessed July 8, 2011.
20. Lovell MR, Collins MW, Fu F, Burke C, Podell K. Neuropsychologic testing in sports: past, present, and future. *Br J Sports Med*. 2001;35:367-372.
21. Maruff P, Collie A, Anderson V, Mollica C, McStephen M, McCrory P. Cognitive development in children: implications for concussion management. *Br J Sports Med*. 2004;38:654-655.
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, Collie A, Anderson V, Davis G. Can we manage sport related concussion in children the same as in adults? *Br J Sports Med*. 2004;38(5):516-519.
24. McCrory P, Johnston K, Meeuwisse W, et al. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. *Br J Sports Med*. 2005;39(4):196-204.
25. McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on concussion in sport: the 3rd International Conference on Concussion in Sport held in Zurich, November 2008. *J Athl Train*. 2009;44: 434-448.
26. Miller JR, Adamson GJ, Pink MM, Sweet JC. Comparison of preseason, midseason, and postseason neurocognitive scores in uninjured collegiate football players. *Am J Sports Med*. 2007;35:1284-1288.
27. Moser RS, Iverson GL, Echemendía RJ, et al. Neuropsychological evaluation in the diagnosis and management of sports-related concussion. *Arch Clin Neuropsychol*. 2007;22:909-916.
28. Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. *The American Statistician*. 1998;42:59-66.
29. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *Am J Sports Med*. 2010;38:47-53.
30. Schatz P, Moser RS, Solomon GS, Ott SD, Karpf R. Incidence of invalid computerized baseline neurocognitive test results in high school and college students. *J Athl Train*. In press.
31. Schatz P, Pardini JE, Lovell MR, Collins MW, Podell K. Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. *Arch Clin Neuropsychol*. 2006;21(1):91-99.
32. Schatz P, Zillmer EA. Computer-based assessment of sports-related concussion. *Appl Neuropsychol*. 2003;10(1):42-47.
33. Valovich TC, Perrin DH, Gansneder BM. Repeat administration elicits a practice effect with the balance error scoring system but not with the standardized assessment of concussion in high school athletes. *J Athl Train*. 2003;38:51-56.
34. Wilk CM, Gold JM, Bartko JJ, Dickerson F, et al. Test-retest stability of the repeatable battery for the assessment of neuropsychological status in schizophrenia. *Am J Psychiatry*. 2002;159(5):838-844.