original research

ImPact Test-Retest Reliability: Reliably Unreliable?

Jacob Resch, PhD, ATC*; Aoife Driscoll, BSc†; Noel McCaffrey, MD†; Cathleen Brown, PhD, ATC‡; Michael S. Ferrara, PhD, ATC, FNATA‡; Stephen Macciocchi, PhD, ABPP§; Ted Baumgartner, PhD‡; Kimberly Walpert, MD‖

*University of Texas at Arlington; †Dublin City University, Ireland; ‡University of Georgia, Athens; §Shepherd Center, Atlanta, GA; ‖Georgia Neurological Surgery, Athens

Context: Computerized neuropsychological testing is commonly used in the assessment and management of sport-related concussion. Even though computerized testing is widespread, psychometric evidence for test-retest reliability is somewhat limited. Additional evidence for test-retest reliability is needed to optimize clinical decision making after concussion.

Objective: To document test-retest reliability for a commercially available computerized neuropsychological test battery (ImPACT) using 2 different clinically relevant time intervals.

Design: Cross-sectional study.

Setting: Two research laboratories.

Patients or Other Participants: Group 1 (n = 46) consisted of 25 men and 21 women (age = 22.4 ± 1.89 years). Group 2 (n = 45) consisted of 17 men and 28 women (age = 20.9 ± 1.72 years).

Intervention(s): Both groups completed ImPACT forms 1, 2, and 3, which were delivered sequentially either at 1-week intervals (group 1) or at baseline, day 45, and day 50 (group 2). Group 2 also completed the Green Word Memory Test (WMT) as a measure of effort.

Main Outcome Measures: Intraclass correlation coefficients (ICCs) were calculated for the composite scores of ImPACT between time points. Repeated-measures analysis of variance was used to evaluate changes in ImPACT and WMT results over time.

Results: The ICC values for group 1 ranged from 0.26 to 0.88 for the 4 ImPACT composite scores. The ICC values for group 2 ranged from 0.37 to 0.76. In group 1, ImPACT classified 37.0% and 46.0% of healthy participants as impaired at time points 2 and 3, respectively. In group 2, ImPACT classified 22.2% and 28.9% of healthy participants as impaired at time points 2 and 3, respectively.

Conclusions: We found variable test-retest reliability for ImPACT metrics. Visual motor speed and reaction time demonstrated greater reliability than verbal and visual memory. Our current data support a multifaceted approach to concussion assessment using clinical examinations, symptom reports, cognitive testing, and balance assessment.

Key Words: intraclass correlation, concussions, mild traumatic brain injuries, neuropsychological testing, athletes

---

Key Points

- ImPACT had strong to weak test-retest reliability over time, consistent with the results of previous studies.
- Reliability was greater for the visual motor speed and reaction time subscores than for the verbal and visual memory subscores.
- Computerized neuropsychological testing is only 1 component of a multifaceted concussion-management program that uses all appropriate tools in clinical decision making.

I n 2001, the Concussion in Sport (CIS) group concluded that neuropsychological testing was one of the "cornerstones" of concussion management.[1] Since that time, the CIS group has emphasized a multifaceted approach that includes neuropsychological testing in the management of sport-related concussion.[2,3]

Computerized neuropsychological tests are readily available and are believed to possess numerous benefits, including standardized and rapid delivery, a centralized means of data storage, and multiple forms to reduce the potential for practice effects while potentially measuring the same neurocognitive constructs as traditional neuropsychological tests.[4,5] Despite the benefits and empirical evidence supporting the use of computerized testing, questions regarding the psychometric properties and the clinical utility of these tests have been raised.[6]

Randolph et al[6] reviewed the psychometric properties of computerized neuropsychological testing platforms and found limited to no evidence of reliability and validity for all existing computerized platforms. Several groups[6–10] investigating the reliability of the ImPACT (ImPACT Applications, Pittsburgh, PA) have found intraclass correlation coefficients (ICCs) ranging from 0.15 to 0.85 for any 1 outcome measure. Specifically, higher ICCs (0.38 to 0.85) were reported for the ImPACT composite visual motor speed and composite reaction time scores and lower ICCs (0.23 to 0.64) for composite visual and verbal memory.[7,10,11] The highest ICC values were for the Web version of ImPACT (0.62 to 0.85), using a 1-year test-retest interval, in participants 13 to 18 years old.[10] The rationale for these discrepancies in ICC values may be varying methods among studies. Schatz[8] and Elbin et al[10] (also P. Schatz, written communication, 2012, and R. J. Elbin,

written communication, 2012) administered the same form (form 1) at 1- and 2-year intervals.[8] Broglio et al[7] administered forms 1, 2, and 3 over a 50-day period.[3] The results from Schatz[8] and Elbin et al[10] are clinically meaningful in reestablishing baseline cognitive values for young athletes and athletes previously diagnosed with a concussion. The methods and results of Broglio et al[7] reflect the clinical use of ImPACT when assessing an athlete with sport concussion. Variable test-retest reliability on computerized cognitive tests enhances the importance of the clinical examination and clinical judgment in the management of sport-related concussion.[6] Although a multifaceted approach to sport-related concussion management is recommended (ie, neuropsychological testing, balance or motor-ability assessment, and monitoring self-reported symptoms), clinicians employed by institutions with limited resources may rely more heavily on computerized neuropsychological testing to determine the state of a concussed athlete and use these data in making return-to-play decisions.[2,4]

Because ImPACT is often used as a stand-alone instrument, we examined the test-retest reliabilities of ImPACT in 2 samples using 2 test-retest time intervals. We hypothesized that ImPACT outcome measures would reflect acceptable ICCs (≥ 0.75) at all time intervals. In addition to our primary hypothesis, we also hypothesized that ImPACT would have low false-positive and false-negative rates and limited practice effects.

METHODS

Our study was conducted at universities in Ireland and the United States. Participants were all nonathlete university students. Participants were excluded if English was not their primary language or they had a self-reported history of learning disability or attention deficit disorder, a psychiatric condition, or a concussion in the 6 months before or during the study. Sample size was calculated to achieve a power of .80 and $d = 0.75$, which is consistent with related literature.[7,12,13]

Participants

For all participants (n = 92), this was their first exposure to ImPACT. Group 1 (n = 46) consisted of students from an Irish university, whereas group 2 (n = 45) consisted of students from a US university. The study was approved by the university ethics or institutional review board. If a participant did not complete all 3 time points or if his or her baseline assessment was determined invalid by the criteria suggested by ImPACT, then the data were removed from subsequent analyses. Removal of incomplete or invalid data allowed for optimal ICC calculations.[14]

ImPACT

A computerized neuropsychological test that tests attention, memory, reaction time, and information-processing speed, ImPACT (version 6.7.723) consists of 8 tasks: immediate and delayed word recall, immediate and delayed design recall, a symbol-match test, a 3-letter recall, the X's and O's test, and a color-match test. Results from 2 or more of the aforementioned tasks are used to calculate 5 subscores: visual and verbal memory, reaction time, visual

motor speed, and impulse control. To determine if examinees provide a good effort, ImPACT uses invalidity criteria. Participants are excluded from analysis if they have an impulse control score of 20 or more, have a score greater than 30 for the number of X's and O's incorrect, correctly respond to less than 69% for word memory, score less than 50% correct for design memory, or recall fewer than 8 letters for the 3-letters test.[14]

Word Memory Test

The Green Word Memory Test (WMT) (Green's Publishing Inc, Edmonton, AB, Canada) measures effort. The WMT presents 20 pairs of words to each participant. The score is based on 4 subtests for immediate recall and delayed recall; the latter occurs after a 30-minute delay. The effort measure is based on scores of immediate recall, delayed recall, consistency of responses, multiple choice, paired associates, and free recall of the word pairs. The test takes approximately 40 minutes to complete, including the 30-minute delay. The ImPACT and WMT were administered via desktop computer and external mouse. Total test time to complete both tests was approximately 40 minutes.[15]

Testing Protocol

The testing protocol for this study is similar to the protocol used by Broglio et al.[7] However, we used only 1 computerized neuropsychological test along with the WMT. Additionally, we supplemented the previous time frame[7] by using a 1-week test-retest interval.

The first session consisted of participants reading and signing the institutional review board–approved informed consent form. Participants then completed a health questionnaire consisting of demographic information, concussion history, and current health status. At this time, we determined if they met the inclusion criteria.

Group 1 (Irish students) completed a baseline test (form 1) and then was reassessed on day 7 with form 2 (time point 2) and on day 14 with form 3 (time point 3). Group 2 (US students) completed a baseline test (form 1) and then was reassessed approximately 45 days later with form 2 (time point 2) and approximately 5 days later (day 50) with form 3 (time point 3).

Statistical Procedures

The ICCs for ImPACT verbal and visual memory, visual motor speed, and reaction time were calculated between time points 1, 2, and 3. Reliability coefficients were calculated for each variable for baseline and time point 2, baseline and time point 3, and time point 2 and time point 3 using a 1-way analysis-of-variance (ANOVA) model. The ICC for a 1-way ANOVA model is defined as $[MS_A - MS_W]/[MS_A + (k - 1)MS_W]$, where MS_A is the mean squares among participants, MS_W is the mean squares within participants, and k is the number of observations per participant.[16,17] Intraclass correlation coefficients range from 0 to 1. Larger coefficients suggest greater reliability.[18]

Effort was assessed using the WMT according to the manufacturer's instructions as well as the ImPACT invalidity criteria (impulse control score greater than 20). A repeated-measures ANOVA was used to detect differ-

Table 1. Participants' Descriptive Data

						Mean ± SD		
Group	Sex	No.	Baseline Age, y	Height, cm	Weight, kg	Cumulative Grade Point Average	Education, y	SAT Score
1	Men	25	22.5 ± 2.15	181.6 ± 7.26	80.7 ± 14.33	NA	17.4 ± 1.12	NA
	Women	21	22.3 ± 1.53	165.9 ± 9.03	59.7 ± 8.31	NA	17.9 ± 0.73	NA
	Total	46	22.4 ± 1.88	174.5 ± 11.28	72.1 ± 16.00	NA	17.6 ± 0.97	NA
2	Men	17	21.1 ± 1.716	179.6 ± 8.74	79.1 ± 14.50	3.4 ± 0.47	13.7 ± 1.05	1223.5 ± 153.67
	Women	28	20.8 ± 1.74	165.8 ± 7.47	59.9 ± 6.81	3.4 ± 0.33	13.5 ± 1.29	1202.9 ± 90.34
	Total	45	20.9 ± 1.72	171.02 ± 10.37	67.1 ± 13.93	3.4 ± 0.39	13.6 ± 1.20	1212.8 ± 122.84

Abbreviation: NA, data not available.

ences in effort across time. The repeated-measures ANOVAs were selected to examine group differences in composite scores across time and to investigate potential practice effects. Greenhouse-Geisser corrections were implemented when sphericity violations occurred. A Bonferroni adjustment was made for multiple pairwise comparisons during post hoc analysis. Effect size was calculated with the Hedges *g*. All data analyses were performed with SPSS (version 17.0; SPSS Inc, Chicago, IL), and statistical significance was set at α ≤ .05.

RESULTS

Descriptive data for both groups are presented in Table 1.

Assessment of Effort

The assessment of group 1's effort was based solely on the ImPACT invalidity data; the WMT data were analyzed only for group 2. The WMT scores for immediate recall, delayed recall, and consistency variables exceeded 85%, indicating that participants provided good effort at each time point.[15] A review of each participant's scores revealed no instance of poor effort at any time point. Group 2 means and standard deviations for each variable at each time point are shown in Table 2.

Group 1 was tested at time point 2, 5.95 ± 1.28 days after time point 1, and at time point 3, 6.84 ± 1.94 days after time point 2. Group 2 was tested at time point 2, 47.27 ± 2.74 days after time point 1, within 44.89 ± 67.87 minutes of their test time during the first session. Time point 3 for group 2 occurred approximately 6.90 ± 1.10 days after time point 2, within 52.22 ± 61.45 minutes of their test time during the first session. One participant was excluded from group 2 because of an invalid effort determined by the ImPACT invalidity criteria. No participants were excluded from this study because they sustained a concussion 6 months before or during the study. An ANOVA revealed that impulse control was the

only variable different between groups ($F_{1,88} = 5.113$, $P = .026$).

Neuropsychological ICC Results. Mean scores and standard deviations for each group and for each ImPACT composite score by time point are presented in Table 3. Calculated ICC values for each ImPACT subscore for baseline to time point 2, baseline to time point 3, and time point 2 to time point 3 are presented in Table 4. The highest ICC values were for composite visual motor speed and reaction time scores. Visual motor speed scores ranged from 0.71 to 0.84 in group 1 and 0.66 to 0.76 for group 2. Composite reaction time scores ranged from 0.78 to 0.88 in group 1 and 0.49 to 0.71 in group 2. The ICC values for composite verbal memory were lower than the other composite scores. Composite verbal memory scores ranged from 0.41 to 0.59 in Group 1 and 0.37 to 0.45 in Group 2. ICC values for visual memory composite ranged from 0.26 to 0.85 in group 1 and 0.52 to 0.55 in Group 2.

Practice Effects. We also assessed practice effects across testing sessions. Repeated-measures ANOVA indicated violations of sphericity for reaction time ($W = 0.835$, $P = .020$) for group 2. Greenhouse-Geisser corrections were performed to account for this violation. In group 1, differences were noted across time for visual motor speed ($F_{1.760,79.198} = 6.100$, $P < .001$, $\eta^2 = 0.12$). A higher score for visual motor speed indicates better performance. Post hoc paired *t* tests revealed increases between time point 1 and time point 2 ($t_{45} = -3.113$, $P = .003$, $g = 0.69$) and time point 2 and time point 3 ($t_{45} = -2.232$, $P = .031$, $g = 0.59$). Decreases in composite reaction time were also observed ($F_{2,90} = 7.106$, $P = .001$, $\eta^2 = 0.14$) with decreases between time point 1 and time point 2 ($t_{45} = 2.114$, $P = .040$, $g = .56$) and time point 1 and time point 3 ($t_{45} = 3.426$, $P = .001$, $g = 0.56$).

In group 2, differences were noted across time for visual motor speed ($F_{2,88} = 4.078$, $P = .020$, $\eta^2 = 0.09$). Post hoc paired *t* tests revealed decreases between time point 1 and

Table 2. Composite Scores on the Green Word Memory Test (Mean ± SD) (n = 45)

	Green Word Memory Test Composite Scores					
Time	Immediate Recall	Delayed Recall	Consistency	Multiple Choice	Pair Associates	Free Recall
Baseline	96.83 ± 14.51	98.67 ± 2.11	97.72 ± 3.01	96.89 ± 4.17	95.67 ± 8.89	70.61 ± 12.76
Day 45	98.84 ± 4.54	99.24 ± 1.47	98.76 ± 1.85[a]	97.00 ± 4.93	98.00 ± 6.43[a]	78.16 ± 16.11[a]
Day 50	99.61 ± 1.06	99.74 ± 0.74[b]	99.36 ± 1.38[a]	98.89 ± 2.80[b]	99.44 ± 3.06[b]	84.90 ± 10.58[b]

[a] $P ≤ .05$.
[b] $P ≤ .01$.

Table 3. ImPACT Composite Scores by Time (Mean ± SD)

Time Point	Memory Composite Verbal	Memory Composite Visual	Visual Motor Speed	Composite Reaction Time	Total Symptom Score
Group 1 (n = 46)					
1	0.90 ± 0.09	0.76 ± 0.14	39.32 ± 8.16	0.56 ± 0.07	6.59 ± 7.67
2	0.92 ± 0.07	0.78 ± 0.11	40.80 ± 6.44	0.54 ± 0.07[a]	5.59 ± 8.27
3	0.92 ± 0.06	0.78 ± 0.12	42.36 ± 6.67[a,b]	0.53 ± 0.07[b]	3.33 ± 5.77
Group 2 (n = 45)					
1	0.90 ± 0.09	0.79 ± 0.125	42.00 ± 6.18	0.56 ± 0.09	4.13 ± 7.02
2	0.92 ± 0.08	0.82 ± 0.10	41.57 ± 5.87	0.55 ± 0.07	3.49 ± 7.42
3	0.92 ± 0.08	0.81 ± 0.12	43.44 ± 6.75[a,b]	0.54 ± 0.07	3.44 ± 6.51

[a] Difference from baseline ($P \leq .05$).
[b] Difference from week 1 ($P \leq .05$).

time point 3 ($t_{44} = -2.122$, $P = .039$, $g = 0.58$) and time point 2 and time point 3 ($t_{44} = -2.521$, $P = .015$, $g = 0.66$).

False-Positive Rate

The ImPACT was designed to identify a significant change in performance compared with baseline measures using reliable change indices.[14] This feature was used to compare time points 2 and 3 with time point 1 for both groups. In group 1, 37% of participants (n = 17) were classified on 1 or more composite scores of ImPACT as impaired at time point 2. At time point 3, 46% of participants (n = 21) were classified as having a change from baseline. Given that our sample consisted of healthy adults, false-positive misclassifications may have reflected either increased or decreased scores compared with baseline performance. Detailed results of this analysis are found in Table 5.

In group 2, 22.2% (n = 10) of participants achieved a different score on 1 or more composite scores at time point 2. At time point 3, 28.9% (n = 13) of participants were considered different from baseline. Detailed results of this analysis are found in Table 6.

DISCUSSION

The purpose of our study was to evaluate the test-retest reliability of ImPACT using 2 different time intervals. We hypothesized that each composite score of ImPACT would demonstrate an ICC value of 0.75 or higher.[19] In accordance with classical test theory, when determining test-retest reliability, time points should be far enough apart to minimize practice effects but not so long as to allow maturational or historical factors to influence test performance. Results from 2 independent groups tested at different time intervals in different countries revealed a range of ICC values similar to those previously reported in the literature.[7,8] For both groups, we found higher ICC values for composite visual motor speed and reaction time and lower ICC values for visual and verbal memory. Despite 3 ICC values exceeding what is considered optimal (>0.75), approximately half of the values fell below acceptable reliability for use in clinical decision making.[6,19,20] The remaining values met or exceeded the suggested values for clinical utility. Another finding was that the majority of ICC values in group 2 were less than those in group 1, indicating that the reliability of ImPACT may decrease over time and possibly reflect a need for more frequent baseline testing.

Our study adds to the evidence suggesting that ImPACT has varying reliability. Overall, 50% of ICC values met our definition of acceptable reliability ($\geq$0.75) for 1-week intervals. Unfortunately, this time frame is not realistic and is not typical of how ImPACT is used clinically. With clinically relevant time points, most (92%) of the test-retest coefficients were suboptimal but slightly better than those in a previous study[7] using similar methods. In addition, their findings were similar to ours when using a longer test-retest time interval and the same test form.[8,9,21]

We controlled for random error by testing individuals at approximately the same time of day; testing in the same environment; and following a standardized testing procedure, using the ImPACT manufacturer's instructions and controlling for effort. Although we accounted for sources of random error, systematic error may have influenced our results through hardware and software applications.[22] All computers were equipped with the Windows XP operating system (Microsoft Corporation, Redmond, WA) and the most current version of Adobe Flash (Adobe Systems Inc, San Jose, CA), with only 2 programs running concurrently: the ImPACT and the Green WMT (group 2). We also

Table 4. Intraclass Correlation Coefficient Values for Each ImPACT Composite Score

	Intraclass Correlation Coefficient			
Time Points	Memory Composite Verbal	Memory Composite Visual	Visual Motor Speed	Composite Reaction Time
Group 1 (n = 46)				
1 to 2	0.56	0.26	0.78	0.84
1 to 3	0.59	0.53	0.71	0.78
2 to 3	0.41	0.85	0.84	0.88
Group 2 (n = 45)				
1 to 2	0.45	0.52	0.76	0.57
1 to 3	0.37	0.52	0.74	0.49
2 to 3	0.40	0.55	0.66	0.71

Table 5. Misclassifications by ImPACT Indicating Statistical Change From Time Point 1 for Group 1 (n = 46), No. (%)

	ImPACT Variable				
Time Point	Memory Composite Verbal	Memory Composite Visual	Visual Motor Speed	Composite Reaction Time	Impaired on Any Variable
2	8 (17)	3 (7)	5 (11)	7 (15)	17 (37)
3	7 (15)	4 (9)	11 (24)	8 (17)	21 (46)

controlled for suboptimal effort in 2 groups using a free-standing effort test, although the results were highly consistent across groups. Data for participants identified as having 1 or more invalid baseline measures on ImPACT or impaired scores on the WMT were removed from the analysis.[23] In our sample, 1% met the ImPACT invalidity criteria. Related literature reported an invalidity rate of 6% to 4%.[7] Explanations for the decreased error rate may include testing in a distraction-free environment and the relatively high academic achievement level of the participants.[21] Future authors should address the various sources of error that influence test-retest reliability, including random and systematic error.

Moderate practice effects (0.56 to 0.66) were noted for the ImPACT visual motor speed in both groups and for reaction time in group 1. For group 2, the score for visual motor speed improved (decreased) over time. An improvement in performance on the ImPACT would only make differences between the baseline and postconcussion evaluations more apparent. A worse performance, as observed in our results, may be more problematic for the clinician in that it may be a contributing source of test-retest error. However, this decreased performance was not statistically significant and may not be clinically meaningful.

A point of contention with respect to computerized neuropsychological testing is the effect of academic performance on neuropsychological test results. Because their sample's mean SAT score was higher than the national average, Broglio et al[7] suggested that their results may not accurately reflect the clinical population to whom this testing is administered. In a related study, Brown et al[21] reported that SAT scores in varsity collegiate athletes were positively related to the results of computerized neuropsychological testing. In the current study, SAT scores for group 2 met or exceeded the 2009 national average. High scores on the SAT positively affect computerized neuropsychological test results, and group 2's SAT scores were well within the range of the scores of most collegiate student-athletes,[21] so our findings are likely representative of those in a sample of student-athletes.

Computerized neuropsychological testing, specifically ImPACT, exhibits a sensitivity of approximately 79% to 93% when classifying individuals as concussed, but ImPACT misclassified 22% to 46% of our healthy college-aged adult sample as impaired on 1 or more indices at 1 or both time points after baseline testing.[11,24,25] The clinician must be aware of these potential misclassifications (either false-positives or false-negatives) when using the ImPACT battery. Specifically, a false-positive result would lead to more conservative management of a concussed athlete. A false-negative result might lead to premature progression through a concussion-management protocol and an inappropriate return-to-play decision. We did not assess false-negative results and cannot make any conclusions about this variable within the context of this study.

CONCLUSIONS

Computerized neuropsychological testing has been adopted as a core component of many concussion-management programs. We found that ImPACT had varying test-retest reliability on several metrics using different time frames for reassessment. Clinicians should recognize that a computerized neuropsychological test such as ImPACT is only 1 component of a concussion-management protocol and use all appropriate tools in clinical decision making and return-to-play decisions.[2,7,8,24]

REFERENCES

1. 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 Sport Med*. 2002;12(1):6–11.
2. 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. *Br J Sports Med*. 2009;43(1 suppl):i76–i90.
3. 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.
4. Guskiewicz KM, Bruce SL, Cantu RC, et al. National Athletic Trainers' Association position statement: management of sport-related concussion. *J Athl Train*. 2004;39(3):280–297.
5. 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.
6. Randolph C, McCrea M, Barr WB. Is neuropsychological testing useful in the management of sport-related concussion? *J Athl Train*. 2005;40(3):139–152.
7. Broglio SP, Ferrara MS, Macciocchi SN, Baumgartner TA, Elliott R. Test-retest reliability of computerized concussion assessment programs. *J Athl Train*. 2007;42(4):509–514.
8. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using imPACT. *Am J Sports Med*. 2009;10:38(1):47–53.

Table 6. Misclassifications by ImPACT Indicating Statistical Change From Time Point 1 for Group 2 (n = 45), No. (%)

	ImPACT Variable				
Time Point	Memory Composite Verbal	Memory Composite Visual	Visual Motor Speed	Composite Reaction Time	Impaired on Any Variable
2	4 (8.9)	1 (2.2)	7 (15.6)	2 (4.4)	10 (22.2)
3	5 (11.1)	4 (8.9)	3 (6.7)	4 (8.9)	13 (28.9)

9. Erlanger D, Feldman D, Kutner K, et al. Development and validation of a web-based neuropsychological test protocol for sports-related return-to-play decision-making. *Arch Clin Neuropsychol*. 2003; 18(3):293–316.

10. 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(11):2319–2324.

11. 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.

12. Maxwell SE, Delaney HD. *Designing Experiments and Analyzing Data*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2003.

13. Baumgartner TA, Chung H. Confidence limits for intraclass reliability coefficients. *Measure Phys Educ Exerc Sci*. 2001;5(3): 179–188.

14. Lovell M. *ImPACT 2007 (6.0) Clinical Interpretation Manual*. Pittsburgh, PA: ImPACT Applications Inc; 2007.

15. Green P. *Green's Word Memory Test for Windows: User's Manual and Program 2005*. Edmonton, AB: Green's Publishing Inc; 2005.

16. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30–46.

17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.

18. Baumgartner TA, Jackson AS, Mahar MT, Rowe DA. *Measurement for Evaluation In Physical Education and Exercise Science*. 8th ed. New York, NY: McGraw Hill; 2007.

19. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Norwalk, CT: Appleton & Lange; 1993.

20. Pedhazu, EJ, Pedhazur Schmelkin L. *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1991.

21. Brown CN, Guskiewicz KM, Bleiberg J. Athlete characteristics and outcome scores for computerized neuropsychological assessment: a preliminary analysis. *J Athl Train*. 2007;42(4):515–523.

22. Cernich AN, Brennana DM, Barker LM, Bleiberg J. Sources of error in computerized neuropsychological assessment. *Arch Clin Neuropsychol*. 2007;22(1 suppl):S39–S48.

23. Lovell M. *ImPACT 2007 (6.0) Software User's Manual*. Pittsburgh, PA; ImPACT Applications Inc; 2007.

24. Broglio SP, Macciocchi SN, Ferrara MS. Sensitivity of the concussion assessment battery. *Neurosurgery*. 2007;60(6):1050–1058.

25. 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(10):1630–1635.

Address correspondence to Jacob Resch, PhD, ATC, University of Texas at Arlington, Box 19259, 113 Maverick Activities Center, Arlington, TX 76102. Address e-mail to resch@uta.edu.