REVIEW

Computerized Neurocognitive Testing in the Management of Sport-Related Concussion: An Update

Jacob E. Resch · Michael A. McCrea · C. Munro Cullum

Abstract Since the late nineties, computerized neurocognitive testing has become a central component of sport-related concussion (SRC) management at all levels of sport. In 2005, a review of the available evidence on the psychometric properties of four computerized neuropsychological test batteries concluded that the tests did not possess the necessary criteria to warrant clinical application. Since the publication of that review, several more computerized neurocognitive tests have entered the market place. The purpose of this review is to summarize the body of published studies on psychometric properties and clinical utility of computerized neurocognitive tests available for use in the assessment of SRC. A review of the literature from 2005 to 2013 was conducted to gather evidence of test-retest reliability and clinical validity of these instruments. Reviewed articles included both prospective and retrospective studies of primarily sport-based adult and pediatric samples. Summaries are provided regarding the available evidence of reliability and validity for the most commonly used computerized neurocognitive tests in sports settings.

J. E. Resch (✉)
Brain Injury Laboratory, Department of Kinesiology, The University of Texas at Arlington, Box 19259, 113 Maverick Activities Center, Arlington, TX 76019-0259, USA
e-mail: resch@uta.edu

M. A. McCrea
Departments of Neurosurgery and Neurology, Medical College of Wisconsin, Neuroscience Research Center, 8701 W. Watertown Plank Road, Milwaukee, WI 53266, USA
e-mail: mmccrea@mcw.edu

C. M. Cullum
Department of Psychiatry, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-9044, USA
e-mail: Munro.cullum@utsouthwestern.edu

Introduction

Since the late nineties, computerized neurocognitive testing (CNT) has become a central component of sport-related concussion (SRC) management at all levels of sport (Ferrara et al. 2001; Meehan et al. 2012; Notebaert and Guskiewicz 2005; Randolph 2011). In 2001, the Concussion in Sport Group (CISG) suggested that neuropsychological testing was one of the "cornerstones" of concussion evaluation. The CISG supported the use of brief cognitive screening tools such as the McGill Acute Concussion Evaluation and the Standardized Assessment of Concussion (SAC) (Aubry et al. 2002; McCrory et al. 2005). In addition, emphasis was placed on the utility of conventional neuropsychological tests for the objective assessment of cognitive functioning after concussion. At that time, the CISG also mentioned several relatively new CNTs, including the Immediate Postconcussion and Cognitive Testing test battery (ImPACT), CogSport, Automated Neuropsychological Assessment Metrics (ANAM), and HeadMinder Concussion Resolution Index (CRI) (Aubry et al. 2002).

Over the past decade, CNTs have gained considerable traction as an alternative to traditional paper and pencil-based neuropsychological testing in sports applications due to a number of presumptive advantages: 1) the ability to baseline test groups of athletes concurrently, 2) wide availability via the internet and other electronic platforms, 3) ease of administration, 4) ready access to alternate test forms to reduce practice effects, and 5) creation of centralized data repositories for ready access by users (Guskiewicz et al. 2004; Woodard and Rahman 2012). Albeit the topic of debate, another advantage often cited is that CNTs do not require direct involvement of a neuropsychologist in the administration, scoring and interpretation of cognitive testing, thereby

allowing implementation of these tests more widely (e.g., in geographic areas with no access to neuropsychological services). CNTs have been integrated into youth, secondary, post-secondary and professional sports of all types as well as the military (Cole et al. 2013; Covassin et al. 2009; Ferrara et al. 2001; Meehan et al. 2012; Notebaert and Guskiewicz 2005).

In 2005, Randolph and colleagues published a review of the literature on both traditional neuropsychological tests and CNTs employed in the management of SRC (Randolph et al. 2005). Based upon the evidence available at that time, the authors concluded that CNTs did not yet possess the psychometric properties necessary to warrant their clinical use in the management of SRC. Recommendations for future research and test development were offered. Published guidelines from the CISG and numerous other health care governing bodies, including the National Athletic Trainers' Association, the American Academy of Neurology, and the American Medical Society for Sports Medicine have subsequently recommended that, while neuropsychological testing is an important component of concussion assessment, it should not be the sole basis of SRC management decisions (Giza et al. 2013; Guskiewicz et al. 2004; Harmon et al. 2013; McCrory et al. 2009, 2013).

Collectively, the aforementioned governing bodies and consensus panels suggested that a multi-dimensional approach increases diagnostic accuracy when assessing SRC (Giza et al. 2013; Guskiewicz et al. 2004; Harmon et al. 2013; McCrory et al. 2004, 2009, 2013). In addition to neuropsychological testing, the recommended multi-dimensional approach consists of brief screening measures to assess symptoms, cognitive functioning and balance. The CISG has most recently developed the Sports Concussion Assessment Tool—3rd Edition (SCAT3) as a multi-dimensional assessment appropriate for use in the competitive sporting environment (McCrory et al. 2013). The SCAT3 includes a postconcussion symptom scale, the SAC (McCrea 2001) and the Balance Error Scoring System (BESS) (Guskiewicz et al. 1997; McCrory et al. 2013).

Recent studies have investigated how CNTs influence clinical decision making in the management of SRC. In a survey of athletic trainers that addressed the clinical practice associated with a commonly used CNT, ImPACT (Lovell 2007), the majority (95.5 %) of respondents indicated they would not return an athlete to competition if the athlete was still experiencing postconcussion symptoms, regardless of CNT results (Covassin et al. 2009). A relatively small percentage (4.5 %) of surveyed athletic trainers reported that they would return a concussed athlete to play despite being symptomatic, if the athlete scored within normal limits on the ImPACT. Surveyed athletic trainers were also presented a scenario in which an athlete reported being symptom free following a concussion but scored below pre-injury baseline values on the ImPACT. While 86.5 % of athletic trainers indicated they would not return this athlete to play, 9.8 % of respondents reported they would return the athlete to competition based on

reported symptom recovery, and an additional 3.8 % responded it depended on the importance of the competition (Covassin et al. 2009). In keeping with the evidence based recommendations toward the multi-dimensional assessment approach, the majority of respondents (77.2 % to 100 %) reported using clinical examination, computerized neurocognitive testing (ImPACT), and a symptom checklist to assess SRC. To a lesser extent, respondents incorporated sideline assessment measures, traditional paper and pencil neuropsychological testing, clinical balance measures (e.g., the BESS) and neuroimaging to assess concussed athletes, which is similar to the findings of previous reports (Ferrara et al. 2001; Notebaert and Guskiewicz 2005).

Some of the most commonly employed CNTs now include the ImPACT, CogSport/State (Axon Sport), HeadMinder, and ANAM (Meehan et al. 2012). ImPACT is the primary CNT of many professional sports teams. In a recent study, 39.9 % of athletic trainers reported that CNTs were a part of their SRC management protocol. Of those, the overwhelming majority (93 %) reported using the ImPACT, followed by CogSport/ CogState (2.8 %), unspecified software (2.8 %), and the HeadMinder CRI (1.4 %) (Meehan et al. 2012).

Approximately 86 % to 95.7 % of athletic trainers who use CNTs reported incorporating pre-participation baseline CNT assessments into their SRC management protocol (Covassin et al. 2009; Guskiewicz et al. 2004; Meehan et al. 2012). In terms of CNT interpretation, surveyed athletic trainers indicated that the majority of CNT reports were interpreted by the athletic trainer, a physician, or both, with the minority (~3 % to 17 %) interpreted by a licensed neuropsychologist (Comstock et al. 2012; Covassin et al. 2009; Meehan et al. 2012).

Since 2005, a growing body of research has begun to address the psychometric properties of commonly used CNTs in the management of SRC. Regardless of background, it is imperative that those who employ CNTs for the management of SRC to be aware of their psychometric properties and limitations (Harmon et al. 2013). The purpose of this review is to provide an overview of the published studies on the psychometric properties associated with some of the most commonly used, commercially available CNTs for the management of SRC with particular emphasis on contributions to the literature since the earlier 2005 review by Randolph and colleagues. Our review focused on peer reviewed articles that specifically involved CNTs in the assessment of SRC.

Clinical Use of Computerized Neurocognitive Tests

Since 1996, obtaining a pre-injury *or* baseline neurocognitive assessment has often been recommended for the management of SRC (Aubry et al. 2002; Guskiewicz et al. 2004; Macciocchi et al. 1996; McCrory et al. 2005, 2009, 2013). The goal of a baseline is to allow clinicians to better account

for individual neurocognitive variability for each athlete. The baseline is then used as an additional individual comparison metric for that person in a non-injured state. Importantly, athletes with learning disabilities, attention deficit/hyperactivity disorder and other preexisting cognitive or neuropsychiatric conditions may have greater variability in CNT performance compared with those without such comorbid conditions. Individuals with very high or very low intellectual levels can also be expected to perform differently on CNTs, which can influence post-injury test interpretations. These caveats support the value of the baseline assessment in the interpretation of CNT results following SRC and underscore the relevance of specific pre-existing conditions that may adversely affect CNT performance in some individuals.

Recently, the value and practice of baseline testing has been brought into question (Randolph 2011). Critiques of the baseline assessment model argue that baseline testing has many limitations insofar as it does not modify risk, lacks sufficient psychometric evidence to support clinical utility, is neither time nor cost effective, is influenced by numerous sources of random error (environmental distractions, amount of sleep the night prior to taking the test, caffeine consumption, acute psychological distress, and/or sub-optimal or variable effort), and ultimately may create a false sense of confidence in the clinician who relies on the baseline results in interpreting the clinical significance of post-injury test findings (Hunt et al. 2007; McCrory et al. 2013; Randolph 2011; Schatz and Glatts 2013; Schmidt et al. 2012). Rather than perform baseline CNT assessment, some assert that comparing CNT results to normative data is sufficient. This is common in the practice of clinical neuropsychology (Schmidt et al. 2012). Most SRC-related guidelines cite the potential added value of baseline testing, but do not recommend it as a routine standard of care for SRC and emphasize the need for further research to determine its relative value (Covassin et al. 2010; Notebaert and Guskiewicz 2005). Nevertheless, baseline assessment of athletes continues as common practice at all levels of sport.

Following a diagnosed SRC, the results from a CNT may be interpreted based upon several methods using various statistical approaches and units of measurement. These methods include but are not limited to the simple difference, standard deviation, standard error of measurement, effect sizes (e.g., Cohen's d), reliable change index (RCI) and regression-based methods (Collie et al. 2004; Lovell et al. 2004). Each CNT uses one or more of these approaches to aid the clinician in determining clinically meaningful neurocognitive change following a SRC. Test-retest variability consists of true variance or *normal* fluctuation and variability due to individual human performance factors, as well as inherent measurement error related to the administration of the CNT. Change methods such as the simple difference and standard deviation methods *do not* account for sources of error, while standard error of measurement, Cohen's d, and RCIs do (Collie et al. 2004).

RCIs are the most commonly used method to determine change on CNTs following concussion (Collie et al. 2003b, 2004; Hinton-Bayre 2012; Iverson et al. 2003; Roebuck-Spencer et al. 2007). The calculation of RCIs results in a reference criterion intended to guide clinicians on the clinical meaningfulness of post-injury test results, accounts for various sources of error such as time elapsed since the last test administration or fluctuation in a test taker's performance due to extraneous factors. Several studies support the use of RCIs to interpret CNT reports in the clinical setting (Collie et al.; Hinton-Bayre 2012; Randolph et al. 2005). That said, numerous RCI models exist to differentiate clinically-meaningful change from error variance in a concussed athlete. A recent report comparing four commonly used RCI models to classify concussed rugby players and healthy controls and concluded that no one RCI model is superior to another; rather, clinicians were left to choose one statistical model and be aware of its advantages and disadvantages when classifying concussed athletes (Hinton-Bayre 2012). The authors also asserted that incorrect use or limited understanding of the model used to determine neurocognitive change may lead to erroneous clinical decisions (e.g., false positives, false negatives) which may directly affect the health and safety of the athlete (Hinton-Bayre 2012).

Key Psychometric Properties of Neurocognitive Tests

The development of any neurocognitive test, whether conventional or computerized, requires the sequential establishment of objectivity, feasibility, reliability, and validity evidence. These steps are described in detail in the American Education Research Association Standards for Psychological Testing and are briefly outline herein (Association, A. E. R. 1999).

Objectivity refers to the degree to which multiple examiners or scorers agree on the values of collected metrics or scores. Also known as inter-rater reliability (Baumgartner and Hensley 2006), objectivity is achieved when two independent clinicians administer the same test to the same individual and achieve similar scores. With the advancement of technology, the introduction of novel CNTs to the marketplace, and greater awareness of various sources of test error, objectivity is an increasingly important consideration (Cernich et al. 2007). For example, Cernich et al. (2007) and Woodard and Rahman (2012) suggest that a computer's operating system, hardware (display, keyboard, and mouse) as well as computer programs and Internet connectivity may introduce error to CNTs, thereby decreasing inter-rater reliability (perhaps unknowingly to the user). As such, establishing an environment and mindset geared toward eliciting maximum effort from athletes undergoing CNTs is imperative.

Methods associated with CNT administration may also lead to increased variability in CNT performance. Schatz and colleagues (2010) found that environmental distractions, difficulty with test instructions, and computer problems

resulted in significant increases in symptom reporting and reaction times. Additionally, the number of participants tested at one time may influence CNT performance and effort. For example, athletes completing ImPACT in a group (≥20 student-athletes) setting rather than individually scored significantly lower on all neurocognitive indices and had a greater test invalidity rate (Moser et al. 2011). Such findings raise concern with the claim that CNTs allow for the ability for mass testing, and this is compounded when baseline results are not carefully reviewed for validity only (Covassin et al. 2009). Ultimately, each of these factors contributes to sub-optimal objectivity. Motivation of the test taker is also a factor that can influence objectivity. As is the case with conventional neuropsychological testing, establishing an environment and mindset geared toward eliciting maximum effort from athletes undergoing CNTs (and observing them in the process) is imperative.

Reliability refers to the consistency of test scores (Bauer et al. 2012; Crocker and Algina 2008). In the context of this article, the focus will be on test-retest reliability or *stability*. Adequate test-retest reliability is critical when determining clinically meaningful neurocognitive change following brain injury. Reliability coefficients range from "0" to "1," with values closer to "1" suggesting more stable scores. Determining what is "acceptable" reliability for clinical utility is difficult and often dependent upon the clinical scenario in question. Intraclass correlation values ranging from .60 to .90 have been recommended, with some evidence suggesting .75 as appropriate for clinical decision making (Broglio et al. 2007a; Portney and Watkins 1993; Randolph et al. 2005). Reliability values <.60 are suggested to be reflective of a measure which is still in development rather than one ready for routine clinical use (Nunnally 1978; Pedhazur and Pedhazur Schmelkin 1991).

Validity refers to the extent to which a test measures what it is intended to measure. Validity can also be defined as the degree to which interpretations or inferences of test score derived from an instrument lead to correct conclusions (Baumgartner and Hensley 2006; Crocker and Algina 2008; Schatz and Sandel 2013). *Construct validity* refers to how well a test taps and correlates with an assumed underlying construct. For example, most CNTs consist of several subtests that assess aspects of reaction time, processing speed, and memory. In order for a CNT to possess construct validity, its neurocognitive indices must possess moderate to strong correlations with a concussed athlete's performance on standard clinical measures of the same construct. *Concurrent or convergent* validity is defined as the relationship between specific test scores and a criterion measurement made at the time the test in question was given. For example, an athlete would take a computerized version of a verbal memory test and then be administered a known, or "gold standard," measure of verbal memory (e.g., California Verbal Learning Test, Hopkins Verbal Learning Test, etc.). If the correlation between the computerized test and the "gold standard" is high, then concurrent

validity has been demonstrated (Crocker and Algina 2008). *Discriminant validity* coefficients are correlations between measures of different constructs using the same method or test of measurement. Most CNTs have several neurocognitive indices which purportedly measure different constructs but are determined based on the completion of one subtest. For instance, administration of the ImPACT test battery results in four neurocognitive indices. If the neurocognitive indices possess discriminant validity, some weaker correlations should be observed, suggesting each index taps a different construct (Crocker and Algina 2008). Because of the reliance of CNTs upon similar cognitive skills such as reaction time and working memory, however, it may be difficult to demonstrate discriminant validity of subtests, although it remains important to determine which measures are most sensitive to SRC and recovery.

The most common forms of validity associated with CNTs and SRC reference metrics of sensitivity and specificity. *Sensitivity* has been classically defined as how well a test can differentiate clinical cases (i.e., athletes with concussion or *true positives*) from healthy controls. Tests with high sensitivity are able to detect abnormalities in the affected cases, with a corresponding low occurrence of *false negatives* or incorrectly classifying a concussed athlete as normal. *Specificity* refers to a test's ability to identify healthy controls as healthy or *true negatives*. There are several ways to calculate sensitivity and specificity (Broglio et al. 2007b; Schatz et al. 2006; Schatz and Sandel 2013), but each revolves around the premise in the current context that a test can accurately depict an athlete as cognitively impaired or intact when healthy.

Clinical Utility A clinical test may meet all the aforementioned psychometric criteria but lack clinical utility within a particular setting. Given that athletic trainers and physicians most often administer CNTs in SRC (Covassin et al. 2009; Meehan et al. 2012), it should be kept in mind that a test must ideally be time and cost effective, easy to administer, and ultimately add value to concussion management (Randolph et al. 2005). Of course, a test should be proven sensitive in detecting neurocognitive impairment, particularly in athletes who are reporting to be asymptomatic and seeking clearance for return to participation. Last, interpretation of the CNT should be based on empirically supported data and algorithms that optimize the sensitivity and specificity of the test to SRC, in addition to the clinician's knowledge of the psychometric properties and limitations of the instruments being used.

With respect to the current state of interpretation, surveyed athletic trainers indicated that the majority of CNT reports were interpreted by the athletic trainer, a physician, or both, with the minority (~3 % to 17 %) interpreted by a licensed neuropsychologist (Comstock et al. 2012; Covassin et al. 2009; Meehan et al. 2012). Furthermore, Covassin and colleagues reported that of those athletic trainers (18 %) who reported interpreting the post-concussion CNT report, only 26.4 % had formal

training (e.g., ImPACT Training Workshop) in the interpretation of CNT results (Covassin et al. 2010). As such, the background training and expertise of individuals interpreting CNT results in many settings appears questionable, unlike applications in the NFL and NHL, wherein all teams are required to have a consulting neuropsychologist.

Psychometric Properties of Computerized Neurocognitive Tests

The section below provides a review of the published data available on four CNTs that are commonly used to assess SRC. Unlike the previous review in 2005, both prospective and retrospective studies were considered in the development of this manuscript. The reviewed CNTs included ANAM, CogSport/State, HeadMinder, and ImPACT. Each of the reviewed CNTs is commercially available for use by and marketed to health care professionals for the management of SRC. While we recognize that several other CNTs have recently entered the marketplace, no published studies on the psychometric properties of those instruments were available at the time of our review. Most CNTs assess common cognitive domains, including reaction time, working memory, verbal and visual recognition memory, and various forms of information processing speed. However, each test employs different methods for the assessment of each construct of interest (i.e. different verbal and visual stimuli, reaction time tasks). For comparison, Table 1 describes each CNT and their respective composite scores.

Though the primary focus of this review was on data published since the review by Randolph et al. (2005), a comprehensive review was conducted that included articles published prior to- and after 2005. Literature was retrieved from PubMed, Medline, Google Scholar, and each CNT's respective website. Search criteria consisted of a combination of each specific CNT name and general psychometric terms (e.g. reliability and validity).

For the context of this article, it is important to define the psychometric criteria used to evaluate the published findings from each study. Regarding test-retest reliability, Intraclass correlation (ICC) values of .60 to .90 have been reported as adequate to deem a test clinically useful (Anastasi 1998; Portney and Watkins 1993; Randolph et al. 2005).

Determination of acceptable validity depends on the types of validity evidence available. One of the more common forms of validity reported has been concurrent/convergent validity wherein a CNT is compared to a "gold standard" neuropsychological test using a Pearson correlation. The correlation is considered weak when $r = 0$ to .30, moderate when $r = .31$ to .59, or strong when $r = .60$ to 1 (Dancey and Reidy 2004). As with reliability, determining what is "clinically acceptable" is difficult and based on the amount of error a clinician is willing

to tolerate, the clinical purpose of the test, as well as practical considerations such as time and cost (Pedhazur and Pedhazur Schmelkin 1991; Randolph et al. 2005). Other forms of validity (e.g. discriminant validity, divergent validity, sensitivity and specificity) are more difficult to discern by level of acceptability and are described in each section below.

The following sections present the psychometric evidence with regard to SRC for each reviewed CNT (reviewed in alphabetical order). Published data on the reliability, validity, sensitivity and specificity for each CNT are presented (see Table 2). Summaries of test-retest reliability coefficients for CNT summary scores may be found in Table 3 and for each CNT's respective subtest scores may be found in Table 4.

Automated Neuropsychological Assessment Metrics (ANAM)

Reliability Two published manuscripts were reviewed which addressed the test-retest reliability of ANAM. The first study reported moderate to strong Pearson r coefficients (.50 to .81) and intraclass correlation (ICC) values (.44 to .72) for the ANAM *Throughput* scores (composite scores based on speed and accuracy for each testing module) using 29 high school students tested with a 1 week retest interval (Register-Mihalik et al. 2012a, b; Segalowitz et al. 2007). A separate study addressed the reliability of ANAM using 132 college student athletes over a variable time period ranging from 2 weeks to 4 months with an average retest period of 46.7±20.02 days. The authors observed predominantly weak reliability coefficients for Pearson r's and ICCs (.14 to .38), with the exception of the Math Processing Throughput score, which showed strong reliability (ICC=.86) (*ANAM4 Sports Medicine Battery: Automated Neuropsychological Assessment Metrics* 2010; Register-Mihalik et al. 2012a, b). Overall, most of ANAM's subtests did not meet the minimum criteria (>.60) acceptable for clinical utility.

Validity One peer reviewed paper addressed the validity of ANAM in SRC. Regarding discriminant validity, Segalowitz et al. (2007) reported a significant correlation (.60) between the Continuous Performance Task and Mathematical Processing scores as well as an expected relationship between Code Substitution and Code Substitution Delayed Throughput scores ($r = .65$). Last, a significant correlation (.50) was observed between the two components of Simple Reaction Time throughput scores. Some of these correlations suggest overlap and a lack of discriminant validity between two purportedly distinct cognitive constructs.

Sensitivity/Specificity: Two prospective studies have reported data on the sensitivity and specificity of the ANAM in SRC. Register-Mihalik and colleagues (2012a, b) administered the ANAM to 132 concussed male and female college athletes. Interpretation of ANAM was performed using the RCI with 95 %, 90 % and 80 % confidence intervals. Since a composite

Table 1 Descriptions of each reviewed computerized neurocognitive test

Test	Description	Subtests	Summary Scores
ANAM (ANAM4 Sports Medicine Battery Automated Neuropsychological Assessment Metrics 2010)	Program Type: Desktop PC-based Approximate Time: 25 to 30 min Assessment of Change: Reliable Change Index.	• Simple Reaction Time • Simple Reaction Time Delayed • Matching to Sample • Mathematical Processing • Procedural Reaction Time • Code Substitution • Code Substitution Delayed • Sternberg Procedure • Go-NoGo • Spatial Processing • Memory Search • Sternberg Procedure	NA
CogSport/State (CogState 2011)	Program Type: Web based Approximate Time: 20 min Assessment of Change: Simple difference method and effect sizes	• Detection Task • Identification Task • One Card Learning • One Back Memory • One Back Accuracy • Complex Reaction Time • Continuous Memory • Matching Task • Incidental Memory • Associative Memory	• Processing Speed • Attention • Visual Memory • Working Memory • Simple Reaction Time • Choice Reaction Time • Executive Function
HeadMinder CRI Erlanger (2002)	Program Type: Web based Approximate Time: 25 min Assessment of Change: Reliable Change Index	• Animal Decoding • Symbol Scanning • Continuous Recognition Task • Cued Reaction Time • Visual Recognition (Immediate and Delayed Tasks)	• Processing Speed • Simple Reaction Time • Complex reaction Time • Simple Reaction Time Errors • Complex Reaction Time Errors
ImPACT (Lovell 2013)	Program Type: Desktop or Web based Approximate Time: 30 min Assessment of Change: Reliable Change Index	• Word Memory • Design Memory • X's and O's • Symbol Match • Color Match • Three Letters	• Verbal Memory • Visual Memory • Visual Motor Speed • Reaction Time

Approximate time refers to the approximate time to administer the CNT inclusive of health history and demographic information. *Assessment of change* refers to which statistical method is used to determine neurocognitive change. *(Reference dates)* are reflective of the most recent user's manual. *NA* represents not available.

score does not yet exist for the ANAM battery as a whole, a range of sensitivity values of each throughput score at varying levels of confidence was presented for each subtest. Regardless of the level of confidence employed, use of single-subtest scores alone resulted in correct classification of 1 % to 15 % of concussed athletes tested approximately 1 to 4 days following their injury. The sensitivity of the ANAM when considering two or more throughput scores was not presented. The authors reported a relatively high specificity for each of the ANAM's throughput scores, however, with values ranging from 86 % to 100 %. The authors concluded that clinicians should exercise caution when using the ANAM in the evaluation and management of SRC (Register-Mihalik et al. 2012a).

CogSport/State/Axon Sports

In 2011 CogSport/State acquired Axon Sports, which is the trade name under which the product is distributed in the United States. Despite the name change, the computerized battery of neurocognitive tests was unchanged. To remain consistent with the literature, we will refer to the platform as reported in the literature as CogSport/State (CogState 2011).

Reliability Three published articles were identified that examined the test-retest reliability of CogSport/State. Collie and colleagues (2003a) administered CogSport/State to 60 young volunteers (21.7±3.3 years of age) at two time points. CogSport/State was observed to have moderate to strong ICC values ranging from .69 to .90 at both 1 h and 1 week test-retest intervals. Similar ICC values were reported (.69 to .82) for the 1 week interval compared to 1 h comparison. Another study also examined the test-retest reliability of CogSport/State administering the test four times at 10 min increments and then again 1 week later to 63 healthy participants 21.6±3.80 years of age (Falleti et al. 2006). A second group was administered the test at 10 minute increments and

Table 2 Summary of psychometric properties of each reviewed computerized neurocognitive test

Test	Reliability	Validity	Sensitivity
ANAM	Two peer reviewed papers addressed test-retest reliability over 1 week to 4 months. Most reliability coefficients did not meet minimum criterion for acceptable reliability (.60) Reliability coefficients ranged from .14 to .86 depending on the re-test interval and subtest.	One peer reviewed paper addressed validity, which demonstrated correlation values ranging from −.01 to .65 amongst ANAM subtest scores.	One peer reviewed paper addressed sensitivity/specificity. Sensitivity ranged from 1 % to 15 %, depending on subtest and methodology. Specificity has ranged from 86 % to 100 % depending upon subtest.
CogSport/State	Three peer reviewed papers addressed test-retest reliability. Test-retest intervals ranged from 10 min to 1 month. Most reliability coefficients did not meet the minimum criteria for acceptable reliability. Reliability coefficients ranged from .45 to .90, depending on the retest interval and subtest.	Four peer reviewed papers addressed validity, showing correlation coefficients that ranged from .23 to .83 when compared to relevant traditional neuropsychological tests.	One peer reviewed paper reported 70.8 % sensitivity for the overall test battery. This study did not a control group, limiting the ability to calculate specificity.
HeadMinder CRI	One peer reviewed paper addressed the test-retest reliability of the CRI over approximately 50 days. Most reliability coefficients did not meet minimum criteria for acceptable reliability. Stability coefficients ranged from .03 to .66 depending on subtest.	Two peer reviewed papers addressed validity. Correlation coefficients ranged from .37 to .70 compared to traditional neuropsychological tests.	Two peer reviewed papers reported the sensitivity ranged from 69 % to 78.6 %. Neither study used control groups, limiting the ability to calculate specificity.
ImPACT	Six peer reviewed papers addressed the test-retest reliability with intervals ranged from 1 day to 2 years. Most reliability coefficients did not meet minimum criteria for acceptable reliability. Reliability coefficients ranged from .23 to .88 depending on the retest interval and subtest.	Three peer reviewed papers addressed validity. Correlation coefficients ranged from .20 to .88 compared to traditional neuropsychological tests.	Four peer reviewed papers addressed sensitivity/specificity. Sensitivity values ranged from 79.2 % to 94.6 % depending on whether based on cognitive subtests only or a combination of subtests and symptom score. Specificity has ranged from 89.4 % to 97.3 %.

then again one month later. The authors reported 7 participants were excluded from group 1 but did not report the average age of the final sample used for data analysis. Reliability coefficients were calculated between each successive time point, but not among all time points. The authors observed weak (.35 to .38) to strong (.70 to .94) ICCs for CogSport/State's reaction time and weak (0 to .39) to moderate (.41 to .66) reliability values for subtest accuracy scores, although the majority (52.5 %) of ICC values fell above the minimum criterion value acceptable for clinical utility (Falleti et al. 2006). Straume-Naesheim and colleagues assessed the test-retest reliability of CogSport/State in Norwegian elite football (soccer) players. A total of 232 athletes were administered CogSport/State consecutively over 2 days and moderate to strong ICC values (.45 to .79) for mean reaction time and weak reliability values (.14 to .31) for subtest accuracy were observed (Straume-Naesheim et al. 2005).

Validity Four published studies that examined the validity of CogSport/State in SRC were reviewed. Collie and colleagues (2003a, b) examined the concurrent validity of CogSport/State by comparing its composite scores to the Digit Symbol Substitution Test and Trail Making Test—Part B when administered to 240 elite athletes who were 22.0±4.1 years of age.

CogSport/State possessed moderate to strong (.42 to .50) ICCs when compared to the Digit Symbol Substitution Test and weak to moderate values (.23 to .44) for the Trail Making Test. Maruff and colleagues (2009) examined both construct and criterion validity in both concussed (*n*=50) and healthy (*n*= 50) participants. To establish convergent validity, CogSport/State was compared to well-evidenced traditional neuropsychological tests in 215 healthy adults between 35 and 50 years of age. The CogSport/State was observed to have weak to strong (.04 to .83) reliability coefficients when compared to traditional measures, although significant correlations were observed for each CogSport/State neurocognitive score and its traditional paper and pencil test equivalent. The traditional measures included the Grooved Peg Board Test, the Trail Making Test, the Symbol Digit Modalities Test (.57 to .74), the Wechsler Memory Scale 3rd Edition Spatial Span Subtest (.69 to .80), the Brief Visuospatial Memory Test (.54 to .83), and the Rey Complex Figure Test-Delayed Recall (.79). Additionally, weak correlation coefficients were observed for computerized and traditional neurocognitive measures that would be assumed to have little to no relationship, demonstrating divergent validity (Maruff et al. 2009).

The authors reported evidence of criterion validity in the form of *t*-tests, effect size, and non-overlapping statistics for

Table 3 A summary of published test retest reliability coefficients for CNT summary scores

Test	Test retest intervals	Summary score	r	ICC	References
ANAM	NA	NA	NA	NA	NA

CogSport/State	10 min–1 month
	1 h–1 week
	1 day
		• Processing Speed
		  Speed
		  Accuracy
			
			NA
			NA
				
				.76–.90[a]
				.20–.31[a]
					(Falleti et al. 2006)[a]
					(Straume-Naesheim et al. 2005)
					(Collie et al. 2003a)[a]

		• Attention			
		Speed	NA	NA	
		Accuracy	NA	NA	
		• Visual Memory			
		Speed	NA	NA	
		Accuracy	NA	NA	
		• Working Memory			
		Speed	NA	.72–.76[a]	
		Accuracy	NA	.24–.51[a]	
		• Simple Reaction Time			
		Speed	NA	.73–.94[a]	
		Accuracy	NA	−.06–.45[a]	
		• Choice Reaction Time			
		Speed	NA	.38–.81[a]	
		Accuracy	NA	.14–58[a]	
		• Executive Function			
		Speed	NA	NA	
		Accuracy	NA	NA	
HeadMinder CRI	45–50 days	• Reaction Time • Decision Making • Matching • Attention • Working Memory	NA NA NA NA NA	.55–.60[a] .56–.62[a] .23–.66[a] .39–43[a] .64–.65[a]	(Broglio et al. 2007a)
ImPACT	1–3 days 4 weeks 45–50 day 45–50 days 1.2 years 2 years	• Verbal Memory • Visual Memory • Visual Motor Speed • Reaction Time	.26–.87[a] .43–.55[a] .60–.86[a] .52–.63[a]	.19–78[a] .26–.85[a] .38–.85[a] .39–.88[a]	(Broglio et al. 2007a)[a] (Elbin et al. 2011) (Schatz and Ferris 2013) (Register-Mihalik et al. 2012b)[a] (Resch et al. 2013)[a] (Schatz 2009)

NA represents not available

[a] More than one value was reported for the cited study

both concussed and control participants. Overall, each statistical method provided evidence of criterion validity. One methodological issue with this study was that concussed participants were tested approximately 72 days after injury (Maruff et al. 2009). Thus these results may have limited clinical meaningfulness when working with concussed athletes due to the post-injury assessment occurring well beyond the typical SRC recovery window (Guskiewicz et al. 2003). Schatz and Putz (2006) also investigated the concurrent validity of the CogSport/State with traditional paper and pencil neuropsychological tests. They administered CogSport/State along with the CRI and ImPACT to 30 college aged participants with an average age of 21 years. The authors correlated the composite scores of each CNT with traditional paper and pencil tests. Moderate, but significant correlations (.53 to .54) were observed between CogSport/State and the Trail Making Test (Parts A and B) and the Digit Symbol subtest of the Wechsler Scale-Revised for simple and complex reaction time (Schatz and Putz 2006).

Sensitivity/Specificity: One published study looked at the sensitivity and specificity of CogSport/State. A study by Makdissi and colleagues (2010) employed CogState in 78 concussed Australian football athletes between 16 and 35 years of age who were administered the test ≤48.6 h of their diagnosis. When compared to baseline values, 70.8 % of athletes were observed to have at least one deficiency as measured by CogSport/State (Makdissi et al. 2010).

Table 4 A summary of published test retest reliability coefficients for each CNT subtest score

Test	Test Retest Intervals	Subtest Scores	r	ICC	References
ANAM	7 days 47 days	• Simple Reaction Time	.30–.65	.29–.44	(Register-Mihalik et al. 2012a) (Segalowitz et al. 2007)
		• Simple Reaction Time Delayed	.15–.50	.14–.47	
		• Matching to Sample	.33–.72	.33–.72	
		• Math Processing	.33–.71	.33–.61	
		• Procedural Reaction Time	.33	.33	
		• Code Substitution	.38–.81	.37–.58	
		• Code Subset Delayed	.68	.59	
		• Sternberg Procedure	.31	.31	
CogSport/State	10 min–1 week 1 day	• Detection Task			(Falleti et al. 2006)[a] (Straume-Naesheim et al. 2005) (Collie et al. 2003a)[a]
		Speed	NA	NA	
		Accuracy	NA	NA	
		• One Card Learning			
		Speed	NA	.79–.83[a]	
		Accuracy	NA	.31–.47[a]	
		• One Back Speed			
		Speed	NA	.62–.82[a]	
		Accuracy	NA	.25–.47[a]	
		• Complex Reaction Time			
		Speed	NA	.41–.66[a]	
		Accuracy	NA	.42–.56[a]	
		• Congruent Reaction Time			
		Speed	NA	.69	
		Accurracy	NA	.23	
		• Continuous Monitoring			
		Speed	NA	.45–.74[a]	
		Accuracy	NA	NA	
		• Matching Task			
		Speed	NA	.69–.85	
		Accuracy	NA	.19–.66	
		• Incidental Learning			
		Speed	NA	.35–.74[a]	
		Accuracy	NA	.47–.56[a]	
		• Associative Learning			
		Speed	NA	.77–.86[a]	
		Accuracy	NA	.46–.66[a]	
		• Learning			
		Speed	NA	.79–.83[a]	
		Accuracy	NA	.31–.47[a]	
HeadMinder CRI	NA	NA	NA	NA	
ImPACT	NA	NA	NA	NA	

[a] More than one value was reported for the cited study

HeadMinder Concussion Resolution Index (CRI)

Reliability One study was identified that addressed the test-retest reliability of the CRI as part of a comparison with ImPACT and Concussion Sentinel. Participants were administered the three CNTs at three separate time points which consisted of Days 1, 45, and 50 to healthy college aged participants (Broglio et al. 2007a). The lowest ICC value reported for CRI was for Simple Reaction Time Errors (.03) and the highest value (.66) was for Complex Reaction Time. Overall, the majority of the CRI's neurocognitive values fell below the minimum criterion reliability value (.60) to suggest limited clinical utility (Anastasi 1998; Portney and Watkins 1993). The values reported by Broglio and colleagues were

much lower than those noted by Erlanger et al. (2001) in a previous, unpublished study referenced by Erlanger and colleagues who administered the CRI twice using a 2 week test-retest time interval and reported reliability coefficients ranging from .65 to .90 (Erlanger et al. 2001).

Validity Two articles were identified that addressed the validity of the CRI. Preliminary evidence of concurrent validity was reported comparing the CRI's neurocognitive composite scores to the scores of well-established traditional neuropsychological measures tapping similar domains (Erlanger et al. 2001). Correlation coefficients ranged from .46 to .70 when the CRI's composite scores were compared to measures of processing speed and reaction time (Erlanger et al. 2001). Last, the CRI's measures of simple reaction time and processing speed demonstrated moderate, but significant correlations (−.61 to .37) with traditional paper-and-pencil neuropsychological measures (Schatz and Putz 2006).

Sensitivity/Specificity: Two articles were indentified that reported on the sensitivity of the CRI. Erlanger and colleagues administered baseline and follow-up CRI assessments to 26 high school athletes who were diagnosed as concussed during their respective sport seasons. The CRI was administered ≤48 h following the diagnosis of concussion. The CRI correctly identified 23 of 26 concussed athletes based on self-reported symptoms and/or neurocognitive decline as measured by the CRI. When examining the CRI's neurocognitive results alone, 69 % of concussed athletes demonstrated neurocognitive change following injury. In a separate study, Broglio and colleagues (2007b) administered the CRI to collegiate athletes approximately 24 h after concussion. The CRI correctly identified 78.6 % (22/28) of concussed athletes who were observed to have a neurocognitive deficit on at least one of the CRI's composite neurocognitive scores. Neither Broglio nor Erlanger used control groups, thereby precluding their ability calculate the CRI's specificity.

Immediate Postconcussion and Cognitive Testing Test Battery (ImPACT)

Reliability Six published studies that investigated the test-retest reliability of the ImPACT in high school and college aged controls. Register-Mihalik and colleague (2012a, b) administered the ImPACT along with several traditional paper-and-pencil neuropsychological measures to both college (M age=20.0±.79 years) and high school (M age= 16.0±.86 years) subjects. Participants were assessed at three time points separated by approximately 24 h between sessions. ICC values ranged from weak to strong, depending on the subtest. The weakest ICC value was observed for Verbal Memory (.29), while Processing Speed (.71) demonstrated the highest reliability (Register-Mihalik et al. 2012b).

Broglio and colleagues administered forms 1, 2, and 3 of the ImPACT at three time points (Days 1, 45, and 50) in a study of college students designed to reflect the typical time between baseline, post-injury, and follow-up assessment. ImPACT was administered with two additional neurocognitive measures (CRI and The Concussion Sentinel). The authors reported weak to moderate reliability coefficients across time for each of the ImPACT's neurocognitive indices. Verbal Memory (.23) was observed to be the least reliable measure while Visual-Motor Speed (.61) demonstrated the highest reliability. A critique of the design was that multiple neurocognitive measures were administered in a 60 min testing period which may have lead to reduced effort or fatigue issues and hence, sub-optimal reliability coefficients (Schatz and Sandel 2013). More recently, Resch and colleagues (2013) replicated Broglio's work administering the ImPACT alone to 45 American and 46 Irish college aged students (Resch et al. 2013). The American students were administered the ImPACT at Days 1, 45 and 50. The Irish students completed the ImPACT at Days 1, 7, and 14. Similar to Broglio, Resch and colleagues observed weak to strong reliability coefficients for both the American (.37 to .76) and Irish (.26 to .88) students, suggesting that the findings by Broglio et al. were not due to cognitive fatigue. Schatz administered form 1 of the ImPACT (baseline) at time point 1 and form 2 approximately 30 days later to a group of 25 college aged students. Schatz observed ICC values ranging from moderate (.60) for Visual Memory to strong (.88) for Visual Motor Speed between time points (Schatz and Ferris 2013). Overall, when using time points representative of the management of SRC, the ImPACT's test-retest reliability has been reported to range between .23 and .88 (Broglio et al. 2007a; Resch et al. 2013; Schatz and Ferris 2013).

Reliability of the ImPACT has also been examined using longer test-retest intervals. Schatz administered form 1 of the ImPACT to 95 varsity college athletes (M age=20.8) at baseline and then again approximately 2 years later. Moderate to strong reliability coefficients were reported ranging from .46 for Verbal Memory to .74 for Processing Speed (Schatz 2009). Elbin et al. reported the ImPACT possessed moderate to strong ICCs (.62 to .85) over the 1 year test-retest interval using form 1 in 369 high school athletes (M age=14.8 years) (Elbin et al. 2011). Overall, the ImPACT was observed to possess moderate to strong (.46 to .85) reliability coefficients over 1 to 3 year test-retest intervals (Elbin et al. 2011; Schatz 2009).

Validity Three published studies were reviewed which addressed construct and concurrent validity of ImPACT. Maerlender administered the ImPACT as well as the N-Back and Verbal Continuous Memory tasks in one or two test sessions to 54 college aged (M=19.1 years) athletes. Statistically significant but weak to moderate (.31 to .59),

correlations between each of the ImPACT's neurocognitive domains and the domains measured by each of these two particular neuropsychological measures were reported (Maerlender et al. 2010).

The authors also performed inter-correlations amongst the ImPACT composite scores and reported weak to moderate (.27 to .38), but significant correlations (Maerlender et al. 2010). Similarly, Allen and Gfeller (2011) administered the ImPACT along with the Hopkins Verbal Learning Test-Revised, Brief-Visual Spatial Memory Test-Revised, Trail Making Test A and B, Controlled Oral Word Association Test, and the Symbol Search, Digit Symbol-Coding, and Digit Span subtests of the Wechsler Adult Intelligence Scale, 3rd edition to 100 healthy college aged students ($M=19.7$ years) during one testing session. The authors reported each of the ImPACT's neurocognitive composite scores demonstrated at least two weak to moderate (.20 to .43), but statistically significant correlations with one or more traditional paper-and-pencil neuropsychological tests. Higher correlations were demonstrated between cognitive measures tapping similar neurocognitive domains (Allen and Gfeller 2011). Similar to Maerlender (2010), the authors also reported several weak to strong (.24 to .88), but statistically significant, correlations amongst each of the ImPACT's neurocognitive indices (Allen and Gfeller 2011). Additionally, Schatz and Putz (2006) observed that ImPACT showed moderate, but significant, correlations (.44 to .64) on complex reaction time when compared to the Trail Making Test (Parts A and B) and the Digit Symbol subtest of the Wechsler Adult Intelligence Scale-Revised.

In terms of construct validity, Allen and Gfeller (2011) reported the ImPACT's neurocognitive composite scores explained 69 % of the variance in a factor model consisting of forced choice efficiency, verbal and visual memory, inhibitory cognitive ability, visual processing abilities, and an additional factor comprised of Color Match Total Commissions score.

Sensitivity/Specificity: Four published studies were reviewed which addressed the sensitivity and/or specificity of the ImPACT. The first conducted a retrospective analysis of concussed high school athletes (tested ≤72 h post-injury) compared to healthy controls with no history of concussion. Using descriptive discriminant analysis of the neurocognitive and symptom scale composite scores, ImPACT's sensitivity was reported to be 81.9 % and specificity 89.4 % (Schatz et al. 2006). Similarly, Van Kampen and colleagues (2006) reported ImPACT correctly classified 93 % of concussed high school and college athletes when coupled with self-reported symptoms. The sensitivity of ImPACT was also examined in concussed collegiate athletes evaluated ≤24 h of their diagnosis (Broglio et al. 2007b). In order to determine sensitivity, ImPACT results were compared to preseason baseline values, with results indicating the

sensitivity of ImPACT to be 79.2 % for cognitive and total symptom score (Broglio et al. 2007a, b). None of the three studies reported the sensitivity of solely the neurocognitive measures; rather, Impulse Control and Total Symptom Scores were included in their analyses. This is a critical issue since the value of a cognitive test should be the unique contribution it makes over and above symptom reporting.

More recently, the sensitivity of the online version of ImPACT was examined via retrospective analysis of data obtained from concussed athletes between the ages of 13 and 21 who were administered the test within72 h of injury (Schatz and Sandel 2013). Data from concussed athletes were compared to healthy controls that ranged from 13 to 21 years of age. Overall, the authors reported the online version of ImPACT's cognitive and symptom scores possessed a sensitivity of 91.4 % and specificity of 69.1 %. A subsequent analysis revealed that the online version of the ImPACT correctly classified 94.6 % of asymptomatic concussed athletes and 97.3 % of healthy controls based on cognitive and total symptom composite scores.

Summary and Clinical Considerations

The purpose of the current review was to summarize the published evidence on psychometric properties of several commercially available CNTs used to evaluate SRC. Overall, we reviewed psychometric concepts integral to establishing the reliability, validity and clinical utility of CNTs, as well as the statistical method(s) employed by each program to assess change in neurocognitive performance associated with SRC.

Our review indicated that the body of research addressing the psychometric properties of CNTs has significantly increased since 2005, concurrent with a major increase in CNTs being adopted in clinical practice. Over that same period, the evidence base on CNT test-retest reliability and validity has been broadened to include studies of pediatric, college, and professional athletes (Broglio et al. 2007b; Register-Mihalik et al. 2012; Schatz and Sandel 2013). Since an initial review in 2005, 29 articles have been published addressing one or more psychometric properties of commercially available CNTs in the assessment of SRC. The highest number of publications involved was for ImPACT (13), followed by CogSport/State (7), CRI (5), and ANAM (4). While we recognize that several other CNTs have recently entered the marketplace, no published studies on the psychometric properties of those instruments specific to SRC were available at the time of our review.

Reported reliability values for the reviewed CNTs were highly variable, with factors such as sample composition and test-retest time intervals being partially responsible for the lack of observed consistency. A sizeable number of CNT subtest and/or summary scores met or exceeded the minimum

reliability value recommended for clinical utility (≥.60) but no CNT's met this criterion for all subtest and/or summary scores. In terms of validity, highly variable evidence has been reported for each CNT. In terms of sensitivity and specificity, values ranged from 1 % to 94.6 % for sensitivity and 86 % to 100 % for specificity for each CNT, including metrics for the battery, subtest, and summary scores (see Table 2). Limited data on other forms of validity have been reported in the literature.

Despite the increase in literature addressing CNTs to manage SRC since 2005, the actual number of published studies was rather limited in view of the level of market penetration of CNTs and adoption into clinical practice over the same time frame. Although many of the studies reviewed demonstrate suboptimal reliability and validity of CNTs, computerized testing is widely used in concussion management at all competitive levels. As in the case of traditional neuropsychological tests, clinicians should be familiar with the psychometric evidence that supports the use of their preferred CNT, as well as those factors that affect test performance, results, and interpretation.

A number of limitations affected the interpretation of study findings on the psychometrics of the reviewed CNTs, leaving key questions about the clinical utility of CNTs incompletely answered. First, although CNTs are commonly used for the management of SRC across a wide age span, age-specific tasks and norms are generally lacking. Several of the reviewed studies used samples consisting of a wide range of ages, which may have influenced the values reported for test-retest reliability and validity (Broglio et al. 2007a; Elbin et al. 2011; Schatz 2009; Schatz and Sandel 2013).

Another important factor that varied across studies was the test-retest period used by each study to address reliability. Test-retest intervals in the studies reviewed ranged from 1 h to approximately 2 years (Collie et al. 2003a, b; Schatz 2009). Test-retest intervals that are based on empirical data and are clinically relevant (i.e., parallel to that observed in a real sports medicine setting) are more applicable to SRC.

In terms of establishing sensitivity and specificity, the assessment period following SRC must occur while the neurocognitive effects of concussion are most evident. It is well established that approximately 80–90 % of adult athletes will self-report a full symptom recovery within 7 to 10 days (Guskiewicz et al. 2003), and that effect sizes on cognitive testing are largest during the acute period within the first 7 days after injury (Belanger and Vanderploeg 2005; Broglio and Puetz 2008). Ideally, studies of the sensitivity, specificity, and predictive validity of CNTs would include testing as soon as possible (e.g. within 24 to 48 h) after injury when signal detection is maximized. It will be important for more studies to employ CNTs at different stages of recovery in order to determine accurate sensitivity/specificity values and cognitive recovery curves after SRC.

One important factor in interpreting CNT results is subject effort, as inadequate or inconsistent effort may lead to greater variability in baseline and/or postconcussion CNT performance. Inadequate effort has been addressed by several studies, with reports of poor effort ranging from 4 % to 11 % (Hunt et al. 2007; Schatz et al. 2012).

Along these lines, a recently recognized phenomenon is that athletes at the high school, college, and professional levels of sport may intentionally provide inadequate effort on baseline or post-injury computerized neurocognitive tests. *Sandbagging*, or purposefully putting forth reduced effort to provide a minimally valid test has become a cause for concern. Motivation for sandbagging can be that athletes wish to avoid appearing "impaired" following SRC (and hence may avoid being removed from play or potentially losing their position, etc.), and as a result, they may intentionally put forth reduced effort on tests at baseline. Recent studies have reported that approximately 11 % to 35 % of participants who were either told not to provide their best effort *or* who were provided information regarding the validity measure of the ImPACT were able to successfully complete the test without violating a standard validity indicator (Erdal 2012; Schatz and Glatts 2013). Though only a minority of participants successfully evaded the standard validity criteria, suboptimal effort is still a cause for concern, as poor effort may be evident yet can escape detection by standard CNT invalidity indicators.

This is of further concern insofar as approximately half of athletic trainers using ImPACT do not conduct a quality control review of baseline assessments for validity (Covassin et al. 2009). In our own experience (n=1,028), up to 47 % of high school athletes produce invalid baseline results (as determined by a board-certified clinical neuropsychologist) even when standard invalidity criteria are not met (Resch et al., unpublished data). For example, baseline test results at the first percentile in an otherwise normal and intact athlete should raise suspicion of test invalidity even though such a score may not be identified by a standard CNT validity indicator. As with any other medical or neuropsychological test results, the assessment metric is no more useful or valid than the interpretation provided by the expert reviewing it. Though each testing platform has its own unique validity criteria, a review of each athlete's baseline test performance compared against normative data may assist clinicians in identifying suboptimal effort and performance.

Several factors relevant to user training and credentials deserve consideration when it comes to the administration of CNTs in concussion management protocols. Numerous governing bodies, including the American Academy of Neurology, American Medical Society for Sports Medicine, National Athletic Trainers' Association, Concussion In Sport Group, American Academy of Clinical Neuropsychology, and the National Academy of Neuropsychology have all advocated the inclusion of a licensed clinical neuropsychologist into the

SRC management team (Bauer et al. 2012; Giza et al. 2013; Guskiewicz et al. 2004; Harmon et al. 2013; McCrory et al. 2013). Despite these recommendations, it is estimated that between 2009 and 2012 less than 3 % of SRCs were assessed by neuropsychologists, with the majority of concussed athletes being assessed by ATs and general physicians (Comstock et al. 2012; Meehan et al. 2012).

Athletic trainers and physicians are well versed in both on-the-field management and ongoing care of a concussed athlete, but most do not possess specialized training to accurately detect and interpret the subtle nuances associated with CNTs (Comstock et al. 2012; Meehan et al. 2012). Further, published reports indicate that less than half of athletic trainers who use CNTs obtain the recommended level of CNT-specific training on the instrument (Covassin et al. 2009). Clinical neuropsychologists obtain extensive training in brain-behavior/relationships and are an integral part of the assessment of SRC in terms of providing education and reassurance, assessment of concussed athlete's neurocognitive state, guiding an athlete's return to school, and treatment of emotional or adjustment problems that may occur (Echemendia et al. 2012).

Clinical Recommendations

Overall, health care professionals who implement CNTs for the management of concussions should:

- Make an informed decision when implementing a CNT and understand its limitations when managing SRC.
- Ensure that a CNT is enlisted as part of a multi-dimensional approach to SRC assessment, and that a CNT is not the sole basis for diagnosing SRC or decision-making about an athlete's fitness to return to play after SRC.
- Incorporate a clinical neuropsychologist into the SRC management team in order to assist with the interpretation of both baseline and/or post injury assessments as well as treatment of concussed athletes (Echemendia et al. 2012).
- Ensure thorough training for those who administer CNTs to increase the objectivity, reliability and validity of the test results (Covassin et al. 2009).
- If possible, administer the CNT individually or in small groups as possible (i.e. ≤ 4 athletes) in order to minimize distraction and achieve values closer to an individual's true neurocognitive ability (Moser et al. 2011).
- Conduct quality control reviews of baseline assessment tests in order to identify invalid tests or signs of suboptimal effort (Covassin et al. 2009; Erdal 2012; Schatz et al. 2012).

Future research to address key issues on the utility of CNT in SRC should:

- Further demonstrate and report the psychometric properties of each CNTs in the assessment of SRC through prospective studies of adult and youth athletes
- Investigate the added value of CNTs compared to traditional neuropsychological tests as measures of cognitive recovery at sequential time points following injury (McCrea et al. 2013)
- Investigate additional sources of random and systematic error which may influence CNT performance
- Investigate the psychometric evidence for equivalence of alternate forms provided by each CNT
- Assess the reliability and validity of a baseline versus no baseline model of CNT assessment when measuring the acute effects of SRC (McCrea et al. 2013)
- Establish guidelines for "best practices" in the administration of CNTs across multiple sports, age groups and levels of competition

Conclusion

Over the past 8 years, progress has been made in studying the psychometric properties of CNTs used in the assessment of SRC, although CNT development, marketing and sales appear to have outpaced the clinical evidence base. Psychometric properties such as those reviewed in this article are critical for clinicians to understand when incorporating CNTs into a SRC management protocol. CNT users should apply caution and appropriate clinical expertise when interpreting test results in view of the many caveats and limitations relevant to CNTs, as discussed in this review. As CNTs continue to improve in terms of their diagnostic accuracy and more evidence on their utility becomes available, CNTs are likely to remain a central component of a multidimensional approach to the management of SRC.

References

Allen, B. J., & Gfeller, J. D. (2011). The Immediate Post-Concussion Assessment and Cognitive Testing battery and traditional neuropsychological measures: a construct and concurrent validity study. *Brain Injury, 25*(2), 179–191.

ANAM4 Sports Medicine Battery: Automated Neuropsychological Assessment Metrics (2010). Norman: Center for the Study of Human Operator Performance.

Anastasi, A. (1998). *Psychological testing* (6th ed.). New York: Macmillan.

Association, A. E. R. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Education Research Association.

Aubry, M., Cantu, R., Dvorak, J., Graf-Baumann, T., Johnston, K. M., Kelly, J., et al. (2002). Summary and agreement statement of the 1st International Symposium on Concussion in Sport, Vienna 2001. *Clinical Journal of Sports Medicine, 12*(1), 6–11.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist, 26*(2), 177–196.

Baumgartner, T. A., & Hensley, L. D. (2006). *Conducting and reading research in health and human performance* (4th ed.). New York: McGraw Hill.

Belanger, H. G., & Vanderploeg, R. D. (2005). The neuropsychological impact of sports-related concussion: a meta-analysis. *Journal of International Neuropsychological Society, 11*(4), 345–357.

Broglio, 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*(1), 53–67.

Broglio, S. P., Ferrara, M. S., Macciocchi, S. N., Baumgartner, T. A., & Elliott, R. (2007a). Test-retest reliability of computerized concussion assessment programs. *Journal of Athletic Training, 42*(4), 509–514.

Broglio, S. P., Macciocchi, S. N., & Ferrara, M. S. (2007b). Sensitivity of the concussion assessment battery. *Neurosurgery, 60*(6), 1050–1057.

Cernich, A. N., Brennana, D. M., Barker, L. M., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Archives of Clinical Neuropsychology, 22*(Suppl 1), S39–S48.

CogState (2011). *CogState Research Manual (Version 6.0)*. New Haven: CogState Ltd.

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*. doi:10.1093/arclin/act040.

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

Collie, A., Maruff, P., McStephen, M., & Darby, D. (2003b). Are Reliable Change (RC) calculations appropriate for determining the extent of cognitive change in concussed athletes? *British Journal of Sports Medicine, 37*(4), 370–372.

Collie, A., Maruff, P., Makdissi, M., McStephen, M., Darby, D. G., & McCrory, P. (2004). Statistical procedures for determining the extent of cognitive change following concussion. *British Journal of Sports Medicine, 38*(3), 273–278.

Comstock, R. D., Collins, C. L., Corlette, J. D., & Fletcher, E. N. (2012). *National high school sports- related injury surveillance study: 2011–2012 school year*. Columbus: The Research Institute at Nationwide Children's Hosptial.

Covassin, T., Elbin, R. J., 3rd, 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*(6), 639–644.

Covassin, T., Elbin, R., Kontos, A., & Larson, E. (2010). Investigating baseline neurocognitive performance between male and female athletes with a history of multiple concussion. *Journal of Neurology, Neurosurgery, and Psychiatry, 81*(6), 597–601.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason: Cengage Learning.

Dancey, C., & Reidy, J. (2004). *Statistics without maths for psychology: using SPSS for windows*. London: Prentice Hall.

Echemendia, R. J., Iverson, G. L., McCrea, M., Broshek, D. K., Gioia, G. A., Sautter, S., et al. (2012). Role of neuropsychologists in the evaluation and management of sport-related concussion: an inter-organization position statement. *The Clinical Neuropsychologist, 25*(8), 1289–1294.

Elbin, R., Schatz, P., & Covassin, T. (2011). One-year test-retest reliability of the online version of ImPACT in high school athletes. *The American Journal of Sports Medicine, 39*(11), 2319–2324.

Erdal, K. (2012). Neuropsychological testing for sports-related concussion: how athletes can sandbag their baseline testing without detection. *Archives of Clinical Neuropsychology, 27*(5), 473–479.

Erlanger, D., Saliba, E., Barth, J., Almquist, J., Webright, W., & Freeman, J. (2001). Monitoring resolution of postconcussion symptoms in athletes: preliminary results of a web-based neuropsychological test protocol. *Journal of Athletic Training, 36*(3), 280–287.

Falleti, M. G., Maruff, P., Collie, A., & Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology, 28*(7), 1095–1112.

Ferrara, M. S., McCrea, M., Peterson, C. L., & Guskiewicz, K. M. (2001). A survey of practice patterns in concussion assessment and management. *Journal of Athletic Training, 36*(2), 145–149.

Giza, C. C., Kutcher, J. S., Ashwal, S., Barth, J., Getchius, T. S., Gioia, G. A., et al. (2013). Summary of evidence-based guideline update: evaluation and management of concussion in sports: Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology, 80*(24), 2250–2257.

Guskiewicz, K. M., Riemann, B. L., Perrin, D. H., & Nashner, L. M. (1997). Alternative approaches to the assessment of mild head injury in athletes. *Medicine and Science in Sports and Exercise, 29*(7 Suppl), S213–S221.

Guskiewicz, K. M., McCrea, M., Marshall, S. W., Cantu, R. C., Randolph, C., Barr, W., et al. (2003). Cumulative effects associated with recurrent concussion in collegiate football players: the NCAA Concussion Study. *Journal of the American Medical Association, 290*(19), 2549–2555.

Guskiewicz, K. M., Bruce, S. L., Cantu, R. C., Ferrara, M. S., Kelly, J. P., McCrea, M., et al. (2004). National athletic trainers' association position statement: management of sport-related concussion. *Journal of Athletic Training, 39*(3), 280–297.

Harmon, K. G., Drezner, J. A., Gammons, M., Guskiewicz, K. M., Halstead, M., Herring, S. A., et al. (2013). American Medical Society for Sports Medicine position statement: concussion in sport. *British Journal of Sports Medicine, 47*(1), 15–26.

Erlanger, D. M. (2002). *HeadMinder concussion resolution index (CRI): professional manual*. New York, NY: HeadMinder.

Hinton-Bayre, A. D. (2012). Choice of reliable change model can alter decisions regarding neuropsychological impairment after sports-related concussion. *Clinical Journal of Sport Medicine, 22*(2), 105–108.

Hunt, T. N., Ferrara, M. S., Miller, L. S., & Macciocchi, S. (2007). The effect of effort on baseline neuropsychological test scores in high school football athletes. *Archives of Clinical Neuropsychology, 22*(5), 615–621.

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2003). Interpreting change on ImPACT following sport concussion. *Clinical Neuropsychologist, 17*(4), 460–467.

Lovell, M. (2007). *ImPACT 2007 (6.0) software user's manual*. Pittsburgh: ImPACT Applications Inc.

Lovell, M. R. (2013). *Immediate post-concussion assessment testing (ImPACT) test: clinical interpretive manual online ImPACT 2007–2012*. Pittsburgh: ImPACT Applications Inc.

Lovell, M., Echemendia, R. J., Barth, J. T., & Collins, M. W. (Eds.). (2004). *Traumatic brain injury in sports: A neuropsychological and international perspective*. Lisse: Swets & Zeitlinger B.V.

Macciocchi, S. N., Barth, J. T., Alves, W., Rimel, R. W., & Jane, J. A. (1996). Neurological functioning and recovery after mild head injury in collegiate athletes. *Neursurgery, 39*, 510–514.

Maerlender, A., Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., et al. (2010). Examination of the construct validity of

ImPACT computerized test, traditional, and experimental neuropsychological measures. *The Clinical Neuropsychologist, 24*(8), 1309–1325.

Makdissi, M., Darby, D., Maruff, P., Ugoni, A., Brukner, P., & McCrory, P. R. (2010). Natural history of concussion in sport: markers of severity and implications for management. *The American Journal of Sports Medicine, 38*(3), 464–471.

Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., et al. (2009). Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Archives Clinical Neuropsychology, 24*(2), 165–178.

McCrea, M. (2001). Standardized mental status testing on the sideline after sport-related concussion. *Journal of Athletic Training, 36*(3), 274–279.

McCrea, M., Iverson, G. L., Echemendia, R. J., Makdissi, M. & Raftery, M. (2013). Day of injury assessment of sport-related concussion. *British Journal of Sports Medicine, 47*(5), 272–284.

McCrory, P., Collie, A., Anderson, V., & Davis, G. (2004). Can we manage sport related concussion in children the same as in adults? *British Journal of Sports Medicine, 38*(5), 516–519.

McCrory, P., Johnston, K., Meeuwisse, W., Aubry, M., Cantu, R., Dvorak, J., et al. (2005). Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. *British Journal of Sports Medicine, 39*(4), 196–204.

McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Molloy, M., et al. (2009). Consensus Statement on Concussion in Sport: the 3rd International Conference on Concussion in Sport held in Zurich, November 2008. *British Journal of Sports Medicine, 43*(Suppl 1), i76–i90.

McCrory, P., Meeuwisse, W. H., Aubry, M., Cantu, B., Dvorak, J., Echemendia, R. J., et al. (2013). Consensus statement on concussion in sport: the 4(th) international conference on concussion in sport held in Zurich, November 2012. *Journal of the American College of Surgeons, 216*(5), e55–e71.

Meehan, W. P., 3rd, 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*(1), 38–44.

Moser, R. S., Schatz, P., Neidzwski, K., & Ott, S. D. (2011). Group versus individual administration affects baseline neurocognitive test performance. *The American Journal of Sports Medicine, 39*(11), 2325–2330.

Notebaert, A. J., & Guskiewicz, K. M. (2005). Current trends in athletic training practice for concussion assessment and management. *Journal of Athletic Training, 40*(4), 320–325.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Pedhazur, E. J., & Pedhazur Schmelkin, L. (1991). *Measurement, design and analysis: An integrated approach*. Hillsdale: Lawrence Erilbaum Associates, Inc., Publishers.

Portney, L. G., & Watkins, M. P. (1993). *Foundations of clinical research: Applications to practice*. Norwalk: Appleton & Lange.

Randolph, C. (2011). Baseline neuropsychological testing in managing sport-related concussion: does it modify risk? *Current Sports Medicine Reports, 10*(1), 21–26.

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., Guskiewicz, K. M., Mihalik, J. P., Schmidt, J. D., Kerr, Z. Y., & McCrea, M. A. (2012a). Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: implications for caution in clinical practice. *The Journal of Head Trauma Rehabilitation, 28*(4), 274–283.

Register-Mihalik, J. K., Kontos, D. L., Guskiewicz, K. M., Mihalik, J. P., Conder, R., & Shields, E. W. (2012b). Age-related differences and reliability on computerized and paper-and-pencil neurocognitive assessment batteries. *Journal of Athletic Training, 47*(3), 297–305.

Resch, J., Driscoll, A., McCaffrey, N., Brown, C., Ferrara, M. S., Macciocchi, S., et al. (2013). ImPact test-retest reliability: reliably unreliable? *Journal of Athletic Training, 48*(4), 506–511.

Roebuck-Spencer, T., Sun, W., Cernich, A. N., Farmer, K., & Bleiberg, J. (2007). Assessing change with the Automated Neuropsychological Assessment Metrics (ANAM): issues and challenges. *Archives of Clinical Neuropsychology, 22*(Suppl 1), S79–S87.

Schatz, P. (2009). long-term test-retest reliability of baseline cognitive assessments using ImPACT. *American Journal of Sports Medicine, 38*(1), 47–53.

Schatz, P., & Ferris, C. S. (2013). One-month test-retest reliability of the ImPACT test battery. *Archives of Clinical Neuropsychology, 28*(5), 499–504.

Schatz, P., & Glatts, C. (2013). "Sandbagging" baseline test performance on ImPACT, without detection, is more difficult than it appears. *Archives of Clinical Neuropsychology, 28*(3), 236–244.

Schatz, P., & Putz, B. O. (2006). Cross-validation of measures used for computer-based assessment of concussion. *Applied Neuropsychology, 13*(3), 151–159.

Schatz, P., & Sandel, N. (2013). Sensitivity and specificity of the online version of ImPACT in high school and collegiate athletes. *The American Journal of Sports Medicine, 41*(2), 321–326.

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*(1), 91–99.

Schatz, P., Neidzwski, K., Moser, R. S., & Karpf, R. (2010). Relationship between subjective test feedback provided by high-school athletes during computer-based assessment of baseline cognitive functioning and self-reported symptoms. *Archives of Clinical Neuropsychology, 25*(4), 285–292.

Schatz, P., Moser, R. S., Solomon, G. S., Ott, S. D., & Karpf, R. (2012). Prevalence of invalid computerized baseline neurocognitive test results in high school and collegiate athletes. *Journal of Athletic Training, 47*(3), 289–296.

Schmidt, J. D., Register-Mihalik, J. K., Mihalik, J. P., Kerr, Z. Y., & Guskiewicz, K. M. (2012). Identifying Impairments after concussion: normative data versus individualized baselines. *Medicine and Science in Sports and Exercise, 44*(9), 1621–1628.

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*(3), 243–251.

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.

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*(10), 1630–1635.

Woodard, J. L., & Rahman, A. (2012). The human-computer interface in computer-based concussion assessment. *Journal of Clnicial Sport Psychology, 6*, 385–408.