

Two-year Test–Retest Reliability of ImPACT in High School Athletes

William T. Tsushima^{1,*}, Andrea M. Siu², Annina M. Pearce³, Guangxiang Zhang⁴, Ross S. Oshiro⁵

¹*Straub Clinic and Hospital, Honolulu, HI, USA*

²*Hawaii Pacific Health, Honolulu, HI, USA*

³*Scripps College, Claremont, CA, USA*

⁴*University of Hawaii, Honolulu, HI, USA*

⁵*Queen's Medical Center, Honolulu, HI, USA*

*Corresponding author at: Straub Clinic and Hospital, 888 South King Street, Honolulu, HI 96813, USA.

E-mail address: wtsushima@straub.net (W.T. Tsushima).

Accepted 5 October 2015

Abstract

This research evaluated the 2-year test–retest reliability of the Immediate Postconcussion Assessment and Cognitive Testing (ImPACT) neuropsychological battery, and clarified the need for biennial updated baseline testing of high school athletes. This study compared the baseline test scores of 212 non-concussed athletes that were obtained in Grade 9 and again 2 years later when they were in Grade 11. Regression-based methods indicated that 4 of the 5 ImPACT scores were stable over 2 years, as they fell within the 80% and 95% confidence intervals (CIs). The results suggested that updating baseline testing for high school athletes after 2 years is not necessary. Further research into the consistency of computerized neuropsychological tests over 2 years with high school athletes is recommended.

Keywords: ImPACT; Test–retest reliability

Introduction

In recent years, computer-based neuropsychological test batteries have been regularly administered for the evaluation of sports-related concussions. Test batteries such as the Immediate Postconcussion Assessment and Cognitive Testing (ImPACT) (Lovell, 2006) and CogSport (Collie et al., 2006) have gained widespread acceptance and use among various levels of athletics. Currently, ImPACT is one of the most utilized computer-based neuropsychological test instruments in sports.

The usefulness of the ImPACT test battery as a reliable and valid instrument in the evaluation of a sports-related concussion has been asserted by several sources (Iverson, Lovell, & Collins, 2003; Maerlender et al., 2010; Resch, McCrea, & Cullum, 2013; Schatz & Sandel, 2012). However, the psychometric properties of ImPACT have been questioned (Mayers & Redick, 2012; Randolph, McCrea, & Barr, 2005), including its reliability, which is a fundamental concern when weighing the merits of a psychological test. While there are various forms of reliability indices for psychological tests, such as split-half reliability, alternative form equivalence, and test–retest reliability, the most relevant reliability test for ImPACT is the test–retest reliability or stability test, because of the recommended practice of comparing neurocognitive scores of concussed athletes to their pre-injury or baseline test scores in making return-to-play decisions for the athlete (Louey et al., 2014).

When a neuropsychological test is employed to assist in important return-to-play decisions, sufficient test–retest reliability is critical. According to the Board on Children, Youth, and Families' Committee on Sports-Related Concussions in Youth, there is no agreement on the minimum test–retest values that are considered acceptable for a computerized neuropsychological test, like ImPACT (Graham, Rivara, & Ford, 2014). Some authors accept a minimum correlation coefficient (r) of .60 (Anastasi & Urbana, 1997), while others uphold higher standards of 0.70–0.79 (adequate), 0.80–0.89 (high), or 0.90–0.99 (very high) as an index of a consistent neuropsychological test (Slick, 2006).

A previous test–retest study of ImpACT Version 2.0 examined the composite scores of 56 healthy high school and college students, with an average retest interval of 5.8 days, and obtained the following test–retest correlation coefficients (r) for the composite test scores: Verbal Memory .70, Visual Memory .67, Reaction Time .79, Processing Speed .86, and Total Symptom .65 (Iverson et al., 2003). These reliability measures, while found to be satisfactory, are not representative of real-world use, in that the period between baseline and post-concussion testing is often a week or more rather than a few days (Broglia, Ferrara, Macciocchi, Baumgartner, & Elliot, 2007; Mayers & Redick, 2012).

Additional tests of reliability for ImpACT over longer periods have been conducted. One-month test–retest reliability was assessed with 25 college students, with the following intraclass correlation coefficients ($ICCs$): Verbal Memory 0.79, Visual Memory 0.60, Reaction Time 0.77, and Visual Motor Speed 0.88 (Schatz & Ferris, 2013). The authors also noted significant improvement in Visual Motor Speed, perhaps due to practice effects. Other researchers examined the ImpACT scores of 73 healthy college controls tested 45 days apart and found $ICCs$ of Verbal Memory 0.23, Visual Memory 0.32, Reaction Time 0.39, Visual Motor Speed 0.38, and Impulse Control 0.15 (Broglia et al., 2007). The investigators concluded that the test–retest $ICCs$ fell below the 0.60 level for making clinical decisions. A similar study re-examined the test–retest reliability of the ImpACT between baseline, 45 days after baseline, and 50 days after baseline in a sample of 85 physically active college students (Nakayama, Covassin, Schatz, Nogle, & Kovan, 2014). The respective $ICCs$ for baseline to Day 45 and baseline to Day 50 were Verbal Memory (0.76 and 0.65), Visual Memory (0.72 and 0.60), Visual Motor Speed (0.87 and 0.85), and Reaction Time (0.67 and 0.71), all exceeding the threshold value of 0.60 for acceptable test–retest reliability.

Hunt and Ferrara (2009) observed significant differences on neuropsychological tests (not including ImpACT) between Grade 9 and Grades 11 and 12. Citing the Luria theory of brain development, the authors noted that cognitive maturity continues to develop throughout early adulthood, resulting in differences in neuropsychological performances as seen in this study. Consequently, Hunt and Ferrara recommended that baseline testing of high school athletes should occur twice in high school, first upon entering Grade 9 and retest at entrance into Grade 10, acknowledging the cognitive growth during that period. Others investigated the 1-year reliability of the ImpACT in a sample of 369 high school athletes (Elbin, Schatz, & Covassin, 2011). They reported $ICCs$ of Verbal Memory 0.62, Visual Memory 0.70, Reaction Time 0.76, and Motor Processing Speed 0.85, with findings that suggest that annual ImpACT baseline evaluations are not necessary.

A study of 95 collegiate athletes tested ~2 years apart revealed $ICCs$ of Verbal Memory 0.46, Visual Memory 0.65, Reaction Time 0.68, and Processing Speed 0.74, and observed that ImpACT Version 3.0 performance at baseline remained stable over a 2-year period (Schatz, 2010). Findings of low-to-moderate ImpACT reliability have been noted in studies of 40 high school and college student-athletes (Register-Mihalik et al., 2012), 91 college non-athletes (Resch et al., 2013), 305 professional hockey players (Bruce, Echemendia, Meeuwse, Comper, & Sisco, 2014), and 215 active duty military members (Cole et al., 2013). The inconsistent findings of previous test–retest reliability studies of ImpACT may be due to the varying methods among the studies, such as different intervals between testings (e.g., 1 week, 1 month, 1 year), or administering two other computerized neuropsychological tests in one session as in the Broglia et al. (2007) study.

Differences in interpretation of reliability indices have become a topic of debate in the field and further support the need for additional research. Mayers and Redick (2012) considered the reliability indices in the studies by Broglia et al. (2007) and Schatz (2010) to be “unacceptably low.” In their response to Mayers and Redick, Schatz, Kontos and Elbin (2012) opined that Mayers and Redick presented a limited view of the data that did not consider reliable change and regression-based measures. Reliable change indices ($RCIs$) provide a confidence interval (CI) for predicted change and assess whether a change between repeated measurements is clinically meaningful, while regression-based methods (RBM s) utilize regression analysis to predict retest performance levels, and may provide the most accurate method to assess neuropsychological change (Barr, 2002). A recent review of the reliability of several different computerized neuropsychological tests found that $ICCs$ for different tests were quite variable, with some studies reporting adequate reliability, while others indicating less than adequate reliability (Graham et al., 2014). The authors concluded that all commercial test batteries reviewed, including ImpACT, were supported by some studies indicating acceptable reliability.

Elbin et al. (2011) found that the baseline data of high school athletes were stable across 1- and 2-year time periods, but they also noted the scores were not without change. Recognizing the ongoing cognitive maturation in adolescent athletes, the authors recommended that high school athletes receive updated baseline testing every 2 years, if not annually. In light of the developmental changes that occur in adolescence and the recommendation of Elbin et al., the present large-scale investigation was designed to clarify the necessity of biennial updated baseline testing by comparing the baseline ImpACT scores of Grade 9 students with their baseline test scores 2 years later in Grade 11. This is the first known study to compare the test–retest ImpACT baseline results of high school students following a 2-year interval.

Methods

Procedure

The participants completed a baseline ImPACT examination prior to their Grade 9 and Grade 11 sport seasons. The ImPACT tests were administered in small groups up to 20 athletes by certified athletic trainers trained in the standardized administration of the examination.

Instrument

The ImPACT test battery provides neurocognitive test scores along with relevant biopsychosocial information. A partial list of self-reported biopsychosocial data collected includes age, sex, years of education, primary spoken language, sport played, prior concussion, and history of learning disability and attention deficit disorder. The ImPACT test yields four composite scores, including Verbal Memory, Visual Memory, Visual Motor Speed, and Reaction Time. The testing included the ImPACT Post-Concussion Symptom Scale that yielded Total Symptom scores, based on 22 commonly reported symptoms (e.g., headache, dizziness).

Participants

The pool of athletes included 7987 Grade 9 athletes from 36 public high schools across the state of Hawaii between 2008 and 2014. Athletes were excluded from the study if they had (1) a history of concussion ($n = 1210$, 15.1%), (2) a primary language that was not English ($n = 300$, 3.8%), (3) a history of learning disability or ADD/ADHD ($n = 561$, 7.0%), (4) an invalid baseline testing ($n = 1010$, 12.6%), or (5) no baseline testing at Grade 11 ($n = 4694$, 58.8%). An invalid baseline test was identified by the internal validation measure in the ImPACT examination, e.g., an Impulse Control score ≥ 30 (Lovell, 2006). After applying the exclusionary criteria, the athletes selected for this research were 212 9th graders (126 males, 86 females) who were later re-tested for their baseline scores as 11th graders. The mean age of the Grade 9 athletes was 15.3 years ($SD = 0.5$), while the mean age of the Grade 11 athletes was 17.3 years ($SD = 0.5$).

The student athletes participated in a variety of sports: football ($n = 75$), soccer ($n = 43$), basketball ($n = 25$), volleyball ($n = 14$), cheerleading ($n = 11$), softball ($n = 10$), baseball ($n = 8$), and others ($n = 26$).

Approval for the use of the research data was granted by the State of Hawaii Department of Education. The study was assessed by the Hawaii Pacific Health Research Institute and determined to be exempt from Institutional Review Board review.

Statistical Analyses

As in recent test–retest reliability studies (Elbin et al., 2011; Nakayama et al., 2014; Schatz & Ferris, 2013), three types of statistical analyses were performed to test reliability: *ICC*, *RCI*, and *RBM*. *ICC*, a single measure two-way random effects analysis of variance which takes into account individual variation in scores was calculated to assess the consistency of testing at Grade 9 and 2 years later at Grade 11. *ICC* analyses also yielded an unbiased estimate of reliability (*UER*), which reflected the consistency at both times of the baseline assessments. *RCI* was calculated to assess whether a change between the repeated baseline assessments is meaningful, i.e., it is beyond what would be expected by chance. A 95% CI was used in determining reliable change, i.e., by chance 5% of all scores are expected to show changes greater than the calculated *RCI*. A modified *RCI* formula (Chelune, Naugle, Lüders, Sedlak, & Awad, 1993) was employed, which includes an adjustment for practice effects. *RBM* was applied, yielding a regression equation to predict an athlete's level of performance on ImPACT at retest from the initial testing. Both *RCI* and *RBM* are designed so that no more than 5% of the athletes should be outside the CI, indicating a significant improvement or deterioration (Temkin, Heaton, Grant, & Dikmen, 1999).

The outcome measures for the ImPACT baseline assessments included four composite scores—Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time—and a Total Symptom score. Paired *t*-tests were conducted to evaluate differences in scores between the two baseline exams. With the Bonferroni correction, the α -level for all analyses was set at $p < .01$. Statistical analyses were conducted in SAS version 9.3 and SPSS version 22.

Results

The mean ImPACT composite and Total Symptoms scores and *ICCs* are presented in Table 1. *ICC* reliability indices ranged from 0.21 to 0.72 for the composite scores, and 0.40 for the Total Symptoms score. Paired *t*-tests found significant differences ($p < .0001$) in Visual Memory, Visual Motor Speed, and Reaction Time between Grades 9 and 11, with better performances in

Table 1. Test–retest reliability^a

| ImPACT score | Grade 9 | Grade 11 | ICC | ICC 95% CI | | UER | <i>t</i> ^b | <i>p</i> -value |
|--------------------|------------------------|------------------------|-------|------------|-------|-------|-----------------------|-----------------|
| | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) | | Lower | Upper | | | |
| Verbal Memory | 84.7 (8.5) | 83.9 (10.1) | 0.210 | 0.077 | 0.335 | 0.353 | −1.0 | .33 |
| Visual Memory | 72.8 (12.2) | 76.4 (12.3) | 0.485 | 0.375 | 0.581 | 0.629 | 4.2 | <.0001 |
| Visual Motor Speed | 36.4 (6.4) | 40.2 (6.7) | 0.719 | 0.648 | 0.779 | 0.746 | 11.1 | <.0001 |
| Reaction Time | 0.60 (0.07) | 0.58 (0.08) | 0.458 | 0.345 | 0.558 | 0.598 | −4.6 | <.0001 |
| Total Symptom | 5.7 (10.0) | 4.7 (7.2) | 0.403 | 0.284 | 0.510 | 0.575 | −1.6 | .12 |

Notes: ^aICC=intraclass correlation coefficient; CI=confidence interval; UER=unbiased estimate of reliability; *p*, significance of *t*.

^b*df* = 210; Bonferroni-corrected $\alpha p < .01$.

Table 2. Reliable change indices (*RCI*)

| ImPACT score | Grade 9 <i>M</i> (<i>SD</i>) | Grade 11 <i>M</i> (<i>SD</i>) | <i>SEM</i> ^{9a} | <i>SEM</i> ^{11a} | <i>S</i> _{diff} ^b | 80% CI ^c | 95% CI ^c |
|--------------------|-----------------------------------|------------------------------------|--------------------------|---------------------------|---------------------------------------|---------------------|---------------------|
| Verbal Memory | 84.7 (8.5) | 83.9 (10.1) | 7.53 | 8.99 | 9.92 | 69.34 | 91.04 |
| Visual Memory | 72.8 (12.2) | 76.4 (12.3) | 8.75 | 8.80 | 10.75 | 71.23 | 92.45 |
| Visual Motor Speed | 36.4 (6.4) | 40.2 (6.7) | 3.41 | 3.53 | 4.65 | 66.04 | 88.21 |
| Reaction Time | 0.60 (0.07) | 0.58 (0.08) | 0.05 | 0.06 | 0.07 | 79.25 | 92.45 |
| Total Symptom | 5.7 (10.0) | 4.7 (7.2) | 7.59 | 5.47 | 6.54 | 77.36 | 84.91 |

Notes: ^aStandard error of measure at Grade 9 and Grade 11 = [*SD* × √(1 − *r*_{xy})].

^bStandard error of difference scores based on Chelune et al. (1993) √[(*SEM*₉²) + (*SEM*₁₁²)].

^cCI=confidence interval; numbers report the percentage of subjects with predicted change scores within cut-off (80% CI = 1.28, 95% = 1.96).

Table 3. Regression-based methods (*RBM*)^a

| ImPACT score | Grade 9 <i>M</i> (<i>SD</i>) | Grade 11 <i>M</i> (<i>SD</i>) | Intercept | Slope | <i>S</i> _{xy} | 80% CI | 95% CI |
|--------------------|-----------------------------------|------------------------------------|-----------|-------|------------------------|--------|--------|
| Verbal Memory | 84.7 (8.5) | 83.9 (10.1) | 62.406 | 0.257 | 9.918 | 79.25 | 95.75 |
| Visual Memory | 72.8 (12.2) | 76.4 (12.3) | 40.892 | 0.488 | 10.748 | 81.13 | 94.81 |
| Visual Motor Speed | 36.4 (6.4) | 40.2 (6.7) | 12.957 | 0.747 | 4.648 | 83.02 | 97.64 |
| Reaction Time | .60 (.07) | .58 (.08) | 0.272 | 0.505 | 0.073 | 87.74 | 96.23 |
| Total Symptom | 5.7 (10.0) | 4.7 (7.2) | 2.937 | 0.306 | 6.542 | 85.85 | 94.34 |

Notes: ^a*S*_{xy}=standard error of estimate; CI=confidence interval; numbers represent the percent of participants with change scores within cut-off (80% CI = 1.28, 95% CI = 1.96).

Grade 11. Visual Motor Speed yielded a medium effect size, while effect sizes were small for Visual Memory and Reaction Time. Unbiased estimates of reliability (*UER*) were consistent with the *ICCs*.

RCIs are presented in Table 2 at 80% and 95% CIs, showing significant changes at Grade 11. All Composite and Total Symptoms scores from the second baseline testing were outside the 80% and 95% CIs. In contrast, *RBM*s, seen in Table 3, revealed stability in the follow-up baseline scores. At follow-up baseline assessments, nearly all composite scores fell within the 80% and 95% CIs. Specifically, at Grade 11 baseline testing, 79%–88% of composite scores and 86% of Total Symptom scores fell within the 80% CI. Only 0.8% of Verbal Memory scores fell outside the 80% CI, and only 0.2% of the Visual Memory and 0.7% of Total Symptoms scores fell outside the 95% CI.

Table 4 shows the rates of changes using *RCIs* versus *RBM*s. *RCIs* showed that 21%–34% of Composite and Total Symptoms scores fell outside the 80% CI, while 8%–15% of the scores were outside the 95% CI. On the other hand, with *RBM*s, 12%–21% of the scores were beyond the 80% CI, and only 2%–6% were outside the 95% CI.

Discussion

The present study evaluated the test–retest reliability of the baseline ImPACT Composite and Total Symptoms scores of 212 non-concussed high school athletes tested in Grade 9 and again 2 years later in Grade 11. The results of the *RBM* analyses revealed that the retest scores were stable as nearly all Composite scores fell within the 80% and 95% CIs. In contrast, *ICC* correlation

Table 4. Rates of changes (%) using reliable change indices (*RCI*) vs regression-based model (*RBM*)^a

| ImPACT | <i>RCI</i> ^b | | | | | | <i>RBM</i> | | | | | |
|--------------------|-------------------------|------|-----|---------------------|------|-----|---------------------|------|-----|---------------------|------|-----|
| | 80% CI ^c | | | 95% CI ^c | | | 80% CI ^c | | | 95% CI ^c | | |
| | Impr | Decl | Tot | Impr | Decl | Tot | Impr | Decl | Tot | Impr | Decl | Tot |
| Verbal Memory | 14 | 17 | 31 | 4 | 5 | 9 | 10 | 10 | 21 | 4 | 0 | 4 |
| Visual Memory | 22 | 7 | 29 | 6 | 1 | 8 | 10 | 9 | 19 | 3 | 2 | 5 |
| Visual Motor Speed | 31 | 3 | 34 | 10 | 1 | 12 | 9 | 9 | 17 | 1 | 1 | 2 |
| Reaction Time | 6 | 15 | 21 | 3 | 4 | 8 | 5 | 8 | 12 | 0 | 4 | 4 |
| Total Symptom | 10 | 12 | 23 | 5 | 10 | 15 | 5 | 9 | 14 | 0 | 5 | 6 |

Notes: ^aImpr=improved; Decl=declined; Tot=total.

^b*RCI*: Chelune et al. (1993).

^cCI=confidence interval; numbers represent percent of participants scoring beyond cut-off values (80% CI = 1.28 and 95% CI = 1.96).

analyses suggested a lack of 2-year reliability among the ImPACT scores (except for Visual Motor Speed) due to correlation coefficients that were below the recommended 0.60 level used for clinical purposes (Anastasi & Urbana, 1997). In addition, the *RCI* data indicated significant changes in the Grade 11 ImPACT scores, with 21%–34% falling outside the 80% CI. When the *RCI* and *RBM* approaches have been compared in past studies, *RCI* had wider prediction intervals and poor prediction accuracy, while *RBM* offered more accurate prediction with small prediction-CIs (Temkin et al., 1999). Thus, weight must be given to the *RBM* method which provides a more advanced statistical approach in accounting for confounding factors not completely addressed by the *RCI* methodology (Barr, 2002). Furthermore, the *t*-test differences in scores between Grades 9 and 11 yielded small effect sizes for Visual Memory and Reaction Time and a medium effect size for Visual Motor Speed.

At present, no firm guidelines are available as to how often baseline neuropsychological testing with high school athletes should be administered. The position statement by the American Medical Society for Sports Medicine indicated that there is no ideal interval for repeating baseline neuropsychological tests (Harmon et al., 2013). Further, the 2012 Zurich consensus statement of the 4th International Conference on Concussion in Sport asserted that “there is insufficient evidence to recommend the widespread routine use of baseline neuropsychological testing” (McCrorry et al., 2013). Based on the present *RBM* results that revealed stable 2-year follow-up baseline scores, we assert that baseline testing conducted at Grade 9 yields sufficiently stable ImPACT scores so that repeated baseline testing at Grade 11 appears unnecessary. At the same time, the low correlation coefficients and significant changes in *RCIs* seem to reflect the developmental changes in youths that take place over the 2-year period between Grades 9 and 11, and offer some support to the cautious practice of updating a baseline testing every 2 years to account for any changes in young athletes (Elbin, Schatz, & Covassin, 2011; Hunt & Ferrara, 2009; Schatz and Ferris, 2013). Finally, baseline data used for post-concussion comparison are a major advantage of neurocognitive tests such as ImPACT. But when baseline test scores are not available because of school budgets or limited personnel, the alternative of employing normative comparisons for post-injury assessment is considered reasonable (Echemendia et al., 2012; Schmidt, Register-Mihalik, Mihalik, Kerr, & Guskiewicz, 2012). In such cases, relevant age and sex norms can be used for comparison with post-injury scores.

Limitations

Several limitations of this research are worthy of note. (1) The sample size ($n = 212$) of the high school athletes in this report is one of the largest among test–retest studies of computerized neuropsychological tests. Nonetheless, the present group included only high school athletes, and the results cannot be generalized to younger adolescents or children. Until test–retest reliability research is conducted with pre-high school athletes, it may be advisable to obtain annual baseline assessments with younger adolescents and children who engage in youth contact sports (Hunt & Ferrara, 2009). (2) Another shortcoming in this retrospective study was the reliance on non-verifiable self-reports of personal data by the student-athletes, such as primary language and history of concussion, learning disability, and ADD/ADHD. (3) Although those with invalid ImPACT profiles were excluded from this research, there was no formal effort testing to assure optimal level of neurocognitive performance by the athlete. (4) The high school athletes in Hawaii represent a diverse population of Caucasians, Polynesians, Asians, and other racial/minorities. In view of the composition of the participants in this study, the present results may not be readily generalizable to high school athletes on the U.S. continent. (5) The present investigation employed only the ImPACT test battery to assess the stability of neurocognitive test scores of high school athletes. Examining the test–retest results of other computerized neuropsychological test methods that are also used with high school athletes (e.g., Automated Neuropsychological Assessment Metrics [ANAM], CogSport/Axon) would be valued.

Conclusion

The present mixed findings showed no reliable change of ImPACT scores after a 2-year interval using *RBM* analyses. These results imply that baseline neuropsychological testing updated every 2 years with high school athletes is not necessary. Additional analyses (*ICC* and *RCI*) of the data found low test–retest reliability correlation coefficients and significant *RCIs*, suggesting that a cautious approach of obtaining biennial baseline testing of high school athletes obtained every 2 years is not unreasonable. This is the first known study of test–retest reliability over a 2-year interval with high school athletes; thus, further investigation into the consistency of ImPACT and other computerized neuropsychological tests over 2 years with high school participants is needed.

Funding

The authors wish to acknowledge the support of the Hawaii Concussion Awareness and Management Program. The third author, A.M.P., participated in the Hawaii Pacific Health Summer Student Research Program while she assisted in this research project.

Conflict of Interest

The authors have no financial interest in ImPACT Applications, Inc. and declare no conflict of interest in this study.

References

- Anastasi, A., & Urbana, S. (1997). *Psychological testing* (7th ed.). New York: Prentice-Hall.
- Barr, W. B. (2002). Neuropsychological testing for assessment of treatment effects: Methodological issues. *CNS Spectrums*, 7 (4), 300–306.
- Broglio, S. P., Ferrara, M. S., Macciocchi, S. N., Baumgartner, T. A., & Elliot, R. (2007). Test–retest reliability of computerized concussion assessment programs. *Journal of Athletic Training*, 42 (4), 509–514.
- Bruce, J., Echemendia, R., Meeuwisse, W., Comper, P., & Sisco, A. (2014). 1 year test–retest reliability of ImPACT in professional ice hockey players. *The Clinical Neuropsychologist*, 28 (1), 14–25.
- Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, 7, 41–52.
- Cole, W. R., Arrieux, J. P., Schwab, K., Ivins, B. J., Qashu, F. M., & Lewis, S. C. (2013). Test–retest reliability of four computerized neurocognitive assessment tools in an active duty military population. *Archives of Clinical Neuropsychology*, 28 (7), 732–742.
- Collie, A., Maruff, P., Darby, D., Makdissi, M., McCrory, P., & McStephen, M. (2006). CogSport. In R. J. Echemendia (Ed.), *Sports neuropsychology: Assessment and management of traumatic brain injury* (pp. 176–190). New York: Guilford Press.
- Echemendia, R. J., Bruce, J. M., Bailey, C. M., Sanders, J. F., Arnett, P., & Vargas, G. (2012). The utility of post-concussion neuropsychological data in identifying cognitive change following sports-related MTBI in the absence of baseline data. *The Clinical Neuropsychologist*, 26 (7), 1077–1091.
- Elbin, R. J., Schatz, P., & Covassin, T. (2011). One-year test–retest reliability of the online version of ImPACT in high school athletes. *The American Journal of Sports Medicine*, 39 (11), 2319–2324.
- Graham, R., Rivara, F. P., Ford, M. A., & Spicer, C. M. (2014). *Sports-related concussions in youth: Improving the science, changing the culture*. Washington, DC: The National Academies Press.
- Harmon, K. G., Drezner, J. A., Gammons, M., Guskiewicz, K. M., Halstead, M., et al. (2013). American Medical Society for Sports Medicine position statement: Concussion in sport. *British Journal of Sports Medicine*, 47, 15–26.
- Hunt, T. N., & Ferrara, M. S. (2009). Age-related differences in neuropsychological testing among high school athletes. *Journal of Athletic Training*, 44 (4), 405–499.
- Iverson, G. L., Lovell, M. R., & Collins, M. W. (2003). Interpreting change on ImPACT following sport concussion. *The Clinical Neuropsychologist*, 17 (4), 460–467.
- Louey, A. G., Cromer, J. A., Schember, A. J., Darby, D. G., Maruff, P., et al. (2014). Detecting cognitive impairment after concussion: Sensitivity of change from baseline and normative data methods using the CogSport/Axon Cognitive Test Battery. *Archives of Clinical Neuropsychology*, 29, 432–441.
- Lovell, M. R. (2006). The ImPACT Neuropsychological Test Battery. In R. J. Echemendia (Ed.), *Sports neuropsychology: Assessment and management of traumatic brain injury* (pp. 193–215). New York: Guilford Press.
- Maerlender, A., Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., et al. (2010). Examination of the construct validity of ImPACT™ computerized test, traditional, and experimental neuropsychological measures. *The Clinical Neuropsychologist*, 24 (8), 1309–1325.
- Mayers, L. B., & Redick, T. S. (2012). Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: Psychometric issues. *Journal of Clinical and Experimental Neuropsychology*, 34 (3), 235–242.
- McCrory, P., Meeuwisse, W. H., Aubry, M., Molloy, M., Cantu, B., et al. (2013). Consensus statement on concussion in sport: The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *British Journal of Sports Medicine*, 47, 250–258.
- Nakayama, Y., Covassin, T., Schatz, P., Nogle, S., & Kovan, J. (2014). Examination of the test–retest reliability of a computerized neurocognitive test battery. *The American Journal of Sports Medicine*, 42, 2000–2005.

- Randolph, C. S., McCrea, M., & Barr, W. B. (2005). Is neuropsychological testing useful in the management of sport-related concussion? *Journal of Athletic Training, 40* (3), 139–154.
- Register-Mihalik, J. K., Kontos, D. L., Guskiewicz, K. M., Mihalik, J. P., Conder, R., & Shields, E. W. (2012). Age-related differences and reliability on computerized and paper-and-pencil neurocognitive assessment batteries. *Journal of Athletic Training, 47* (3), 297–305.
- Resch, J., Driscoll, A., McCaffrey, N., Brown, C., Ferrara, M. S., et al. (2013). ImPACT test–retest reliability: Reliably unreliable? *Journal of Athletic Training, 48* (4), 506–511.
- Resch, J. E., McCrea, M. A., & Cullum, C. M. (2013). Computerized neurocognitive testing in the management of sport-related concussion: An update. *Neuropsychological Review, 2*, 335–349.
- Schatz, P. (2010). Long-term test–retest reliability of baseline cognitive assessments using ImPACT. *The American Journal of Sports Medicine, 38* (1), 47–53.
- Schatz, P., & Ferris, C. S. (2013). One-month test–retest reliability of the ImPACT test battery. *Archives of Clinical Neuropsychology, 28* (5), 499–504.
- Schatz, P., Kontos, A., & Elbin, R. J. (2012). Response to Mayers and Redick: “Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: Psychometric issues.” *Journal of Experimental and Clinical Neuropsychology, 34* (4), 428–434.
- Schatz, P., & Sandel, N. (2012). Sensitivity and specificity of the online version of ImPACT in high school and collegiate athletes. *The American Journal of Sports Medicine, 41* (2), 321–326.
- Schmidt, J. D., Register-Mihalik, J. K., Mihalik, J. P., Kerr, Z. Y., & Guskiewicz, K. M. (2012). Identifying impairments after concussion: Normative data versus individualized baselines. *Medicine & Science in Sports & Exercise, 44* (9), 1621–1628.
- Slick, D. J. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, E. M. S. Sherman, & O. Spreen (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed., pp. 3–43). New York: Oxford University Press.
- Temkin, N. R., Heaton, R. K., Grant, I., & Dikmen, S. S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society, 5* (4), 357–369.