APPLIED NEUROPSYCHOLOGY: CHILD
2017, VOL. 6, NO. 2, 166–171
http://dx.doi.org/10.1080/21622965.2016.1138310

# One year test–retest reliability of neurocognitive baseline scores in 10- to 12-year olds

Rosemarie Scolaro Moser[a,b], Philip Schatz[c], Emily Grosner[d], and Kelly Kollias[e]

[a]Sports Concussion Center of NJ, Princeton, New Jersey, USA; [b]RSM Psychology Center, LLC, Princeton, New Jersey, USA; [c]Department of Psychology, St. Joseph's University, Philadelphia, Pennsylvania, USA; [d]Department of Psychology, Pennsylvania State University, University Park, Pennsylvania, USA; [e]Bancroft NeuroRehab, Mt. Laurel, New Jersey, USA

**ABSTRACT**
How often youth athletes 10–12 years of age should undergo neurocognitive baseline testing remains an unanswered question. We sought to examine the test–retest reliability of annual ImPACT data in a sample of middle school athletes. Participants were 30 youth athletes, ages 10–12 years (Mean = 11.6, *SD* = 0.6) selected from a larger database of 10–18 year old athletes, who completed two consecutive annual baseline evaluations using the online version of ImPACT. Athlete assent and parental consent were obtained for all participants. Assessments were conducted either individually or in small groups of 2 to 3 athletes, under the supervision of a neuropsychologist or post-doctoral fellow. Test–retest coefficients were as follows: Verbal Memory .71, Visual Memory .35, Visual Motor Speed .69, Reaction Time .34. Intra-class Correlation Coefficients (single/average) were as follows: Verbal Memory .70/.83, Visual Memory .35/.52, Visual Motor Speed .69/.82, Reaction Time .34/.50. Regression–based measures to correct for practice effects revealed that only a small percentage of cases fell outside 90 and 95% confidence intervals, reflecting stability across assessments. Findings indicate that test–retest reliability of Verbal Memory and Visual Motor Speed are generally stable in 10–12 year old athletes. Nevertheless, Visual Memory Index, Reaction Time Index, and Symptom Checklist scores appear to be less reliable over time, especially compared to published data on high school athletes, suggesting the utility of re-testing on an annual basis in this younger age group.

**KEYWORDS**
Baseline testing; concussion testing; ImPACT; neurocognitive testing

Sport-related concussion in youth continues to be viewed as a widespread public health concern. Currently, all 50 states and the District of Columbia in the United States have implemented some type of legislation mandating concussion management programs for youth. Often referred to as "Return to Play Laws," youth concussion legislation typically includes at least three components: (a) proactive education of sports/school personnel, athletes, and parents/guardians; (b) immediate removal from play of the concussed athlete; and (c) the need for medical clearance before the athlete can return to play. As in professional sports, youth sports in schools and community leagues are adding baseline neurocognitive testing to their concussion management programs. Such baseline testing is often seen as an important way of protecting the youth athlete, although there has been debate as to how effectively these testing programs are administered, overseen, and used (Moser, Schatz, & Lichtenstein, 2015). Nonetheless, baseline test results serve as a measure of the athlete's "normal" level of functioning in the areas that are thought to be most prominently affected by concussion: working memory, attention, processing speed, and reaction time. Post-concussion test results can be compared to baseline test results to assist in decisions regarding the athlete's recovery and readiness to return to play.

For the most part, current legislation does not mandate that youth athletes undergo annual baseline testing. In fact, the Consensus Statement on Concussion in Sport from the fourth International Conference on Concussion in Sport held in Zurich, November, 2012, declared that there was inadequate support for universal, routine baseline neuropsychological testing (Echemendia, 2013; McCrory et al., 2013). Nevertheless, the Centers for Disease Control and Prevention (CDC) recommend baseline testing on a yearly basis, as valid baseline testing allows for a unique and individual comparison of subtle changes in neurocognitive functioning (CDC, 2015).

As a result, the necessity of updating baseline testing on an annual basis is still not scientifically grounded. There is some support for implementing annual

**CONTACT** Rosemarie Scolaro Moser Rmoser@rsmpsychology.com Sports Concussion Center of New Jersey, 281 Witherspoon Street, Ste. 230, Princeton, NJ 08540.

baseline testing during high school, as research has demonstrated that individuals' scores improved between the ninth grade and eleventh/twelfth grades on "traditional" paper-based measures (i.e., Trails A & B, Symbol Digit Modalities Test, Tapping) (Hunt & Ferrara, 2009), as well as improvement during the high school years on computer-based measures (i.e., ImPACT Motor Processing Speed; (Schatz, 2009)), arguably reflecting cognitive growth and maturity during this time. However, using correlational analyses, high school athletes demonstrated low-to-marginal reliability/stability on a computer-based assessment measure (Axon/CogSport; (MacDonald & Duerson, 2015)). Other researchers have documented similar improvements in performance (using repeated-measures analyses) from high school to college-age on both "traditional" paper-based measures (i.e., Trails A & B, Symbol Digit Modalities Test, Hopkins Verbal Learning Test, Brief Visual-Spatial Memory Test), as well as on ImPACT Motor Processing Speed (Register-Mihalik et al., 2012). In contrast, baseline test–retest research has reported 1- and 2-year stability in scores in high school and college-aged athletes on computer-based measures (ImPACT) (Elbin, Schatz, & Covassin, 2011), as well as computer-based and analog measures of reaction time (MacDonald et al., 2015). While the aforementioned research applies solely to high school and college athletes, when testing is implemented in the younger middle school age group, both age and test setting are seen as factors affecting the validity of test score (Lichtenstein, Moser, & Schatz, 2013). Thus, this 10–12 year old group requires increased vigilance and care in test administration and interpretation (Moser et al., 2015).

Few would downplay the importance of establishing the stability and effectiveness of any serially-administered measure used to compare baseline and post-concussion cognitive functioning (Eckner, Kutcher, & Richardson, 2011). However, to date, there has been no research examining the stability of baseline scores in the younger group of athletes, ages 10–12. It has been well-documented that significant developmental changes occur in the brain structure and cognitive functions during the pre- to early adolescent period (Blakemore, 2012; Casey, Jones, & Hare, 2008). Maturational alterations in the cortical gray matter may peak around the age of 12 years, with significant changes in cortical white matter in the 9 to 14 year old age range (Blakemore, 2012). It is longstanding knowledge that around the age of 11, cognitive development shifts from a Piagetian concrete operational stage to the formal (more abstract) operational stage (Gruber & Voneche, 1977). Thus, in the years between 10 and 12, one might expect a significant maturational change with regard to problem-solving, executive functioning, and decision-making skills. If indeed, there is such a shift, then might there be a need to more frequently update neurocognitive baselines during those years? It is in this context that we sought to document the stability of serial assessments in a sample of youth athletes, ages 10–12. Such data may help answer the question regarding how often young athletes should undergo baseline testing.

## Methods

### Participants

Participants were 30 youth, predominantly male ($n = 29$, 96%), athletes, ages 10–12 (Mean = 11.6, $SD = 0.6$), participating in lacrosse, basketball, and ice hockey selected from a larger sample of 156 youth athletes competing in a youth sports league. Athletes were selected if they: (a) were between the ages of 10 and 12 years of age, (b) had no history of concussion, (c) had no history of attention deficit disorder or learning disability (as determined by chart review and ImPACT questionnaire), and (d) had completed two successive preseason annual baseline evaluations (mean time between assessments 364 days, *SD* 8 days).

### Materials

All participants completed the online version of the ImPACT test. The ImPACT instrument is a computer-based program used to assess neurocognitive function and concussion symptoms. It consists of six 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. The ImPACT test manual (Lovell, 2011) provides normative data for individuals as young as 10 years of age. For more detail on the ImPACT test, see Iverson, Brooks, Collins, and Lovell (2006).

### Analyses

Participants completing two baseline assessments comprised 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, Reaction Time composite scores from ImPACT, and the Total Symptom scores. All dependent variables were screened for outliers (>3.0 *SD*), and one case was removed from the analysis of Symptom Scores. Following this, normality was established using the Shapiro-Wilk test of normality. Only Symptom scores were found to be non-normally distributed, so Spearman's Rank-Order correlations were used in place of Pearson's for this variable.

Pearson product-moment correlations (r) were calculated as a general measure of the strength of linear association between variables at Times 1 and 2. Given the weakness of Pearson's r as a measure of test-retest reliability, when coefficients are high, group means are similar, but there is considerable individual variation in scores from Time 1 to Time 2 (Rodgers & Nicewander, 1998). Thus, ICCs, often considered a better measure of association than Pearson's r, (Wilk et al., 2002) were also calculated as the primary indicator of test–retest reliability, as ICC can distinguish those sets of scores that are merely ranked in the same order from test to re-test from those that are not only ranked in the same order but are in low, moderate, or complete agreement (Chicchetti, 1994). ICC model "Two-Way Mixed" type "Consistency" was used, documenting average measure ICCs. In addition, ICC analyses yield an Unbiased Estimate of Reliability (UER), which reflects the general consistency of the baseline assessments (Fleishman & Benson, 1987). The minimum acceptable reliability coefficient (ICC) has been recommended as .60 (Anastasi, 1998) with a coefficient of .75 representing good reliability (Portney & Watkins, 2009). With respect to making clinical decisions regarding an athlete's cognitive status after concussion, it has been suggested that coefficients must be above .90 (Randolph, McCrea, & Barr, 2005), while others have characterized the clinical significance of coefficients below .40 as "poor," between .40 and .59 as "fair," between .60 and .74 as "good," and above .75 as "excellent"(Landis & Koch, 1977).

Finally, regression-based methods (RBM) were applied to the data. In 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 initial performance level, as well as controlling for any regression to the mean (McCrea et al., 2005). With this technique, regression equations can be built to predict a participant's level of performance on a neuropsychological instrument at re-test from the initial testing (Crawford & Garthwaite, 2006). Following Bonferroni correction, alpha level for all analyses was set at $p < .01$.

## Results

Correlations between baseline assessments ranged from .34 to .71, with Verbal Memory and Motor Speed scores at .71 and .69, respectively. Average Measures ICCs reflected higher reliability than Pearson's r, across all measures. Total Symptom scores (ICC =.836) scores showed the most stability, followed by Verbal Memory (.826), Motor Processing Speed (.813), Visual Memory (.516), and Reaction Time (.504). Unbiased Estimates of Reliability were consistent with ICCs: Total Symptom scores (.846), Verbal Memory (.838), Motor Processing Speed (.826), Visual Memory (.550), and Reaction Time (.538). Mean ImPACT composite and symptom scores showed significant improvement between the two assessments on Visual Memory and Motor Processing Speed indices (Table 1).

Regression-based measures using 90 and 95% confidence intervals assume that 90 and 95% of cases will fall within these ranges. Results showed that follow-up baseline scores showed considerable stability (Table 2). All scores from follow-up baseline assessments fell within 95% confidence intervals. In addition, all but one score from follow-up baseline assessments (Visual Memory) fell within a 90% confidence interval.

## Discussion

We documented variable reliability of ImPACT composite scores across serial, 1-year assessments, in a sample of youth athletes, ages 10–12. Using traditional measures of reliability (Pearson's correlation, ICCs),

**Table 1.** One-year test–retest reliability data in 30 athletes ages 10–12.

| Variable | Time 1 | Time 2 | *r* | ICC | UER | $t^a$ | sig | *d* |
|---|---|---|---|---|---|---|---|---|
| Verbal (*M*) | 82.70 | 83.60 | .71 | .826 | .838 | −0.7 | .510 | .12 |
| Memory (*SD*) | 9.90 | 8.60 | | | | | | |
| Visual (*M*) | 69.20 | 76.20 | .35 | .516 | .550 | −3.2 | .003 | .60 |
| Memory (*SD*) | 10.90 | 9.50 | | | | | | |
| Motor (*M*) | 26.60 | 29.10 | .69 | .813 | .826 | −4.7 | .001 | .86 |
| Speed (*SD*) | 3.40 | 4.00 | | | | | | |
| Reaction (*M*) | .71 | .69 | .34 | .504 | .538 | 1.3 | .200 | .21 |
| Time (*SD*) | .08 | .09 | | | | | | |
| Symptom (*M*) | 1.70 | 1.40 | .61 | .836 | .847 | −0.7 | .510 | .16 |
| Scale (*SD*) | 2.10 | 2.00 | | | | | | |

*Note.* ICC = Intra-class correlation coefficient, two-way random, average measures; UER = Unbiased Estimate of Reliability.
[a]df = 29; Bonferonni corrected alpha $p < .01$.

**Table 2.** Rates of impairment using regression based methods.

| RBM | | | | | | |
|---|---|---|---|---|---|---|
| Variable | 90%[a] (Impr/Decl/Tot)% | | | 95%[a] (Impr/Decl/Tot)% | | |
| Verbal Memory | 0 | 3 | 3 | 0 | 0 | 0 |
| Visual Memory | 6 | 6 | 12 | 0 | 3 | 3 |
| Motor Speed | 3 | 3 | 6 | 3 | 0 | 3 |
| Reaction Time | 0 | 9 | 9 | 0 | 6 | 6 |
| Symptom Scale | 0 | 6 | 6 | 0 | 0 | 0 |

*Note.* RBM = Regression-Based Measures.
[a]CI = Confidence Interval; numbers represent percentage of participants scoring beyond cut-off values at 90% (1.64) and 95% (1.96) Confidence Intervals.

reliability was high for Verbal Memory, Visual Motor Speed, and Symptom Scores, but lower for Visual Memory and Reaction Time. This is out of register with previous research on Reaction Time scores, which, along with Visual Motor Speed, have been among the most stable, representing the "speed" component of the ImPACT test (Schatz & Maerlender, 2013). It is not clear if the current results are unique based on the younger age of the sample, although similar "small-N" studies on college-aged students (Schatz & Maerlender, 2013) yielded higher reliability data ($r = .63$, ICC $= .77$) for Reaction Time than the current study ($r = .34$, ICC $= .50$). Other researchers have also documented higher reliability data for Reaction Time ($r = .63$ to .76) and Motor Processing Speed (.65 to .83) as compared to Verbal Memory ($r = .19$ to .40) and Visual Memory ($r = .36$ to .55) scores in high school and college athletes (Register-Mihalik et al., 2012); in general, ImPACT subscales measuring "speed" (Reaction Time, Motor Processing Speed) have higher reliability than those measuring memory (Verbal and Visual Memory) (Schatz & Maerlender, 2013). However, using Regression-based measures, which account for practice effects, very few athletes fell outside 90% confidence intervals, and none outside and 95% confidence intervals reflecting stability. Given that human performance is expected to vary from time-to-time, and youth athletes are expected to display maturation throughout their adolescence, the stability demonstrated by regression-based measures is encouraging. Nonetheless, given the variable reliability using traditional measures, annual updates of baseline assessments remains warranted.

The use of computerized neurocognitive testing has become commonplace in many high schools and youth athletic leagues across the country. While most of the available neurocognitive assessment batteries for baseline and post-concussion testing appear to measure similar constructs (i.e., working memory, processing speed, attention and concentration, reaction time), reaction time has been identified as a unique and sensitive construct (Collie et al., 2003; Norris, Carr, Herzig, Labrie, & Sams, 2013). Athletes with a history of multiple previous concussion have been shown to exhibit deficits in reaction time as compared to those with a history of zero or one previous concussion (Covassin, Moran, & Wilhelm, 2013; Gardner, Shores, & Batchelor, 2010). In this context, increased variability in a youth sample, with respect to Reaction Time data/performance, may translate to inappropriate interpretation of changes in post-concussion reaction time performance, as being reflective of changes in mentation or psychomotor speed, when they are actually due to variation in the measure or athlete over time.

Implementation and administration of testing programs are often costly and time consuming, utilizing personnel resources as well as straining school/organizational budgets. As such, updating of baseline testing is sometimes considered impractical. Furthermore, schools often limit testing to a select group of athletes, considering the factors of level of sport (varsity vs. junior varsity vs. intramural) and/or level of risk (noncontact vs. contact vs. collision). Independent youth sport leagues often have fewer resources and funding, with fewer available computer laboratories, supervising faculty, and so forth. Yet, with increased public concern over youth concussion and in the face of media-fueled fears of future chronic traumatic encephalopathy, baseline concussion testing programs are now filtering down to the middle school and younger athlete level. Despite good intentions, these programs may lack blueprints for implementation, and limited research to guide concussion practice and management in this younger age group.

Because youth brains experience dynamic changes and significant cognitive development (Blakemore, 2012; Casey, 2008; Gruber, 1977), there is a question as to what degree neurocognitive baseline profiles may change over a 1–2 year period. Research on high school and college students has demonstrated that computerized neurocognitive testing, such as ImPACT, remains stable over time (Elbin et al., 2011; Schatz, 2009). However, no parallel studies have been conducted in the younger 10 to 12 year old group. Considering the pressing economic need to avoid overutilization of health resources, and the real limitations of youth sports resources and finances, the important question becomes, "How often should the younger athlete take the baseline test?" Although guidelines have documented the need for re-testing younger athletes on a yearly basis, there has been no evidence until now to suggest that yearly baseline testing may be of any use.

The current results reveal that "improvement" may be present in the domains of visual memory and reaction time across a 1-year period in 10- to 12-year old athletes, as evidenced by lower ICCs and UERs. Furthermore, Symptom scores at baseline times 1 and 2 were significantly different, perhaps reflecting the prepuberty to puberty shift in physical, hormonal body changes that occur during the 10- to 12-year age range. Understanding the change in Symptom scores is important considering that, in concussion research, Symptom scores have often been heavily relied upon to examine recovery. Thus, these findings alone may warrant the use of annual baseline assessments in this younger age group.

An alternative hypothesis for these findings might include examining the characteristics of the instrument or tool itself, suggesting that perhaps it could be improved upon to positively affect its inherent reliability. However, if employing regression-based measures for the tool's analysis, which control for possible practice effects and regression to the mean, the results also suggest that all scores still fall within expected confidence intervals. In this regard, one could argue that annual baseline test data may be somewhat stable in 10- to 12-year old athletes, when controlling for possible practice effects/improvement and regression to the mean, as was found in high school and collegiate athletes.

The present study is preliminary and the generalizability may be hampered by inherent limitations. First, it is retrospective in nature, with restricted representation of sport, gender, and stratification of participants. Second, the sample size is small and reflective of a single site. It is not known whether a larger, multi-site study may have produced similar results. Third, the present results are specific to one type of computerized baseline test, ImPACT. It is possible that other types of baseline testing, whether computerized or other, may have presented different test–retest reliability data.

The present study represents a preliminary examination of an important practical question that has clinical implications as well as economic implications for how baseline testing is implemented and executed with younger athletes. In sum, these data support repeating baseline testing on annual basis in the 10- to 12-year old athlete group, especially with consideration of possible maturational changes in visual memory, reaction time and symptom report over time. Further research is recommended to better examine the multiple neurocognitive domains affected by concussion, with large groups of athletes involved in a multitude of sports.

## Declaration of interest

Dr. Schatz is a consultant to ImPACT. Dr. Moser was once a consultant to ImPACT more than at least five years ago. Dr. Moser is the director and owner of the center in which the study took place.

## Funding

Study was conducted with administrative office support by the private practice setting in which it was conducted: The Sports Concussion Center of New Jersey.

## References

Anastasi, A. (1998). *Psychological testing* (7th ed.). New York, NY: MacMillan.

Blakemore, S. (2012). Imaging brain development: The adolescent brain. *NeuroImage*, *61*(2), 397–406. doi:10.1016/j.neuroimage.2011.11.080

Casey, B., Jones, R. M., & Hare, T. A. (2008). The adolescent brain, *Annals of the New York Academy of Sciences*, *1124*, 111–126.

CDC. (2015). *Sports concussion policy and laws*. Retrieved July 27, 2015, from http://www.cdc.gov/headsup/policy/index.html

Chicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 294–290. doi:10.1037//1040-3590.6.4.284

Collie, A., Maruff, P., Makdissi, M., McCrory, P., McStephen, M., & Darby, D. (2003). CogSport: Reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clinical Journal of Sport Medicine*, *13*(1), 28–32. doi:10.1097/00042752-200301000-00006

Covassin, T., Moran, R., & Wilhelm, K. (2013). Concussion symptoms and neurocognitive performance of high school and college athletes who incur multiple concussions. *The American Journal of Sports Medicine*, *41*(12), 2885–2889. doi:10.1177/0363546513499230

Crawford, J. R., & Garthwaite, P. H. (2006). 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*, *20*(3), 259–271. doi:10.1037/0894-4105.20.3.259

Echemendia, R. (2013). *Legislation update and Zurich II*. Paper presented at the International Conference on Sports Concussion, Sports Neuropsychology Society, Minneapolis, MN, May 3–4.

Eckner, J. T., Kutcher, J. S., & Richardson, J. K. (2011). Between-seasons test-retest reliability of clinically measured reaction time in national collegiate athletic association division I athletes. *Journal of Athletic Training*, *46*(4), 409–414.

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*, 2319–2324. doi:10.1177/0363546511417173

Fleishman, J., & Benson, J. (1987). Using lisrel to evaluate measurement models and scale reliability, *Educational and Psychological Measurement*, *47*, 925–939. doi:10.1177/0013164487474008

Gardner, A., Shores, E. A., & Batchelor, J. (2010). Reduced processing speed in rugby union players reporting three or more previous concussions. *Archives of Clinical Neuropsychology*, *25*(3), 174–181. doi:10.1093/arclin/acq007

Gruber, H., & Voneche, J. J. (1977). *The essential Piaget*. New York, NY: Basic Books.

Hunt, T. N., & Ferrara, M. S. (2009). Age-related differences in neuropsychological testing among high school athletes. *Journal of Athletic Training*, *44*(4), 405–409. doi:10.4085/1062-6050-44.4.405

Iverson, G. L., Brooks, B. L., Collins, M. W., & Lovell, M. R. (2006). Tracking neuropsychological reovery following concussion in sport. *Brain Injury*, *20*(3), 245–252. doi:10.1080/02699050500487910

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. doi:10.2307/2529310

Lichtenstein, J., Moser, R. S., & Schatz, P. (2013). Age and test setting affect the prevalence of invalid baseline scores on neurocognitive tests, *The American Journal of Sports Medicine*, *42*, 479–484. doi:10.1177/0363546513509225

Lovell, M. R. (2011). *ImPACT technical manual: Online ImPACT 2007–2012* (Vol. *2013*). Pittsburgh, PA: ImPACT Applications.

MacDonald, J., & Duerson, D. (2015). Reliability of a computerized neurocognitive test in baseline concussion testing of high school athletes. *Clinical Journal of Sport Medicine*, *25*(4), 367–372. doi:10.1097/JSM.0000000000000139

MacDonald, J., Wilson, J., Young, J., Duerson, D., Swisher, G., Collins, C. L., & Meehan, W. P. 3rd. (2015). Evaluation of a simple test of reaction time for baseline concussion testing in a population of high school athletes. *Clinical Journal of Sport Medicine*, *25*(1), 43–48. doi:10.1097/JSM.0000000000000096

McCrea, M., Barr, W. B., Guskiewicz, K., Randolph, C., Marshall, S. W., Cantu, R.,… Kelly, J. P. (2005). Standard regression-based methods for measuring recovery after sport-related concussion. *Journal of the International Neuropsychological Society*, *11*(1), 58–69. doi:10.1017/s1355617705050083

McCrory, P., Meeuwisse, W. H., Aubry, M., Cantu, B., Dvorak, J., Echemendia, R. J. & Turner, M. (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*(5), 250–258. doi:10.1136/bjsports-2013-092313

Moser, R. S., Schatz, P., & Lichtenstein, J. (2015). The importance of proper administration and interpretation of neuropsychological baseline and post-concussion computerized testing. *Applied Neuropsychology: Child*, *4*(1), 41–48. doi:10.1080/21622965.2013.791825

Norris, J. N., Carr, W., Herzig, T., Labrie, D. W., & Sams, R. (2013). ANAM4 TBI reaction time-based tests have prognostic utility for acute concussion. *Military Medicine*, *178*(7), 767–774. doi:10.7205/milmed-d-12-00493

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Randolph, C., McCrea, M., & Barr, W. B. (2005). Is neuropsychological testing useful in the management of sport-related concussion? *Journal of Athletic Training*, *40*(3), 139–152.

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. doi:10.4085/1062-6050-47.3.13

Rodgers, J. L., & Nicewander, W. A. (1998). Thirteen ways to look at the correlation coefficient, *The American Statistician*, *42*, 59–66. doi:10.1080/00031305.1988.10475524

Schatz, P. (2009). Long-term test-retest reliability of baseline cognitive assessments using ImPACT, *The American Journal of Sports Medicine*, *38*, 47–53. doi:10.1177/0363546509343805

Schatz, P., & Maerlender, A. (2013). A two-factor theory for concussion assessment using ImPACT: Memory and speed. *Archives of Clinical Neuropsychology*, *28*(8), 791–797. doi:10.1093/arclin/act077

Wilk, C. M., Gold, J. M., Bartko, J. J., Dickerson, F., Fenton, W. S., Knable, M. & Buchanan, R. W. (2002). Test-retest stability of the repeatable battery for the assessment of neuropsychological status in schizophrenia. *The American Journal of Psychiatry*, *159*(5), 838–844. doi:10.1176/appi.ajp.159.5.838