

# Prospective, Head-to-Head Study of Three Computerized Neurocognitive Assessment Tools (CNTs): Reliability and Validity for the Assessment of Sport-Related Concussion

Lindsay D. Nelson,<sup>1</sup> Ashley A. LaRoche,<sup>1</sup> Adam Y. Pfaller,<sup>1</sup> E. Brooke Lerner,<sup>1</sup> Thomas A. Hammeke,<sup>1,2</sup> Christopher Randolph,<sup>3</sup> William B. Barr,<sup>4</sup> Kevin Guskiewicz,<sup>5</sup> AND Michael A. McCrea<sup>1,2</sup>

<sup>1</sup>Medical College of Wisconsin, Milwaukee, Wisconsin

<sup>2</sup>Clement J. Zablocki VA Medical Center, Milwaukee, Wisconsin

<sup>3</sup>Loyola University Medical School, Maywood, Illinois

<sup>4</sup>New York University School of Medicine, New York, New York

<sup>5</sup>University of North Carolina at Chapel Hill, Chapel Hill, NC

(RECEIVED May 21, 2015; FINAL REVISION October 8, 2015; ACCEPTED October 13, 2015)

## Abstract

Limited data exist comparing the performance of computerized neurocognitive tests (CNTs) for assessing sport-related concussion. We evaluated the reliability and validity of three CNTs—ANAM, Axon Sports/Cogstate Sport, and ImPACT—in a common sample. High school and collegiate athletes completed two CNTs each at baseline. Concussed ( $n = 165$ ) and matched non-injured control ( $n = 166$ ) subjects repeated testing within 24 hr and at 8, 15, and 45 days post-injury. Roughly a quarter of each CNT's indices had stability coefficients ( $M = 198$  day interval) over .70. Group differences in performance were mostly moderate to large at 24 hr and small by day 8. The sensitivity of reliable change indices (RCIs) was best at 24 hr (67.8%, 60.3%, and 47.6% with one or more significant RCIs for ImPACT, Axon, and ANAM, respectively) but diminished to near the false positive rates thereafter. Across time, the CNTs' sensitivities were highest in those athletes who became asymptomatic within 1 day before neurocognitive testing but was similar to the tests' false positive rates when including athletes who became asymptomatic several days earlier. Test–retest reliability was similar among these three CNTs and below optimal standards for clinical use on many subtests. Analyses of group effect sizes, discrimination, and sensitivity and specificity suggested that the CNTs may add incrementally (beyond symptom scores) to the identification of clinical impairment within 24 hr of injury or within a short time period after symptom resolution but do not add significant value over symptom assessment later. The rapid clinical recovery course from concussion and modest stability probably jointly contribute to limited signal detection capabilities of neurocognitive tests outside a brief post-injury window. (*JINS*, 2016, 22, 24–37)

**Keywords:** Concussion, Sport, Computerized testing, ANAM, Axon Sports, ImPACT

## INTRODUCTION

Neuropsychological testing is recognized as an important component in the assessment of athletes with sport-related concussion (SRC; Echemendia et al., 2013; McCrory et al., 2013; Moser et al., 2007). Over the last 10–15 years, computerized neurocognitive testing (CNT) has become especially popular in the sports medicine community (Covassin, Elbin, & Stiller-Ostrowski, 2009; Meehan, d'Hemecourt, Collins, Taylor, & Comstock, 2012; Resch, McCrea, & Cullum, 2013). CNTs have several purported advantages over traditional

paper-and-pencil neuropsychological tests, including the ability to (1) baseline test multiple athletes simultaneously, (2) administer and interpret tests in the absence of neuropsychologists, (3) maximally standardize components of test administration, (4) readily use alternate test forms (*via* randomized presentation of stimuli), (5) quantify reaction time, and (6) take advantage of centralized data repositories (Collie, Darby, & Maruff, 2001; Rahman-Filipiak & Woodward, 2014).

Although these features have undoubtedly contributed to the rapid adoption of CNTs into routine sports medicine practice, this trend has not occurred without controversy. The major concerns raised revolve around baseline testing practices (e.g., testing athletes in group settings that contribute to poor estimation of premorbid abilities; Lichtenstein,

Correspondence and reprint requests to: Lindsay Nelson, 8701 West Watertown Plank Road, Milwaukee, WI. E-mail: linelson@mcw.edu.

Moser, & Schatz, 2014; Moser, Schatz, Neidzowski, & Ott, 2011), the limited assessment and psychometrics training of some professionals who administer and interpret the tests (Moser, Schatz, & Lichtenstein, 2015), and the fact that much of the research has been conducted by the test developers themselves (Cernich, Reeves, Sun, & Bleiberg, 2007). Most problematic is that the reliability and validity of neurocognitive testing for concussion assessment has not been adequately demonstrated. A 2005 review of neuropsychological testing for sport-related concussion concluded that no neuropsychological tests (paper-and-pencil or computerized) met the minimum criteria needed to establish their utility in SRC assessment due to the very limited base of published research establishing the psychometric properties and performance of any test under conditions that are clinically relevant for concussion management (Randolph, McCrea, & Barr, 2005). While the number of published studies on CNTs has significantly increased since that time (for a review see Resch, McCrea, et al., 2013), there is little published work directly comparing the performance of the currently available CNTs, which precludes informed decision-making about which CNT to use.

This gap in the literature was the impetus for Project Head to Head, an independent, prospective study aimed at comparing the reliability, validity, and clinical utility of several popular CNTs for the assessment of sport-related and civilian concussion (or mild traumatic brain injury, mTBI). The study enrolled athletes in its sport-related concussion (SRC) arm from 2012 to 2014. Here, we present findings on the test–retest reliability, sensitivity, and specificity of the three CNTs (ANAM, Axon, ImPACT) used in the study's athlete sample.

### Test–Retest Reliability of ANAM, Axon, and ImPACT

Reported test–retest reliability coefficients for ANAM, Axon (or CogSport), and ImPACT from prior studies are somewhat difficult to compare, owing to differences in samples, test–retest intervals, and choice of stability coefficient (i.e., Pearson or intraclass correlation, ICC).<sup>1</sup> Several samples have been rather small for correlational analysis, some test–retest intervals used have been too short to be of clinical relevance (e.g., 1 week), and no studies have directly compared the reliability of these three CNTs within the same athlete sample.

<sup>1</sup> It is worth mentioning that there has been debate about whether Pearson or ICCs are more appropriate for the estimation of test–retest reliability. Those who advocate for the use of ICCs tend to cite the statistic's ability to take into account systematic error (e.g., practice effects; Weir, 2005). Further complicating this debate is that numerous formulas for the ICC exist, some of which do not take into account systematic error. This underscores the importance that researchers specify the formula they are using when reporting ICCs. In contrast, proponents of Pearson correlations have pointed out that, given the classic definition of reliability (i.e., the proportion of true score variance over total variance), practice effects could reflect changes in true score variance and therefore should not be accounted for in the denominator of a reliability coefficient (Rousson, Gasser, & Seifert, 2002). The aim of this manuscript is not to contribute to this debate but rather to acknowledge it while summarizing findings from different methods.

Reports of the stability of performance on each CNT have varied widely by study. Across three studies of ANAM, only 9 of 19 (47%) of reported reliability coefficients met minimal standards for clinical use (.60 or more; Cernich et al., 2007; Register-Mihalik et al., 2013; Segalowitz et al., 2007). Reports of Axon's stability have varied from finding only 2 of 5 Pearson coefficients to be over .60 (MacDonald & Duerson, 2015) to reporting strong stability (range, .83–.94) for all 4 indices (Louey et al., 2014); see also (Collie et al., 2003; Eckner, Kutcher, & Richardson, 2011; Straume-Naesheim, Andersen, & Bahr, 2005).<sup>2</sup> A larger number of studies have been published on the reliability of ImPACT in high school (Elbin, Schatz, & Covassin, 2011; Iverson, Lovell, & Collins, 2003; Register-Mihalik, Kontos, et al., 2012), collegiate (Iverson et al., 2003; Nakayama, Covassin, Schatz, Nogle, & Kovan, 2014; Register-Mihalik, Kontos, et al., 2012; Resch, Driscoll, et al., 2013; Schatz, 2010), and professional (Bruce, Echemendia, Meeuwisse, Comper, & Sisco, 2014) athletes as well as non-athlete students (Broglia, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007; Schatz & Sandel, 2013). Reliability coefficients for ImPACT have been uniformly poor in some samples (e.g., ICCs .23–.39 in 73 college students tested 45 days apart; Broglia, Ferrara, et al., 2007) and consistently stronger (over .60) in others (Iverson et al., 2003; Schatz & Ferris, 2013).

Given that correlation coefficients are inherently sensitive to sample-specific factors (e.g., degree of heterogeneity), it is all the more important to obtain these estimates from comparable samples and to use equivalent test–retest intervals before conclusions can be drawn about the relative stability of indices from different CNTs. The one study that evaluated the reliability of these three CNTs (along with CNS-Vital Signs) in a military sample tested approximately 30 days apart reported that, although select subtests from each CNT demonstrated adequate reliability, overall the coefficients appeared lower than is desired for clinical decision-making (Cole et al., 2013).

### Group-Level Sensitivity to Concussion

Publications presenting concussed *versus* control group effect sizes for CNT measures are also similarly difficult to compare due to variability in samples, post-injury time points, and statistical methods across studies. Consistent with findings on the neurocognitive sequelae of concussion for other measures, the literature has revealed moderate to large neurocognitive impairments within 1–3 days post-injury on ImPACT whether concussed athletes are compared to their own baselines (Iverson, Brooks, Collins, & Lovell, 2006; Iverson et al., 2003; McClincy, Lovell, Pardini, Collins, & Spore, 2006) or to non-injured controls (Schatz, Pardini,

<sup>2</sup> In general, reliability coefficients for Axon appear stronger for reaction time *versus* accuracy-based metrics, probably due to range restriction in accuracy measures. Because of its limited psychometric properties, the working memory accuracy measure (reported on in some of these cited studies) has since been dropped as a core clinical measure by Axon.

Lovell, Collins, & Podell, 2006; Schatz & Sandel, 2013), with effect sizes diminishing 1 week or more post-injury. The ANAM battery has limited published data on athletes but has demonstrated statistically significant impairments within 10 days of injury in a small high school sample (Sim, Terryberry-Spohr, & Wilson, 2008) and, in another sample, significant impairments on two (of six) indices 1–2 days post-injury with resolution by 3–7 days (Bleiberg et al., 2004). Axon has also demonstrated large concussed *versus* control group effects ( $d = -.94$  to  $-2.95$ ) in symptomatic Australian Rules Football and Rugby players tested 26–42 hr post-injury (Louey et al., 2014).

### Sensitivity and Specificity of Reliable Change Indices

Because athletes at greatest risk of concussion are readily identified (by virtue of participating in contact and collision sports), many sports medicine professionals baseline test teams of athletes pre-season so that they can apply reliable change indices (RCIs) produced by each CNT to estimate whether concussed athletes have returned to their premorbid levels of functioning (Covassin, Elbin, Stiller-Ostrowski, & Kontos, 2009; Meehan et al., 2012). RCIs were first proposed to estimate whether individual patients benefitted from psychotherapy interventions (Jacobson & Truax, 1991) and are computed by dividing the change in some measure between two time points (e.g., neurocognitive performance from baseline to post-concussion) by the standard error of the difference. This results in a score that can be compared to standard Z score cutoffs to determine whether an individual's change score is statistically unusual after accounting for chance variation. Thus, RCIs provide a theoretical advantage over the application of normative cutoffs in that they facilitate clinical decisions by formally accounting for individuals' pre-injury abilities, measurement error, and in some cases expected practice effects (Chelune, Naugle, Lüders, Sedlak, & Awad, 1993).

However, the sensitivity and specificity of the RCIs provided by the available CNTs have not been adequately documented for all available CNT programs, and no studies have focused analyses of the RCIs' sensitivity in the subpopulation of concussed athletes for which neurocognitive testing could add value to concussion assessments: those who have become asymptomatic and would be otherwise cleared for participation unless clinical testing (neurocognitive or other) indicated lingering impairment that would alter the clinician's decision on the athlete's readiness to return to play. Because current guidelines preclude returning athletes to play until symptom-free (i.e., free of symptoms initiated or exacerbated by the concussive injury), the inclusion of symptomatic athletes in most estimates of sensitivity may overestimate the degree to which neurocognitive test results would alter clinical decision making. Given the time, expense, and expertise needed to properly administer and interpret neurocognitive tests, their added value to concussion assessment relies on demonstrating that they reliably and

validly identify impairments beyond freely and quickly administered symptom measures.

Previous reports of the sensitivity and specificity of CNTs are difficult to compare for a variety of reasons. For example, several studies have reported on concussed athletes only (disregarding specificity) or emphasized the sensitivity and specificity of individual indices within a CNT rather than presenting findings across the set of available scales within each battery. Given that clinicians are faced with interpreting the outcomes of multiple RCIs simultaneously, documenting the joint base rates of impairment in both concussed and non-concussed athletes is essential to determining the validity of the measures. Furthermore, reports that have aggregated neurocognitive and symptom measures do not directly address the added value of neurocognitive measures over symptom scores. Finally, since the confidence levels applied to the RCIs to determine significance vary by test manufacturer [90% confidence intervals (CIs) for ANAM and Axon; 80% CIs for ImPACT], the expected specificities (and by extension, sensitivities) are not equal across all measures.

The majority of published studies on this topic have focused on ImPACT, which is the most widely used CNT in athletic settings (Meehan et al., 2012). Perhaps in part due to the reasons cited above, the sensitivity and specificity of ImPACT's RCI criteria have varied across studies. The percentage of concussed athletes with one or more significantly declined RCIs on ImPACT has ranged from 62.5–83% at 1–2 days post-injury (Broglia, Macciocchi, & Ferrara, 2007; Iverson et al., 2003; Van Kampen, Lovell, Pardini, Collins, & Fu, 2006), with 90% of concussed athletes showing 2 or more significant RCIs in another sample (Iverson et al., 2006). Specificity values have also varied quite a bit by sample and, as expected, have improved as criteria for significant change were made more stringent (Iverson et al., 2003; Resch, Driscoll, et al., 2013). Reports of the RCIs used by ANAM and Axon are more limited in scope. One study of ANAM reported 0–11% sensitivity (90% CIs) on each subtest of the battery, with only 50% sensitivity (and 95% specificity) across a battery incorporating ANAM data with that of a symptom checklist and the Sensory Organization Test (Register-Mihalik, Guskiewicz, et al., 2012). A single study of Axon found 100% sensitivity to SRC (one or more significant RCIs with 90% CIs) but only 50.8% specificity (Louey et al., 2014).

### Current Study

The aim of this study was to quantify and compare the reliability and validity of three CNTs—ANAM, Axon, and ImPACT—in the context of sport-related concussion assessment. More specifically, we were interested in characterizing the psychometric properties and clinical performance of the CNTs under conditions in which they are used in routine sports medicine practice, including using relevant test–retest intervals as well as examining the RCIs produced by each CNT's standard software package. Consistent with

prior research, we hypothesized that (1) test–retest reliability coefficients in the control sample would vary across indices within each CNT and would be larger for shorter *versus* longer test–retest intervals, (2) concussed *versus* control group effect sizes would be moderate to large within 24 hr of injury on some indices from each CNT and would diminish in magnitude further out from injury, (3) the sensitivity of each CNT’s RCIs would be moderately strong within 24 hr of injury and would substantially diminish at the day 8 assessment, and (4) given the multiple indices that are provided in each CNT’s score report and associated issues with multiple comparisons, that the base rates of one or more impairments (per the RCI criteria) in non-injured control sample would be relatively high and would diminish with more stringent criteria for significant change (i.e., two or more significant RCIs within a CNT).

## METHOD

### Participants

Participants were contact and collision sport athletes from 9 high schools and 4 colleges in southeastern Wisconsin enrolled in Project Head to Head between August, 2012 and October, 2014 (see also LaRoche, Nelson, Connelly, Walter, & McCrea, 2015; Nelson, Pfaller, Rein, & McCrea, 2015). Among the 2,148 participants who consented to participate, 166 were concussed during the study and were enrolled in post-injury testing. Ten of those athletes sustained a repeat concussion during their study participation. A sample of 166 non-injured controls were selected to match injured athletes on school, sports team (and by extension gender), estimated premorbid verbal intellectual ability (Wechsler Test of Adult Reading; see baseline testing protocol), cumulative self-reported GPA, and age. Because of limited controls on some sports teams and the numerous matching criteria, 22 injured subjects were matched to a control from another institution. Athletes who had failed to produce any valid CNT at baseline ( $n = 1$ ) were excluded from the analysis, yielding 165 concussed athletes and 166 controls for analysis.

Adult athletes and parents of minor athletes completed informed consent, and minor participants completed assent before their first evaluation. Participants were compensated \$30 for their time and effort in completing baseline assessments and received \$50 for each post-injury assessment. All testing procedures were approved by the Institutional Review Board at the Medical College of Wisconsin.

### Definition of Injury and Acute Injury Characteristics

The definition of concussion used in this study was based on that of the study sponsor, the U.S. Department of Defense: “mTBI is defined as an injury to the brain resulting from an external force and/or acceleration/deceleration mechanism from an event such as a blast, fall, direct impact, or motor

vehicle accident which causes an alteration in mental status typically resulting in the temporally related onset of symptoms such as headache, nausea, vomiting, dizziness/balance problems, fatigue, insomnia/sleep disturbances, drowsiness, sensitivity to light/noise, blurred vision, difficulty remembering, and/or difficulty concentrating” (Helmick et al., 2006).

### Baseline and Post-Injury Test Battery

The study protocol involved testing athletes at pre-season baseline examinations and retesting concussed athletes within 24 hr and at 8 ( $\pm 1$ ), 15 ( $\pm 2$ ), and 45 ( $\pm 5$ ) days post-injury. Occasionally, examinations were scheduled outside the target window to avoid missing data. For the concussed sample, the  $M$  ( $SD$ ) time from injury to the 24-hr assessment was 19.09 (5.09) hr, with  $M$  ( $SD$ ) number of days from injury to the day 8, day 15, and day 45 assessments = 8.16 (.96), 15.37 (1.55), and 45.39 (3.67), respectively. For controls, testing was done as soon after identification as possible and then 7 ( $M$  [ $SD$ ] = 7.10 [.88]), 14 (14.28 [1.22]), and 44 (43.82 [4.15]) days after their initial evaluation. The baseline testing protocol consisted of, in order: Contact Information, Demographics/Health History (gathered by one-on-one interview), Wechsler Test of Adult Reading (WTAR; Wechsler, 2001), CNT #1, Standardized Assessment of Concussion (SAC; McCrea et al., 1998), Sport Concussion Assessment Tool – 3<sup>rd</sup> edition (SCAT3) symptom checklist (McCroory et al., 2013), CNT #2, Green’s Medical Symptom Validity Test (MSVT; Green, 2003),<sup>3</sup> Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985), Brief Symptom Inventory-18 (BSI-18; Derogatis, 2001), and the Balance Error Scoring System (BESS; Guskiewicz, Ross, & Marshall, 2001). Tests were individually proctored by a research assistant in quiet settings with computers positioned to minimize distractions. Baseline testing group sizes ranged from 1–20 athletes; post-injury testing was conducted one-on-one. Each athlete was read a standardized script at the beginning of the baseline testing session and before each of the CNTs about the importance of valid baseline tests. Follow-up protocols began with an interview of recovery information and then followed the same procedure as listed above starting with CNT#1. Baseline testing sessions lasted approximately 90 min and post-injury testing sessions lasted approximately 60 min.

Each athlete took two of three CNTs: Automatic Neuropsychological Assessment Metrics (ANAM v. 4.3; Vista Life Sciences), Axon Sports (Axon/Cogstate Sport;

<sup>3</sup> As has been thoroughly examined in another report on the larger baseline sample from this study (Nelson et al., 2015), MSVT failure was rare and demonstrated poor agreement with the validity output of any CNT. Thus, given our goal to examine the performance of these CNTs in their typical clinical context (in which only the CNT validity criteria are available, and because the MSVT does not appear to measure the same construct as related to performance validity as the CNTs measure, and, we did not exclude subjects from the primary analyses due to failure to pass the MSVT at baseline ( $n = 3$  in this sample). However, see the Supplemental Materials for evidence that the major study findings were not affected by these subjects (Supplementary Tables S1–S2).

Cogstate Ltd.), and Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT, Online version; ImPACT Applications Inc.). These were selected by the study Principal Investigator and study advisors to match the most widely used CNTs in sports medicine at the time of study design. The decision to administer two CNTs to each participant was made to balance the benefits of increased statistical power using a within-subjects, head-to-head design while minimizing the potential for cognitive fatigue associated with performing multiple neurocognitive tests in a single session. CNT pairing groups were assigned to each school with the aim of balancing the demographic distribution across CNTs. Because controls were selected from the same sports teams as the injured subjects they were selected to match, each concussed-control pair took the same two CNTs at each assessment (less 11 pairs who were selected from different institutions that had only one of two CNTs in common). The overall distribution of CNT pairings across the sample evaluated in this manuscript was: 27.2% ANAM-Axon, 40.8% ANAM-ImPACT, and 32.0% Axon-ImPACT. For each subject, order of administration was selected at random by a computer algorithm at the first assessment and repeated for that individual at all follow-up examinations.

## Computerized Neurocognitive Tests

### ANAM

The version of ANAM used in this study included eight subtests: Simple Reaction Time, Code Substitution-Learning, Procedural Reaction Time, Mathematical Processing, Matching to Sample, Code Substitution-Delayed, Simple Reaction Time 2, and Go/No-Go. The score summary produced for the study also included a Composite Score previously derived to aggregate the throughput scores from each subtest (Vincent et al., 2012). ANAM forms used for baseline and post-injury assessments were, in order, forms 1, 2, 3, 4, and 5.

### Axon

The Axon Sports (Cogstate Sport) CNT is comprised of four tasks: Processing Speed (simple reaction time), Attention (choice reaction time), Learning (LN; visual recognition memory) and Working Memory (one-back). Axon baseline and post-injury test protocols are equivalent with stimulus order randomized for every administration.

### ImPACT

ImPACT is comprised of six tasks, Word Memory, Design Memory, X's and O's, Symbol Match, Color Match, and Three Letters, which yield the following neurocognitive composite scores: Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control. The Impulse Control Composite was not included in the analyses because it appears to be intended for the assessment of performance validity. ImPACT alternate forms used for baseline

and post-injury assessments were, in order, the Baseline and Post-Injury forms 1, 2, 3, and 4.

## Data Analysis

### Sample considerations and measures

The majority of the concussed sample ( $n = 133$ ) and the entire control sample enrolled in the study at pre-season baseline testing; an additional 33 concussed athletes enrolled post-injury. As concussed athletes with and without baseline data were statistically equivalent on markers of injury severity (differences on acute injury characteristics and 24-hr symptoms and neurocognitive performance; all unadjusted  $ps > .10$ ), all available subjects were included in the analyses. Repeat injuries ( $n = 10$ ) during the study were not included.

Analyses involving symptom data used the SCAT3 symptom checklist, a 22-item checklist of common post-concussive symptoms in which athletes rate the degree to which they are experiencing each item on a 0–6 (none to severe) scale. Symptom severity scores represent the sum of the item-level scores (range, 0–132), with higher scores reflecting more severe symptoms. Analysis of the CNT data used throughput scores for all ANAM subtests except Go/No-Go, for which d-prime was used, scaled scores for all Axon subtests ( $M = 100$ ;  $SD = 10$ ), and composite scores for all ImPACT subtests. Although some CNTs have embedded symptom checklists, these were excluded from analyses to focus on neurocognitive testing. Preliminary analyses indicated that all measures were reasonably normally distributed (skewness  $\leq \pm 1$ ). Subjects were excluded from analyses of a CNT if they did not produce a valid baseline for that test.

### Test–retest reliability

Reliability for each CNT subscale was quantified for the non-injured control sample using both Pearson correlations ( $r$ ) and Intraclass Correlations (ICC; 2-way mixed, absolute agreement). Test–retest intervals were selected from varying combinations of the available time points to yield a range of retest intervals and to include retest intervals with clinical relevance to sports medicine practice. This yielded the following test–retest intervals: 7 days (24-hr vs. day 8 assessment), 14 days (24-hr vs. day 15), 30 days (day 15 vs. day 45), 44 days (24-hr vs. day 45), and 198 days ( $M$  time interval between pre-season baseline and first repeat examination).

### Group-level sensitivity

Group (concussed, control)  $\times$  Time (baseline, 24 hr, day 8, day 15, day 45) repeated measures analyses of variance (ANOVAs) were computed for each CNT index. Follow-up ANOVAs examined the main effect of Group at each time point within each measure. Adjustment for multiple comparisons was performed using the false discovery rate method (Benjamini & Hochberg, 1995). This approach is a sequential Bonferroni-type procedure that, unlike traditional Bonferroni

correction (which controls the familywise error rate), is aimed at controlling the expected proportion of incorrectly rejected null hypotheses (“false discoveries”) and, consequently, better preserves statistical power while also providing a reasonable degree of control of type I errors (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001). Cohen’s  $d$  was computed from the groups’ descriptive statistics to provide a comparable metric of effect size across the measures. Because concussion histories differed between groups, steps were taken to ensure that this variable did not moderate the reported group differences. In particular, correlations between number of prior concussions and each CNT measure (at each time point) found only 4 comparisons (<5% of unadjusted  $p$ -values) to be statistically significant. Adding concussion history as a covariate in the ANOVA models described above did not in any case change the significance status of the comparison and had no marked influence on the effect sizes reported. Thus the data presented below reflect those of the models computed without the inclusion of concussion history as a covariate. Next, to illustrate how the effect sizes reported translate into utility for individual decision making, receiver operating characteristic (ROC) curves were produced for each index and the area under the curve (AUC) reported.

### Performance of reliable change indices

Finally, a set of analyses were conducted to document the sensitivity and specificity of the standard neurocognitive RCI output for each CNT. The RCIs produced by each CNT software package were selected over sample-derived RCIs to document the performance of the indices routinely used in clinical practice. However, it should be noted that because the manufacturer’s standard RCIs reflect different confidence levels (90% CIs for ANAM and Axon; 80% CIs for ImpACT) and produce differing numbers of RCIs (seven for ANAM and four for Axon and ImpACT), the expected false positive rates are not equivalent and should be interpreted in that context. The version of ANAM used in the study did not provide an RCI for the Go/No-go subtest.

Sensitivity values were computed both for individual subtests/subscales as well as summated across the RCIs for each CNT. To retain a large  $n$  at each time point and maintain consistency with most published literature on these measures, we first computed sensitivity values for the entire concussed sample. However, we also separately computed the sensitivity of each test in asymptomatic concussed athletes, with each athlete classified as symptom-free at each assessment point if they reported feeling recovered of any postconcussive symptoms in our recovery interview.<sup>4</sup> Note that very few

subjects reported recovery within 24 hr of injury ( $ns$  for ANAM, Axon, and ImpACT at 24 hr = 7, 8, and 13, respectively, vs. day 8  $ns$  = 56, 37, and 61). Second, because athletes identified through the first approach (particularly for day 8 and beyond) were tested at variable time points with regard to the number of days since they became asymptomatic, we aggregated all concussed subjects (across all time points) who were tested within 1 day of becoming asymptomatic (based on their self-reported symptom duration in a recovery interview) to estimate the degree to which the CNTs would alter clinical decision making at this important time point. This yielded  $ns$  of “recently” asymptomatic athletes for ANAM, Axon, and ImpACT of 18, 19, and 32, respectively.

## RESULTS

### Sample Characteristics and Course of Symptom Recovery

Table 1 displays the sample characteristics and degree of matching between the concussed and control groups. A total of 162 (97.6%) of the control subjects had been selected as a matched control for one of concussed athletes in the final study sample. The groups were closely matched on age, sex, race, sport, estimated verbal intellectual ability (WTAR score), socioeconomic status, history of neurodevelopmental disorder, grade point average, height, and weight. As described under Data Analysis, the difference in concussion history between groups did not moderate the effects reported below. Among our injured sample, 6.1% exhibited observed loss of consciousness, 10.4% posttraumatic amnesia, and 9.8% retrograde amnesia, consistent with the acute injury characteristics in our other published work on SRC (e.g., McCrea et al., 2003).

Symptom severity scores for the concussed versus control groups were equivalent at baseline and elevated at 24 hr and day 8 (baseline  $M$  [ $SD$ ] = 6.52 [10.23] vs. 5.88 [7.36],  $p$  = .534 [ $d$  = -.07]; 24 hr  $M$  [ $SD$ ] = 24.80 [18.26] vs. 4.48 [5.03],  $p$  < .001 [ $d$  = -1.52]; day 8  $M$  [ $SD$ ] = 7.44 [14.32] vs. 3.19 [5.09],  $p$  < .001 [ $d$  = -.40]). Symptom scores were equivalent by the day 15 assessment ( $p$  = .287;  $d$  = -.12). The percentage of concussed athletes who reported on interview that they had achieved symptom recovery was 10.6% within 24 hr of injury and 64.6%, 85.2%, and 98.6% at the day 8, 15, and 45 assessments, respectively.

### Analysis of Test Order

Because each athlete took two CNTs, analyses were undertaken to ensure that the primary analyses reported were not influenced by test order. To summarize these findings (documented more completely in Supplementary Tables S3–S4, which are available online), we found very little evidence for any effects of test order on the reliability and validity of any of

<sup>4</sup> Classification of symptom status was also conducted by evaluating whether athletes had returned to their reported levels of baseline symptoms on the SCAT3. Analyses of the CNTs’ sensitivity in symptom-free athletes using this approach produced results that were highly consistent with the interview-based approach ( $M$  difference between sensitivity using the interview vs. SCAT3-based classification of symptom recovery was 0.1% across CNTs and time points).

**Table 1.** Sample characteristics

	Concussed	Control	<i>P</i> Value
	<i>N</i> = 165	<i>N</i> = 166	
	<i>M</i> ( <i>SD</i> ) or %	<i>M</i> ( <i>SD</i> ) or %	
Male (vs. female)	83.6%	83.1%	.902
Age (years)	17.46 (1.99)	17.64 (1.80)	.391
College (vs. high school)	60.0%	61.4%	.788
Sport			.999
Football	65.5%	65.7%	
Soccer	23.0%	23.5%	
Field hockey	0.6%	0.6%	
Wrestling	2.4%	1.8%	
Lacrosse	4.2%	3.6%	
Rugby	1.8%	1.8%	
Ice hockey	2.4%	3.0%	
Race			.500
White	85.8%	84.7%	
Black	12.3%	11.7%	
Asian	1.2%	0.6%	
Native Hawaiian/ Pacific Islander	0.0%	1.2%	
Other/unknown	0.6%	1.8%	
Height	70.11 (3.28)	70.42 (3.42)	.404
Weight	188.94 (48.87)	180.69 (37.66)	.089
Grade point average	3.24 (.54)	3.30 (.49)	.331
WTAR standard score	100.87 (12.79)	101.05 (12.03)	.894
Household SES	49.10 (10.70)	47.47 (9.80)	.159
Number of prior concussions	.95 (1.01)	.48 (.79)	<.001
ADHD	9.9%	4.9%	.087
Learning disability	3.7%	3.1%	.751

Note. WTAR = Wechsler Test of Adult Reading standard score; SES = Hollingshead socioeconomic status; ADHD = attention deficit-hyperactivity disorder.

the three CNTs. In regards to test–retest reliability, there was not a consistent advantage for tests administered first or second: the median difference in reliability for each subtest for Order 1–Order 2 was .05 (for both Pearson *r*s and ICCs) and 9 of 17 indices showed higher Pearson reliability coefficients (10 of 17 for ICCs) for Order 1 versus 2. Analyses of overall test performance also revealed no evidence of meaningful effects of test order on performance (no concussion Group × Order interactions, very few main effects of test order that were not in a consistent direction, and no consistent influence of order on the magnitude of concussed vs. control group differences).

### Test–Retest Reliability of CNT Indices

Table 2 displays the test–retest reliability for each CNT subtest for a range of test–retest intervals (7, 14, 30, 45, and 198 days) using both Pearson *r*s and ICCs. Coefficients were

**Table 2.** Test–retest reliability among non-injured controls on ANAM, Axon, and ImPACT: Pearson (and intraclass) correlations

	Test-retest interval				
	7 days	14 days	30 days	44 days	198 days
<b>ANAM</b>					
Composite	.79 (.78)	.72 (.71)	.71 (.70)	.83 (.82)	.71 (.70)
SRT	.55 (.53)	.52 (.50)	.54 (.55)	.57 (.56)	.35 (.36)
CDS	.73 (.72)	.68 (.79)	.65 (.59)	.72 (.69)	.73 (.72)
PRO	.62 (.61)	.43 (.43)	.43 (.41)	.31 (.30)	.57 (.53)
MTH	.74 (.65)	.68 (.58)	.72 (.60)	.78 (.77)	.68 (.66)
M2S	.73 (.72)	.65 (.64)	.55 (.53)	.77 (.75)	.61 (.61)
CSD	.69 (.60)	.64 (.61)	.55 (.55)	.71 (.67)	.63 (.56)
SR2	.66 (.65)	.63 (.62)	.54 (.53)	.61 (.59)	.48 (.48)
GNG	.35 (.34)	.44 (.44)	.35 (.34)	.46 (.45)	.35 (.34)
Mean	.65 (.62)	.60 (.59)	.56 (.53)	.64 (.62)	.57 (.55)
<b>Axon</b>					
PS Speed	.44 (.42)	.59 (.54)	.58 (.32)	.57 (.57)	.49 (.40)
AT Speed	.66 (.65)	.66 (.66)	.64 (.73)	.66 (.66)	.61 (.61)
LN Acc.	.49 (.46)	.40 (.39)	.40 (.34)	.74 (.71)	.51 (.41)
WM Speed	.81 (.80)	.71 (.70)	.69 (.72)	.62 (.63)	.76 (.76)
Mean	.60 (.58)	.59 (.57)	.58 (.53)	.65 (.64)	.59 (.55)
<b>ImPACT</b>					
VERM	.53 (.53)	.50 (.50)	.54 (.55)	.60 (.59)	.51 (.48)
VISM	.50 (.48)	.60 (.58)	.54 (.52)	.59 (.59)	.49 (.49)
VMS	.75 (.75)	.75 (.74)	.66 (.65)	.78 (.78)	.75 (.73)
RT	.67 (.67)	.65 (.64)	.60 (.60)	.67 (.67)	.60 (.60)
Mean	.61 (.61)	.63 (.62)	.59 (.58)	.66 (.66)	.59 (.58)

Note. The 7-day interval = 24-hr to day 8 assessment; 14-day interval = 24-hr to day 15 assessment; 30-day interval = day 15 to day 45 assessment; 44 day interval = 24-hr to day 45 assessment; 198 day interval = baseline to 24-hr (first repeat) assessment. SRT = Simple reaction time; CDS = code substitution-learning; PRO = procedural reaction time; MTH = mathematical processing; M2S = matching to sample; CDD = code substitution-delayed; SR2 = simple reaction time 2; GNG = go no-go; PS = processing speed; AT = attention; LN acc. = learning accuracy; WM = working memory; VERM = verbal memory composite; VISM = visual memory composite; VMS = visual motor speed composite; RT = reaction time composite.

similar between CNTs, with roughly half of the reliability coefficients for each CNT (198-day interval) over .6 (5 of 9 for ANAM and 2 of 4 for both Axon and ImPACT) and roughly a quarter were over .7 (2 for ANAM and 1 for Axon and ImPACT). Counter to expectation, there was not a consistent advantage of a shorter retest interval, *M* Pearson *r* for the 7-day/198-day intervals: ANAM .65/.57, Axon .60/.59, and ImPACT .61/.59.<sup>5</sup>

<sup>5</sup> Additional analyses of age group (high school vs. college) were undertaken on reliability and validity. Stability coefficients were highly comparable for high school versus college athletes. For the baseline to first follow-up test-retest interval (*M* = 198 days), the median High School - College difference in stability for both Pearson and ICCs was .02; 9 of 17 coefficients favored the High School and 8 favored the Collegiate cohort. Furthermore, ANOVAs were performed of each CNT variable (at each time point) using age Level (high school, college) and concussion Group as independent variables. This revealed no significant interactions between Level and Group for any CNT measure at any time point, suggesting that the reported concussion effects of interest were not affected by age group.

**Table 3.** Concussed vs. control group effect sizes (Cohen's *d*)

	BL	24 hr	Day 8	Day 15	Day 45
SCAT3 Symptom Severity	-.07	<b>-1.53</b>	<b>-.41</b>	-.13	.22
ANAM					
Composite	-.17	<b>-.84</b>	-.33	-.30	-.21
SRT	-.04	<b>-.58</b>	-.18	-.12	-.10
CDS	-.06	<b>-.61</b>	-.26	-.17	-.19
PRO	-.32	<b>-.68</b>	-.35	-.26	.03
MTH	-.19	-.19	-.02	-.22	-.09
M2S	-.32	<b>-.89</b>	<b>-.47</b>	<b>-.40</b>	-.36
CDD	.06	<b>-.66</b>	-.10	-.16	-.36
SR2	.08	<b>-.64</b>	-.27	-.15	-.02
GNG	-.13	-.30	-.22	.03	.18
Mean	-.12	<b>-.60</b>	<b>-.25</b>	<b>-.19</b>	<b>-.13</b>
Axon					
PS Speed	-.07	<b>-.53</b>	-.12	-.11	-.03
AT Speed	-.09	<b>-.72</b>	<b>-.41</b>	-.22	-.08
LN Acc.	-.22	<b>-.51</b>	<b>-.39</b>	-.12	-.28
WM Speed	-.07	<b>-.51</b>	-.25	.01	.00
Mean	-.11	<b>-.57</b>	<b>-.29</b>	<b>-.11</b>	<b>-.10</b>
ImPACT					
VERM	.02	<b>-.76</b>	<b>-.40</b>	-.18	-.18
VISM	-.21	<b>-.76</b>	-.17	-.21	-.26
VMS	-.27	<b>-.80</b>	-.31	-.29	-.21
RT	.08	<b>-.70</b>	-.24	-.23	-.18
Mean	-.10	<b>-.76</b>	<b>-.28</b>	<b>-.23</b>	<b>-.21</b>

Note. Bolded where  $p < .05$  after adjustment for multiple comparisons. Comparisons are all scaled such that negative values reflect worse performance in the concussed group. BL = baseline; SRT = Simple reaction time; CDS = code substitution-learning; PRO = procedural reaction time; MTH = mathematical processing; M2S = matching to sample; CDD = code substitution-delayed; SR2 = simple reaction time 2; GNG = go no-go; PS = processing speed; AT = attention; LN acc. = learning accuracy; WM = working memory; VERM = verbal memory composite; VISM = visual memory composite; VMS = visual motor speed composite; RT = reaction time composite.

### Group Performance and Effect Sizes of CNT Measures at Baseline and Follow-Up Assessments

Supplementary Tables S5–S7 display the descriptive statistics and statistical significance of Group × Time and Group ANOVAs for ANAM, Axon, and ImPACT. Table 3 displays the concussion by control group effect sizes (Cohen's *d*) for each CNT index at each assessment (with *ds* all scaled such that negative values indicate worse performance in the concussed group). Effect sizes of SCAT3 symptom ratings are provided in Table 3 for comparison to neurocognitive measures and to clarify the subjective recovery of this sample.

The groups were statistically equivalent on baseline performance for all CNT indices. The vast majority of indices (7/8 for ANAM, 4/4 for Axon, 4/4 for ImPACT) demonstrated statistically significant differences between groups at 24 hr and most effect sizes were moderate in size (ANAM *ds* = .19 to .89; Axon *ds* = .51 to .72; ImPACT *ds* = .70 to .80). Only 4 of 17 neurocognitive indices (ANAM Matching to Sample, Axon Attention and Learning, and ImPACT Verbal Memory) were significantly different

**Table 4.** Area under the receiver operating characteristic (ROC) curve

	BL	24 hr	Day 8	Day 15	Day 45
SCAT3 Symptom Severity	.50	<b>.87</b>	.53	.48	.44
ANAM					
Composite	.56	<b>.72</b>	.59	.59	.56
SRT	.50	<b>.63</b>	.54	.55	.53
CDS	.54	<b>.67</b>	.56	.55	.56
PRO	.58	<b>.66</b>	.58	.56	.51
MTH	.56	.57	.52	.57	.53
M2S	.59	<b>.73</b>	<b>.64</b>	<b>.61</b>	.60
CDD	.48	<b>.68</b>	.53	.55	.60
SR2	.46	<b>.65</b>	.57	.53	.50
GNG	.53	.58	.55	.47	.45
Mean	.53	<b>.65</b>	<b>.56</b>	<b>.55</b>	<b>.54</b>
Axon					
PS Speed	.51	<b>.65</b>	.50	.51	.52
AT Speed	.55	<b>.69</b>	.60	.55	.54
LN Acc.	.56	<b>.65</b>	.60	.52	.59
WM Speed	.53	<b>.64</b>	.55	.48	.51
Mean	.54	<b>.66</b>	<b>.56</b>	<b>.52</b>	<b>.54</b>
ImPACT					
VERM	.50	<b>.71</b>	<b>.62</b>	.56	.55
VISM	.57	<b>.70</b>	.55	.57	.57
VMS	.58	<b>.71</b>	.58	.59	.57
RT	.48	<b>.70</b>	.55	.57	.54
Mean	.53	<b>.71</b>	<b>.58</b>	<b>.57</b>	<b>.56</b>

Note. Bolded where  $p < .05$  after adjustment for multiple comparisons. BL = baseline; SRT = Simple reaction time; CDS = code substitution-learning; PRO = procedural reaction time; MTH = mathematical processing; M2S = matching to sample; CDD = code substitution-delayed; SR2 = simple reaction time 2; GNG = go no-go; PS = processing speed; AT = attention; LN acc. = learning accuracy; WM = working memory; VERM = verbal memory composite; VISM = visual memory composite; VMS = visual motor speed composite; RT = reaction time composite.

between groups (*ds* = .39 to .47) at day 8, and only the ANAM Matching to Sample was significant at day 15 (*d* = .40).

### Receiver Operating Characteristic Curves of CNT Subscales

Table 4 displays the AUC values from the ROC curve for the SCAT3 symptom severity score and each CNT index. Across the three CNTs, all AUC values within 24 hr of injury were in the poor ( $\leq .69$ ) to fair (.70–.73) range. AUCs at day 8 were all in the poor range. The SCAT3 symptom score demonstrated good (AUC = .87; 95% CI = .82–.91) discrimination within 24 hr, with discrimination falling to chance levels at day 8 (AUC = .53; 95% CI = .47–.60).

### Joint Rates of Impairment Across All RCIs for Each CNT

Table 5 displays the percentage of all concussed (All), symptom-free concussed (Sx-), and control subjects who were classified as impaired on 1 or more (1+) and 2 or more



**Table 5.** Percentage of concussed (All), asymptomatic concussed (Sx-), and non-injured controls with 1 or more (1+) and 2 or more (2+) significant declines according to reliable change index (RCI) criteria

	24 hr			Day 8			Day 15			Day 45		
	Concussed		Control	Concussed		Control	Concussed		Control	Concussed		Control
	All	Sx-	(false +)	All	Sx-	(false +)	All	Sx-	(false +)	All	Sx-	(false +)
<b>ANAM</b>												
1+ decline	47.6	42.9	25.0	31.0	28.6	30.3	35.4	30.8	27.9	25.7	26.1	25.3
2+ decline	31.0	0.0	6.3	10.7	7.1	14.7	7.6	6.2	11.5	11.4	11.6	8.0
<b>Axon</b>												
1+ decline	60.3	50.0	22.0	38.9	32.4	26.7	26.5	23.5	22.9	26.3	28.6	20.8
2+ decline	34.2	0.0	4.4	12.5	8.1	7.0	7.4	7.8	7.2	14.0	14.3	5.6
<b>ImPACT</b>												
1+ decline	67.8	53.8	29.6	48.8	39.3	42.7	47.6	46.2	40.8	39.7	39.0	40.0
2+ decline	34.5	15.4	4.0	8.3	4.9	6.5	10.7	9.0	10.8	6.4	6.5	4.8

Note. Symptom-free (Sx-) *ns* at 24 hr were small (7 for ANAM, 8 for Axon, and 13 for ImPACT). Symptomatic *ns* were small at day 45 (2 for ANAM; 1 for Axon/ImPACT). The number of neurocognitive RCIs available for each CNT was 7 for ANAM, 5 for Axon, and 4 for ImPACT. ImPACT uses 80% confidence intervals around RCIs, whereas ANAM and Axon use 90% CIs.

(2+) RCIs. Symptom-free was classified according to athletes' self-report of recovery of any postconcussive symptoms during the recovery interview. As expected, the sensitivity of each CNT to concussion (All) was highest within 24 hr of injury (47.6% ANAM, 60.3% Axon, and 67.8% ImPACT with one or more significant RCIs) and lower for day 8 and beyond (25.7–35.4% for ANAM; 26.3–38.9% for Axon, and 39.7%–48.8% for ImPACT). The false positive rate (percentage of controls with 1+ impaired RCIs) across all time points ranged from 25.0–30.3% for ANAM, 20.8–26.7% for Axon, and 29.6–42.7% for ImPACT. At 24 hr, the sensitivity for symptom-free concussed athletes was similar to that of the entire concussed sample for ANAM (42.9%) and was somewhat lower for Axon (50.0%) and ImPACT (53.8%). Sensitivities in symptom-free athletes at 8 days and beyond were comparable to the false positive rates although, as we address below (see Table 6 and second to last section of the Results), this could have been due to the fact that many athletes tested at these later time points had been asymptomatic for several days. Finally, as expected, both sensitivity values and false positive rates decreased when examining only athletes with 2 or more significant RCIs (e.g., ANAM sensitivity/false positive rate at 24 hr: 31.0/6.3; Axon: 34.2/4.4, and ImPACT 34.5/4.0).

### Sensitivity and Specificity of RCIs by Subtest

Although the joint rates of impairment across each test's set of RCIs is most relevant to clinical decision making, it may also be useful to examine the performance of RCIs for individual subtests within each CNT to determine the subtests with the best (and worst) discrimination between concussed and control athletes. Supplementary Table S8 displays the percentage of all concussed (All), symptom-free concussed (Sx-), and control subjects who were classified as impaired on each RCI within each test battery.

Sensitivity to concussion (All) at 24 hr ranged from 6.0–23.8% for ANAM's seven subtests, 6.8–48.6% for Axon's four subtests, and 24.4–39.5% for ImPACT's four clinical composite scales (*M* difference between the hit and false positive rate for ANAM, Axon, and ImPACT was 13.4%, 21.0%, and 23.2%, respectively). Sensitivity to concussion (All) diminished substantially at day 8 and beyond (*M* difference between the hit and false positive rate at day 8 for ANAM, Axon, and ImPACT = 0.4%, 4.9%, and 2.4%, respectively). Sensitivity for most tests generally also diminished when considering only symptom-free athletes, with the *M* difference at 24 hr between the hit and false positive rate for ANAM, Axon, and ImPACT = 1.5%, 3.4%, and 5.2%, respectively (*M* sensitivity for asymptomatic athletes at day 8 was lower than the false positive rate for ANAM and ImPACT and only 1.1% higher than the false positive rate for Axon).

### Sensitivity of RCIs in Recently Asymptomatic Athletes

As the study design involved fixed assessment time points, the prior analysis of athletes who were symptom-free at each assessment point may not have optimal ecological validity. This is because in many concussion management programs, sports medicine professionals are likely to test their athletes soon after they report becoming symptom-free, and many athletes who were identified as asymptomatic at days 8, 15, and 45 had become asymptomatic several days before these assessment points. To the degree that neurocognitive impairment diminishes rapidly over the course of several days, aggregating athletes who became symptom free recently *versus* more remotely (as was the case in the day 8 and later time points for Table 5) could underestimate the frequency of neurocognitive impairment at the time when

**Table 6.** Percentage of concussed athletes with self-reported symptom resolution within 1 day of testing classified as impaired Per RCI criteria as compared to non-injured controls

	Concussed		Net gain (hits – FP)
	Sx- within 1 day of assessment	Control	
ANAM			
1+ decline	44.4	27.9	16.5
2+ decline	22.2	11.0	11.2
Axon			
1+ decline	52.6	24.4	28.2
2+ decline	5.3	6.0	-0.7
ImPACT			
1+ decline	56.3	37.2	19.1
2+ decline	21.9	5.9	16.0

Note. FP = False positives. Asymptomatic concussed group aggregates all follow-up time points, selecting any subject who self-reported symptom resolution within 1 day of any follow-up exam. Control data represent a weighted average of the false positive rates observed at each time point, weighted to match the percentage of 24-hr, day 8, and day 15 time points used in the concussed athlete column. “1+ decline” (and “2+ decline”) indicate the percentage of subjects with 1 or more (and 2 or more) significant declines from baseline across each test’s set of RCIs.

many athletes would be likely to first take a CNT. To determine whether this was the case, we defined a group of concussed athletes who, based on their self-reported symptom duration in a recovery interview, reported having become asymptomatic within 1 day of any follow-up examination. Table 6 provides the sensitivity of each CNT to concussion for this subset of recently asymptomatic athletes (across 24-hr, day 8, and day 15 assessments; no athletes fell into this category at the day 45 assessment). False positive rates observed in the non-injured controls at the 24-hr, day 8, and day 15 time points were weighted to match the proportion concussed data pulled from each assessment. Consistent with expectation, sensitivity values were generally higher using this approach, with the sensitivity (1 or more decline) of ANAM = 44.4%, Axon = 52.6%, and ImPACT = 56.3% (the false positive rates were 27.9%, 24.4%, and 37.2%, respectively, yielding *M* differences between hit and false positive rates = 16.5%, 28.2%, and 19.1%).

### Positive and Negative Predictive Value of CNT RCIs

Finally, positive predictive value (PPV) and negative predictive value (NPV) was computed to illustrate the relationship between the sensitivity, specificity, and clinical utility of the CNTs’ RCI profiles over time. Given that symptom reporting is the gold standard metric of clinical impairment for SRC, base rates reflect the percentage of concussed athletes reporting symptom impairment at each time point. Accordingly, in the interest of establishing the degree to which the CNT’s correctly classify concussed athletes into symptomatic *versus*

**Table 7.** Positive (PPV) and negative (NPV) predictive values of CNT RCIs profiles by time (%)

	24 hr		Day 8		Day 15	
	PPV	NPV	PPV	NPV	PPV	NPV
ANAM						
1+ decline	90.0	11.1	43.8	68.4	23.0	89.5
2+ decline	100.0	14.9	12.6	37.9	7.9	75.6
Axon						
1+ decline	91.8	14.9	44.4	70.3	24.1	88.6
2+ decline	100.0	16.9	18.0	31.7	21.3	85.8
ImPACT						
1+ decline	91.7	15.6	50.7	80.9	20.1	90.3
2+ decline	95.4	13.9	16.3	53.0	39.1	88.7
Base rate of impairment	89.4		35.4		14.8	

Note. Base rate = percentage of concussed athletes reporting being symptomatic at each assessment point. Given the outcome of interest involved predicting who from the concussed group was impaired from a symptom standpoint, sensitivities and specificity values were extracted from the symptomatic and symptom-free concussed athletes, respectively (which did not allow for computation at day 45 given the small sample of symptomatic subjects at this time point). 1+ (and 2+) decline reflects profiles with 1 or more (and 2 or more) RCIs demonstrated significantly worse performance as compared to an athletes’ pre-injury baseline.

asymptomatic categories, sensitivity was extracted from symptomatic concussed athletes, and specificity from asymptomatic (“recovered”) concussed athletes for these computations (this did not allow for computation of PPV/NPV at day 45, given that only 2 athletes remained symptomatic at this time point). Although multiple approaches to selecting base rates could have been implemented, this approach was targeted to provide an illustration of the relationship between test psychometrics and clinical utility using a clinically relevant anchor of recovery. Table 7 depicts the resultant PPV/NPV values. Given the high base rate of symptom impairment at 24 hr, it is not surprising that PPV was uniformly high at this assessment point (>90% across all CNTs and thresholds for impairment). NPV, however, was low at this time point (<17% across all CNTs). At day 8, PPV was lower and only over 50% for one metric: ImPACT using a threshold for impairment requiring 1 or more significant RCIs. NPV at day 8 was relatively high (>68%) across all CNTs using this 1+ impairment criteria.

### DISCUSSION

In this large-scale, prospective study of the utility of three CNTs for the assessment of SRC, we found that ANAM, Axon, and ImPACT manifested variable and generally modest test–retest reliability and moderate group-level sensitivity soon (<24 hr) after SRC. At 8 days post-injury and beyond, concussed *versus* control group effect sizes were generally small. The test–retest reliability values reported are consistent with a recent review of this topic (Resch, McCrea, et al., 2013) and were generally lower than is considered needed to contribute meaningfully to clinical decisions.

In particular, only approximately a quarter of indices from each CNT had stability coefficients over  $r = .70$ . Similarly, although concussed *versus* control group differences for each CNT were moderate to large within 24 hr of injury according to convention ( $M$  Cohen's  $d$  for ANAM, Axon, and ImpACT =  $-.60$ ,  $-.57$ , and  $-.76$ , respectively), these effect sizes translated to fair to poor discrimination between groups, even at this early post-injury time point ( $M$  AUC for ANAM, Axon, and ImpACT =  $.65$ ,  $.66$ , and  $.71$ , respectively). In contrast, effect sizes for SCAT3 symptom checklist were large within 24 hr ( $d = 1.53$ ) and manifested good discrimination between groups at this time point (AUC =  $.87$ ).

Analyses of the sensitivity and specificity of the CNT's reliable change index output told a similar story, with sensitivities best within 24 hr of injury (47.6%, 60.3%, and 67.8% for ANAM, Axon, and ImpACT, respectively) and diminished substantially to at or near the false positive rate observed in non-injured controls for each measure by the day 8 assessment and beyond. The overall sensitivity rate for ImpACT within 24 hr of injury (67.8% of all concussed athletes showed declines on one or more neurocognitive RCIs) was consistent with the lower bound of previously reported rates (Broglia, Macciocchi, et al., 2007) and lower than some other published estimates (Iverson et al., 2006, 2003; Van Kampen et al., 2006). Although prior data on ANAM's performance in the context of SRC is limited, our overall sensitivity rate was consistent with that of one prior report (with false positive rates in our sample somewhat higher; Register-Mihalik, Guskiewicz, et al., 2012). Our sample yielded lower sensitivity but higher specificity than a previously published study of Axon (Louey et al., 2014).

Our findings of modest reliability and validity may be explained by several factors. First, the clinical manifestations of SRC are most prominent immediately after injury and demonstrate rapid recovery even within the first hours post-injury at a group level (McCrea et al., 2003). Indeed, our findings are consistent with prior meta-analyses of the magnitude of neurocognitive changes after SRC (Belanger & Vanderploeg, 2005; Broglia & Puetz, 2008) and with what is known about the rapid clinical recovery course after concussion (for a review, see Nelson, Janeczek, & McCrea, 2013). An alternative viewpoint is that impairments persist further out from injury but that these CNTs simply lack the sensitivity to detect the abnormal signal. That the cognitive domains most affected by SRC (e.g., processing speed, attention) may be more sensitive than others (e.g., "hold" measures) to state factors (e.g., effort, motivation, fatigue) could limit the stability of measures of these constructs and, by extension, magnify difficulties detecting what become very subtle impairments within hours after injury. It is also possible that testing conditions (e.g., group size at baseline examinations) could have increased variability in performance at this time point and affected results pertaining to the baseline data, although limited recording of group size precluded formal analysis of this (Moser et al., 2011).

An important contribution of this paper was its emphasis on presenting joint base rates of impairment for both

concussed and control athletes. Much prior work on the performance of these CNTs has emphasized the sensitivity of individual subtests or the sensitivity of sets of indices for concussed athletes alone. However, given that clinicians using these multi-index batteries are faced with interpreting the results of sets of indices simultaneously, it is critical to know the joint base rates of impairment in healthy controls (i.e., false positives) to fairly judge the utility of the tests and to identify optimal decision rules for classifying individuals as impaired. Although the false positive rates of individual RCIs can be predicted from their confidence levels (e.g., 10% using an 80% CI; 5% using a 90% CI), as with any set of neuropsychological tests, the base rates of impairment across multiple tests may be much higher depending on the number of indices being jointly interpreted and their intercorrelations (Crawford, Garthwaite, & Gault, 2007; Nelson, in press; Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008).

Consistent with this, the false positive rates in our sample (using 1 or more significant RCIs as the threshold for impairment) ranged (across time points) from 25.0–30.3% ( $M = 27.1%$ ) for ANAM, 20.8–26.7% ( $M = 23.1%$ ) for Axon, and 29.6–42.7% ( $M = 38.3%$ ) for ImpACT. False positive rates were significantly reduced when considering controls with 2 or more significant RCIs ( $M$  false positive rates for ANAM, Axon, and ImpACT using this criterion = 10.1, 6.1, and 6.5%, respectively). These data can serve as important reference points for clinicians who are faced with determining the best impairment criteria given how they weigh different decision making errors.

The current study findings highlight the psychometric limitations of neurocognitive tests for SRC assessment at a group level, yet it has been suggested that such analyses obscure the contribution of neurocognitive testing for the minority of individuals who appear to show more prolonged clinical recovery (Iverson et al., 2006). In support of this idea, our data suggest that CNTs may be more sensitive than athletes' subjective symptom ratings for a short window of time post-symptom resolution and therefore could alter clinical return-to-play decision making for some concussed athletes. However, because of the relatively high false positive rates in the CNTs, the added value of these neurocognitive measures appears rather modest even for individual-level analyses. A limitation of these analyses is that, because the primary aim of this study was to compare the properties and performance of these three CNTs in a common sample, we used fixed assessment time points that were not overtly tied to symptom recovery. This resulted in diminished  $n$ s available for supplementary analyses of symptom-free athletes at some time points and underscores the importance of replicating these results in other samples. Future studies using floating study designs that explicitly perform CNT testing after athletes become asymptomatic would be valuable to garner more power to evaluate the performance of these tests in this clinically-relevant subgroup of athletes.

Even if neurocognitive deficits persist after symptom resolution for some athletes, it is not known to what extent delaying their return-to-play due to these findings would

modify their short-term risk of re-injury, underlying neural recovery, or longer-term prognosis. In fact, a recent randomized controlled trial found that extended strict rest (5 days) resulted in longer symptom recovery (and equivalent neurocognitive and balance recovery) as compared to a shorter period of rest (1–2 days) followed by a graduated return to normal activity (Thomas, Apps, Hoffmann, McCrea, & Hammeke, 2015). It is also not known to what extent clinical recovery intersects with that of underlying neural systems, as a growing neuroimaging literature is finding neurophysiological deficits that persist after the point of clinical recovery (Broglia, Pontifex, O'Connor, & Hillman, 2009; Dettwiler et al., 2014; Pritchep, McCrea, Barr, Powell, & Chabot, 2013; Zhu et al., 2015). It will be important for future research to elucidate the mechanisms underlying these effects, establish which athletes are at greatest risk for extended neurocognitive and neurophysiologic recovery, and establish to what degree changes in clinical decisions mediate individuals' immediate recovery and long-term outcomes.

Further complicating this research is that there is no universally agreed upon way to define concussion and, consequently, its diagnosis relies on athletes' subjective reporting of nonspecific signs and symptoms. This likely leads to research samples being comprised of individuals with heterogeneous injuries that could unknowingly diminish the effects of neurocognitive and other clinical measures. Emerging research is beginning to identify neurophysiologic markers of concussion with the hope of developing more objective definitions of injury (Mondello et al., 2014; Yuh, Hawryluk, & Manley, 2014). To the extent that the construct of concussion becomes better operationalized, our ability to study under what conditions neuropsychological testing contributes meaningful clinical information will improve. However, even with more objective ways to identify concussion, individual athletes will vary in their propensity to develop clinical symptoms of injury and in their recovery courses.

Overall, our findings suggest that the clinical utility of CNTs in the context of SRC management is maximal very soon (within 24 hr) after injury or after symptom resolution and quite limited at later time points (day 8 and beyond). These findings are consistent with current consensus within the broader community that, although neurocognitive tests *can* contribute to the overall clinical picture, they should not be considered in isolation or favored over multidimensional clinical assessment approaches. Future research that improve the objective diagnosis of concussion and that illuminates the interplay between the individual risk factors, patterns of clinical recovery, and interactions with underlying neurophysiological processes will inform best practice in the use of neurocognitive testing in concussion management programs.

## ACKNOWLEDGMENTS

This work was supported by the U.S. Army Medical Research and Materiel Command under award number W81XWH-12-1-0004. Opinions, interpretations, conclusions and recommendations are

those of the author and are not necessarily endorsed by the U.S. Army. The REDCap electronic database service used for the study was supported by the Clinical and Translational Science Institute grant IUL1-RR031973 (-01) and by the National Center for Advancing Translational Sciences, National Institutes of Health grant 8UL1TR000055. The manuscript's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The authors have no conflicts of interest to report.

## Supplementary Materials

Supplementary materials can be found online. Please visit [journals.cambridge.org/jid\\_INS](http://journals.cambridge.org/jid_INS).

## REFERENCES

- Belanger, H.G., & Vanderploeg, R.D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*, 345–357.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 57*, 289–300.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics, 29*, 1165–1188.
- Bleiberg, J., Cernich, A.N., Cameron, K., Sun, W., Peck, K., Ecklund, P.J., ... Warden, D.L. (2004). Duration of cognitive impairment after sports concussion. *Neurosurgery, 54*, 1073–1078.
- Broglia, S.P., Ferrara, M.S., Macciocchi, S.N., Baumgartner, T.A., & Elliott, R. (2007). Test-retest reliability of computerized concussion assessment programs. *Journal of Athletic Training, 42*, 509–514.
- Broglia, S.P., Macciocchi, S.N., & Ferrara, M.S. (2007). Sensitivity of the concussion assessment battery. *Neurosurgery, 60*, 1050–1057. doi:10.1227/01.NEU.0000255479.90999.C0
- Broglia, S.P., Pontifex, M.B., O'Connor, P., & Hillman, C.H. (2009). The persistent effects of concussion on neuroelectric indices of attention. *Journal of Neurotrauma, 26*, 1463–1470. doi:10.1089/neu.2008-0766
- Broglia, S.P., & Puetz, T.W. (2008). The effect of sport concussion on neurocognitive function, self-report symptoms and postural control: A meta-analysis. *Sports Medicine, 38*, 53–67.
- 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*, 14–25. doi:10.1080/13854046.2013.866272
- Cernich, A., Reeves, D., Sun, W., & Bleiberg, J. (2007). Automated Neuropsychological Assessment Metrics sports medicine battery. *Archives of Clinical Neuropsychology, 22*(Suppl. 1), S101–S114. doi:10.1016/j.acn.2006.10.008
- Chelune, G.J., Nagle, 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*, 732–742. doi:10.1093/arclin/act040

- Collie, A., Darby, D., & Maruff, P. (2001). Computerised cognitive assessment of athletes with sports related head injury. *British Journal of Sports Medicine, 35*, 297–302.
- 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*, 28–32.
- Covassin, T., Elbin, R., III, & Stiller-Ostrowski, J.L. (2009). Current sport-related concussion teaching and clinical practices of sports medicine professionals. *Journal of Athletic Training, 44*, 400–404. doi:10.4085/1062-6050-44.4.400
- Covassin, T., Elbin, R.J., III, Stiller-Ostrowski, J.L., & Kontos, A.P. (2009). Immediate post-concussion assessment and cognitive testing (ImPACT) practices of sports medicine professionals. *Journal of Athletic Training, 44*, 639–644. doi:10.4085/1062-6050-44.6.639
- Crawford, J.R., Garthwaite, P.H., & Gault, C.B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology, 21*, 419–430. doi:10.1037/0894-4105.21.4.419
- Derogatis, L.R. (2001). *Brief Symptom Inventory 18 (BSI-18): Administration, scoring, and procedures manual*. Bloomington, MN: Pearson.
- Dettwiler, A., Murugavel, M., Putukian, M., Cubon, V., Furtado, J., & Osherson, D. (2014). Persistent differences in patterns of brain activation after sports-related concussion: A longitudinal functional magnetic resonance imaging study. *Journal of Neurotrauma, 31*, 180–188. doi:10.1089/neu.2013.2983
- Diener, E., Emmons, R.A., Larsen, R.J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71–75. doi:10.1207/s15327752jpa4901\_13
- Echemendia, R.J., Iverson, G.L., McCrea, M., Macciocchi, S.N., Gioia, G.A., Putukian, M., & Comper, P. (2013). Advances in neuropsychological assessment of sport-related concussion. *British Journal of Sports Medicine, 47*, 294–298. doi:10.1136/bjsports-2013-092186
- 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*, 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. *American Journal of Sports Medicine, 39*, 2319–2324. doi:10.1177/0363546511417173
- Green, P. (2003). *Green's Medical Symptom Validity Test for Windows*. Edmonton, Alberta, Canada: Green's Publishing, Inc.
- Guskiewicz, K.M., Ross, S.E., & Marshall, S.W. (2001). Postural stability and neuropsychological deficits after concussion in collegiate athletes. *Journal of Athletic Training, 36*, 263–273.
- Helmick, K., Guskiewicz, K., Barth, J., Cantu, R., Kelly, J., McDonald, E., ... Warden, D. (2006). *Defense and Veterans Brain Injury Center Working Group on the acute management of mild traumatic brain injury in military operational settings: Clinical practice guideline and recommendations*. Washington, DC: Defense and Veteran Brain Injury Center. Retrieved from [http://www.pdhealth.mil/downloads/clinical\\_practice\\_guideline\\_recommendations.pdf](http://www.pdhealth.mil/downloads/clinical_practice_guideline_recommendations.pdf)
- Iverson, G.L., Brooks, B.L., Collins, M.W., & Lovell, M.R. (2006). Tracking neuropsychological recovery following concussion in sport. *Brain Injury, 20*, 245–252. doi:10.1080/02699050500487910
- Iverson, G.L., Lovell, M.R., & Collins, M.W. (2003). Interpreting change on ImPACT following sport concussion. *Clinical Neuropsychologist, 17*, 460–467. doi:10.1076/clin.17.4.460.27934
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
- LaRoche, A.A., Nelson, L.D., Connelly, P.K., Walter, K.D., & McCrea, M.A. (2015). Sport-related concussion reporting and state legislative effects. *Clinical Journal of Sport Medicine*. doi:10.1097/JSM.0000000000000192
- Lichtenstein, J.D., Moser, R.S., & Schatz, P. (2014). Age and test setting affect the prevalence of invalid baseline scores on neurocognitive tests. *American Journal of Sports Medicine, 42*, 479–484. doi:10.1177/0363546513509225
- Louey, A.G., Cromer, J.A., Schembri, A.J., Darby, D.G., Maruff, P., Makdissi, M., & McCrory, P. (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. doi:10.1093/arclin/acu020
- 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*, 367–372. doi:10.1097/JSM.0000000000000139
- McClincy, M.P., Lovell, M.R., Pardini, J., Collins, M.W., & Spore, M.K. (2006). Recovery from sports concussion in high school and collegiate athletes. *Brain Injury, 20*, 33–39. doi:10.1080/02699050500309817
- McCrea, M., Guskiewicz, K.M., Marshall, S.W., Barr, W., Randolph, C., Cantu, R.C., & Kelly, J.P. (2003). Acute effects and recovery time following concussion in collegiate football players: The NCAA Concussion Study. *Journal of the American Medical Association, 290*, 2556–2563. doi:10.1001/jama.290.19.2556
- McCrea, M., Kelly, J.P., Randolph, C., Kluge, J., Bartolic, E., Finn, G., ... Baxter, B. (1998). Standardized assessment of concussion (SAC): On-site mental status evaluation of the athlete. *Journal of Head Trauma Rehabilitation, 13*, 27–35.
- McCrory, P., Meeuwisse, W., Aubry, M., Cantu, B., Dvorak, J., Echemendia, R.J., ... Tator, C.H. (2013). Consensus statement on concussion in sport—the 4th International Conference on Concussion in Sport held in Zurich, November 2012. *Clinical Journal of Sport Medicine, 23*, 89–117. doi:10.1097/JSM.0b013e31828b67cf
- Meehan, W.P., III, d'Hemecourt, P., Collins, C.L., Taylor, A.M., & Comstock, R.D. (2012). Computerized neurocognitive testing for the management of sport-related concussions. *Pediatrics, 129*, 38–44. doi:10.1542/peds.2011-1972
- Mondello, S., Schmid, K., Berger, R., Kobeissy, F., Italiono, D., Jeromin, A., ... Buki, A. (2014). The challenge of mild traumatic brain injury: Role of biochemical markers in diagnosis of brain damage. *Medical Research Reviews, 34*, 503–531.
- Moser, R.S., Iverson, G.L., Echemendia, R.J., Lovell, M.R., Schatz, P., Webbe, F.M., ... Barth, J.T. (2007). Neuropsychological evaluation in the diagnosis and management of sports-related concussion. *Archives of Clinical Neuropsychology, 22*, 909–916. doi:10.1016/j.acn.2007.09.004
- Moser, R.S., Schatz, P., & Lichtenstein, J.D. (2015). The importance of proper administration and interpretation of neuropsychological baseline and postconcussion computerized testing. *Applied Neuropsychology, Child, 4*, 41–48. doi:10.1080/21622965.2013.791825

- Moser, R.S., Schatz, P., Neidzowski, K., & Ott, S.D. (2011). Group versus individual administration affects baseline neurocognitive test performance. *American Journal of Sports Medicine*, *39*, 2325–2330. doi:10.1177/0363546511417114
- Nakayama, Y., Covassin, T., Schatz, P., Nogle, S., & Kovan, J. (2014). Examination of the test-retest reliability of a computerized neurocognitive test battery. *American Journal of Sports Medicine*, *42*, 2000–2005. doi:10.1177/0363546514535901
- Nelson, L.D. (in press). False positive rates of reliable change indices for concussion test batteries: A Monte Carlo simulation. *Journal of Athletic Training*.
- Nelson, L.D., Janecek, J.K., & McCrea, M.A. (2013). Acute clinical recovery from sport-related concussion. *Neuropsychology Review*, *23*, 285–299. doi:10.1007/s11065-013-9240-7
- Nelson, L.D., Pfaller, A.Y., Rein, L., & McCrea, M.A. (2015). Rates and predictors of invalid baseline test performance for three computerized neurocognitive tests (CNTs): ANAM, Axon, and ImPACT. *American Journal of Sports Medicine*, *43*, 2018–2026. doi:10.1177/0363546515587714
- Pritchep, L.S., McCrea, M., Barr, W., Powell, M., & Chabot, R.J. (2013). Time course of clinical and electrophysiological recovery after sport-related concussion. *Journal of Head Trauma Rehabilitation*, *28*, 266–273. doi:10.1097/HTR.0b013e318247b54e
- Rahman-Filipiak, A.A.M., & Woodward, J.L. (2014). Administration and environment considerations in computer-based sports-concussion assessment. *Neuropsychology Review*, *23*, 314–334.
- Randolph, C., McCrea, M., & Barr, W.B. (2005). Is neuropsychological testing useful in the management of sport-related concussion? *Journal of Athletic Training*, *40*, 139–152.
- Register-Mihalik, J.K., Guskiewicz, K.M., Mihalik, J.P., Schmidt, J.D., Kerr, Z.Y., & McCrea, M.A. (2012). Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: Implications for caution in clinical practice. *Journal of Head Trauma Rehabilitation*. doi:10.1097/HTR.0b013e3182585d37
- Register-Mihalik, J.K., Guskiewicz, K.M., Mihalik, J.P., Schmidt, J.D., Kerr, Z.Y., & McCrea, M.A. (2013). Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: Implications for caution in clinical practice. *Journal of Head Trauma Rehabilitation*, *28*, 274–283. doi:10.1097/HTR.0b013e3182585d37
- 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*, 297–305. doi:10.4085/1062-6050-47.3.13
- Resch, J.E., Driscoll, A., McCaffrey, N., Brown, C., Ferrara, M.S., Macciocchi, S., ... Walpert, K. (2013). ImPact test-retest reliability: Reliably unreliable? *Journal of Athletic Training*, *48*, 506–511. doi:10.4085/1062-6050-48.3.09
- Resch, J.E., McCrea, M.A., & Cullum, C.M. (2013). Computerized neurocognitive testing in the management of sport-related concussion: An update. *Neuropsychology Review*, *23*, 335–349. doi:10.1007/s11065-013-9242-5
- Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intratester, interrater and test-retest reliability of continuous measurements. *Statistics in Medicine*, *21*, 3431–3446. doi:10.1002/sim.1253
- Schatz, P. (2010). Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *American Journal of Sports Medicine*, *38*, 47–53. doi:10.1177/0363546509343805
- Schatz, P., & Ferris, C.S. (2013). One-month test-retest reliability of the ImPACT test battery. *Archives of Clinical Neuropsychology*, *28*, 499–504. doi:10.1093/arclin/act034
- Schatz, P., Pardini, J.E., Lovell, M.R., Collins, M.W., & Podell, K. (2006). Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. *Archives of Clinical Neuropsychology*, *21*, 91–99. doi:10.1016/j.acn.2005.08.001
- Schatz, P., & Sandel, N. (2013). Sensitivity and specificity of the online version of ImPACT in high school and collegiate athletes. *American Journal of Sports Medicine*, *41*, 321–326. doi:10.1177/0363546512466038
- Schretlen, D.J., Testa, S.M., Winicki, J.M., Pearlson, G.D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society*, *14*, 436–445. doi:10.1017/S1355617708080387
- Segalowitz, S.J., Mahaney, P., Santesso, D.L., MacGregor, L., Dywan, J., & Willer, B. (2007). Retest reliability in adolescents of a computerized neuropsychological battery used to assess recovery from concussion. *NeuroRehabilitation*, *22*, 243–251.
- Sim, A., Terryberry-Spohr, L., & Wilson, K.R. (2008). Prolonged recovery of memory functioning after mild traumatic brain injury in adolescent athletes. *Journal of Neurosurgery*, *108*, 511–516. doi:10.3171/JNS/2008/108/3/0511
- Straume-Naesheim, T.M., Andersen, T.E., & Bahr, R. (2005). Reproducibility of computer based neuropsychological testing among Norwegian elite football players. *British Journal of Sports Medicine*, *39*(Suppl. 1), i64–i69. doi:10.1136/bjism.2005.019620
- Thomas, D.G., Apps, J.N., Hoffmann, R.G., McCrea, M., & Hammeke, T. (2015). Benefits of strict rest after acute concussion: A randomized controlled trial. *Pediatrics*, *135*, 213–223. doi:10.1542/peds.2014-0966
- Van Kampen, D.A., Lovell, M.R., Pardini, J.E., Collins, M.W., & Fu, F.H. (2006). The “value added” of neurocognitive testing after sports-related concussion. *American Journal of Sports Medicine*, *34*, 1630–1635. doi:10.1177/0363546506288677
- Vincent, A.S., Roebuck-Spencer, T., Lopez, M.S., Twillie, D.A., Logan, B.W., Grate, S.J., ... Gilliland, K. (2012). Effects of military deployment on cognitive functioning. *Military Medicine*, *177*, 248–255.
- Wechsler, D. (2001). *Wechsler Test of Adult Reading: WTAR*. San Antonio, TX: The Psychological Corporation.
- Weir, J.P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength & Conditioning Research*, *19*, 231–240. doi:10.1519/15184.1
- Yuh, E.L., Hawryluk, G.W., & Manley, G.T. (2014). Imaging concussion: A review. *Neurosurgery*, *75*(Suppl. 4), S50–S63. doi:10.1227/NEU.0000000000000491
- Zhu, D.C., Covassin, T., Nogle, S., Doyle, S., Russell, D., Pearson, R.L., ... Kaufman, D.I. (2015). A potential biomarker in sports-related concussion: Brain functional connectivity alteration of the default-mode network measured with longitudinal resting-state fMRI over thirty days. *Journal of Neurotrauma*, *32*, 327–341. doi:10.1089/neu.2014.3413