*Journal of Athletic Training* 2005;40(3):139–154
© by the National Athletic Trainers' Association, Inc
www.journalofathletictraining.org



# Is Neuropsychological Testing Useful in the Management of Sport-Related Concussion?

**Christopher Randolph*; Michael McCrea†; William B. Barr‡**

*Loyola University Medical Center, Maywood, IL; †Waukesha Memorial Hospital, Waukesha, WI, and Medical College of Wisconsin, Milwaukee, WI; ‡New York University School of Medicine, New York, NY

*Christopher Randolph, PhD, contributed to conception and design and drafting, critical revision, and final approval of the article. Michael McCrea, PhD, and William B. Barr, PhD, contributed to conception and design and critical revision and final approval of the article.*

*Address correspondence to Christopher Randolph, PhD, 1 East Erie, Suite 355, Chicago, IL 60613. Address e-mail to crandol@lumc.edu.*

***Objective:*** Neuropsychological (NP) testing has been used for several years as a way of detecting the effects of sport-related concussion in order to aid in return-to-play determinations. In addition to standard pencil-and-paper tests, computerized NP tests are being commercially marketed for this purpose to professional, collegiate, high school, and elementary school programs. However, a number of important questions regarding the clinical validity and utility of these tests remain unanswered, and these questions present serious challenges to the applicability of NP testing for the management of sport-related concussion. Our purpose is to outline the criteria that should be met in order to establish the utility of NP instruments as a tool in the management of sport-related concussion and to review the degree to which existing tests have met these criteria.

***Data Sources:*** A comprehensive literature review of MEDLINE and PsychLit from 1990 to 2004, including all prospective, controlled studies of NP testing in sport-related concussion.

***Data Synthesis:*** The effects of concussion on NP test performance are so subtle even during the acute phase of injury (1–3 days postinjury) that they often fail to reach statistical significance in group studies. Thus, this method may lack utility in individual decision making because of a lack of sensitivity. In addition, most of these tests fail to meet other psychometric criteria (eg, adequate reliability) necessary for this purpose. Finally, it is unclear that NP testing can detect impairment in players once concussion-related symptoms (eg, headache) have resolved. Because no current guideline for the management of sport-related concussion allows a symptomatic player to return to sport, the incremental utility of NP testing remains questionable.

***Conclusions/Recommendations:*** Despite the theoretic rationale for the use of NP testing in the management of sport-related concussion, no NP tests have met the necessary criteria to support a clinical application at this time. Additional research is necessary to establish the utility of these tests before they can be considered part of a routine standard of care, and concussion recovery should be monitored via the standard clinical examination and subjective symptom checklists until NP testing or other methods are proven effective for this purpose.

***Key Words:*** neurocognitive function, traumatic brain injury, athletic injury

The medical management of sport-related traumatic brain injury can be conceptualized as having 2 distinct components. The first involves the acute care management of the injured athlete at the time of injury; its primary purpose is to identify and treat any potential neurosurgical emergencies (eg, cerebral hemorrhage). This type of management is quite rarely necessary, as most sport-related brain injuries involve uncomplicated concussions that do not constitute acute neurologic emergencies.[1,2] The second component of sport-related concussion management is much more commonly required of team medical personnel. This involves monitoring the symptoms of concussion over time for the purposes of tracking recovery and making return-to-play decisions. In the case of very mild concussions, often referred to as ''ding'' injuries, recovery may be complete within a few minutes, allowing for return to play that day under most proposed guidelines and (often) obviating any further workup.

Many sport-related concussions, however, produce a number of subjective symptoms (eg, headaches, wooziness, changes in balance/coordination, memory impairment) that may last for days after the injury.[3] Although more than a dozen concussion rating scales compete with separate return-to-play guidelines, they are all in agreement that *athletes should be symptom free before returning to play*.[2,4–8] The various guidelines differ only in the factors involved in rating the severity of a concussion and in how long a player should be symptom free before returning to competition. A review of the relative merits of these various rating scales is beyond the scope of this paper, as is an extensive discussion of the rationale for ensuring a symptom-free waiting period before returning to competition. The primary concern, however, is that players may be at an elevated risk for repeat concussion during the symptomatic postconcussive period. Some evidence suggests that such a period of vulnerability exists[9] and that recovery after a second concussion may be more prolonged. A less well-validated concern is the risk of second-impact syndrome, or uncontrollable brain swelling, believed to be due to cerebrovascular congestion.[10] This can be a life-threatening condition, but it is extremely rare and the causative mechanism remains unclear. It has not been established that closely spaced concussions are

necessary to produce this syndrome, and the name may, therefore, be a misnomer[11]; in fact, this syndrome may be attributable to a genetic abnormality related to familial hemiplegic migraine.[12]

Although there is general agreement to date that athletes who suffer a concussion should be symptom free before returning to play, the risks of premature return to play remain poorly defined, and little consensus exists on exactly how to measure concussion-related symptoms or impairments. Various approaches have been employed to date, including the use of concussion symptom checklists that rely upon player self-report information,[8] the use of brief neurocognitive testing developed for sideline evaluations,[13] postural stability measurements,[14] and more extensive neuropsychological (NP) testing.[15–17] The latter form of testing, which typically involves a 20- to 30-minute battery of tests measuring attention, memory, and other cognitive functions, is the focus of this paper.

This type of NP testing, initially used in studies of college athletes,[18] has been routinely employed as part of sport-related concussion management programs in the National Football League and National Hockey League for several years, as well as in a large number of collegiate programs. The use of these batteries has proliferated rapidly. Testing was initially limited to pencil-and-paper batteries. Several computerized NP test batteries have been developed and are now being commercially marketed to athletic programs in high schools and colleges, including ImPACT (University of Pittsburgh, Pittsburgh, PA),[19] CogSport (CogState Ltd, Victoria, Australia),[20] and HeadMinder Concussion Resolution Index (HeadMinder, Inc, New York, NY).[21] Athletic trainers and other sports medicine clinicians typically lack a sufficient background in psychometrics to make an informed decision about the utility of such instruments, and no peer-reviewed guidelines exist for the selection of these instruments. Although neuropsychologists are trained in identifying psychometric criteria necessary for the implementation of a given instrument for the purpose of clinical assessment, athletic trainers may lack the services of an NP consultant to aid in such decision making. In addition, the application of neurocognitive testing to determine recovery from concussion has some unique characteristics, further underscoring the need for a review of the factors involved in decision making regarding instruments marketed for this purpose. This need was the impetus for this review, and the objectives of this paper are to (1) briefly review the literature substantiating the potential utility of NP testing in concussion management, (2) acquaint athletic trainers and other medical staff with the existing tests used for this purpose, (3) review criteria necessary to establish clinical validity and utility for any test battery proposed or marketed for this purpose, and (4) determine the degree to which existing tests have met these criteria.

## BACKGROUND, RATIONALE, AND TERMINOLOGY

In a standard clinical setting, NP assessment involves the administration of various tests of cognitive abilities (eg, memory, attention, language, visuospatial skills, etc), tests of psychological functioning (eg, personality inventories, psychiatric symptom scales), and some limited testing of sensory and motor functioning. The results of these tests are combined with information from other sources (clinical history, neuroimaging, laboratory data) to allow neuropsychologists to reach diagnostic conclusions regarding the presence and nature of various developmental and acquired disorders of the central nervous system. The NP assessment is an established method for detecting and quantifying residual cognitive or behavioral deficits that may ensue from traumatic brain injury (TBI).[22]

In all forms of TBI, cognitive impairments are typically most severe and easily detected in the acute/subacute postinjury phase of recovery, with continued improvement over time and eventual stabilization. Mild, uncomplicated (eg, no associated bleeding or swelling) TBIs, including concussions, are generally not expected to produce any detectable permanent impairments of cognitive functioning.[23,24] Even in mild concussion, however, transient impairments of cognition are measurable in the immediate postinjury phase of recovery. The types of cognitive impairments that are most consistently reported after concussion (and TBI in general) include deficits in memory, cognitive processing speed, and certain types of executive functions (typically, measures of verbal fluency or response inhibition).[25,26] Recently published data from sport concussion studies suggest that these impairments may be detectable in group studies for up to 5 to 7 days.[27]

Although clinical NP assessments, as noted above, involve psychological instruments to screen for conditions such as depression or somatoform tendencies that could be diagnostically important, these instruments are not typically employed (or indicated) in the management of sport-related concussion. This is because most athletes are motivated to return to play, and psychological/motivational factors with the potential to affect subjective symptoms or actual test performance are generally uncommon in this setting. Therefore, most NP protocols for this purpose consist of only *neurocognitive* measures (ie, specific tests of memory, attention, and other cognitive domains). For the purpose of this article, however, the terms *neurocognitive* and *neuropsychological* are used interchangeably.

### Sideline and Baseline Testing

Standardized neurocognitive tests have been employed in the management of sport-related concussion in two ways. The first approach is through the use of a brief measurement tool, designed for the sideline assessment of players after a concussion, for the purposes of quantifying the severity of impairment during the acute postinjury phase and, in conjunction with other clinical information, determining eligibility to return to play in the same game or practice session.[37–39] The most completely studied and well validated of these is the Standardized Assessment of Concussion (SAC) (CNS Inc, Waukesha, WI), which takes approximately 5 minutes to administer. Sensitivity, reliability, and change-score analyses of SAC data have been reasonably well explored.[13,40–42]

The SAC is a relatively cursory neurocognitive screening tool, however, with ceiling effects that potentially limit its usefulness in detecting subtle changes in neurocognitive functions due to concussive brain injury. The primary role of this type of instrument is as one component in decision making regarding same-day return to play. On the other hand, standard clinical NP evaluations typically require several hours to complete, and the length and nature of these evaluations are inappropriate for application in a sports medicine context. Finally, the relatively subtle nature of the cognitive impairments associated with concussion suggests that the best method for detecting these is to compare an injured player's performance with his or her preinjury baseline test scores. This approach

**Table 1. Traditional Pencil-and-Paper Neurocognitive Tests Used in Baseline Batteries**

| Test | Domain Measured | Description |
|---|---|---|
| Hopkins Verbal Learning Test[28] | Memory | This test consists of a 12-word list that is repeated over 4 learning trials for immediate recall. After a delay, free recall and recognition of the words are tested. |
| Brief Visuospatial Memory Test[29] | Memory | This is a visual learning test with 3 learning trials and a delayed free-recall trial. Subjects have to learn and reproduce 6 abstract designs in a 2-by-3 matrix. |
| WAIS-3 Digit Symbol subtest[30] | Processing speed | This WAIS-III subtest is a measure of cognitive processing speed. Subjects are asked to fill in a long series of boxes underneath numbers with symbols, using a key to identify which symbol goes with each number. The objective is to correctly fill in as many boxes as possible within the (2-minute) time limit. |
| Symbol Digit Modalities Test[31] | Processing speed | This test is similar to the WAIS-III Digit Symbol subtest described above. The only significant difference between the tests is that the location of the numbers and symbols is reversed. |
| Trail Making Test[32] | Processing speed, executive | This test requires subjects to search an array of circles that each contain a letter or number and to connect the circles in order by drawing lines between them. The score is the time required to complete the task. |
| Controlled Oral Word Association Test[33] | Processing speed, executive | Subjects are given a letter of the alphabet and asked to generate as many words as they can that start with that letter over a 60-minute period. The test typically involves 3 separate trials with different letters. |
| Stroop Color Word Test[34] | Executive | This is a measure of response inhibition. Subjects are given a page with columns of color names (red, green, blue). The names are printed in various colors, and subjects are asked to say the name of the color in which each word is printed, inhibitiing the natural tendency to read the words themselves. The test is timed and scored for the total number correct (typically within 45 seconds). |
| WAIS-III Digit Span Test[34] | Working memory | A subtest of the Wechsler Adult Intelligence Scale-III, this test involves repeating strings of numbers of increasing string length. This includes both forward span (exact repetition) and backward span (repeating the strings in reverse order). |
| WAIS-III Letter-Number Sequencing Test[30] | Working memory | This test is also from the WAIS-III, somewhat more demanding than simple digit span. Subjects are given strings of numbers and letters in random order and must rearrange them mentally, repeating the numbers in order and then the letters in order. The string lengths increase until the failure of all strings at a given length. |
| Paced Auditory Serial Addition Test[35] | Working memory, speed of processing | Subjects perform a series of mental arithmetic computations, with the stimuli delivered via audiotape. |
| Repeatable Battery for the Assessment of Neuropsychological Status[36] | Neurocognitive battery | A brief (20- to 25-minute) neurocognitive test battery with alternate forms, designed as a neurocognitive screening tool or stand-alone battery for the evaluation of dementia. It includes measures in 5 neurocognitive domains, including some domains not believed to be routinely affected by concussion (ie, language and visuospatial skills). |

is important as a means to control for individual baseline variation in cognitive abilities. In sport settings, the most common approach has been to employ a focused, 20- to 40-minute NP battery that is administered to obtain preseason baseline scores against which to compare postinjury performance. The widespread use of this method has resulted in the use of the term *baseline battery* to identify groups of tests used for this purpose and to distinguish this approach from sideline testing or traditional NP evaluations.[15]

## Existing Tests Available for Baseline Test Protocols

A variety of neurocognitive tests have been borrowed or adapted from the clinical NP armamentarium for the purpose of baseline testing. Table 1 lists the most common such tests, with a brief description of each measure. These tests all measure various aspects of memory (new learning), cognitive processing speed, working memory, or executive functions. The rationale for choosing tests from these domains is that these are the functions typically affected by traumatic brain injury, as opposed to functions such as language or visuospatial skills, which are more resistant to the effects of brain injury. The first large-scale study employing a set of pencil-and paper tests was completed in the early 1980s by Macciocchi et al[18] at the University of Virginia. This method was adopted by Lovell et al[16] in creating a battery for baseline testing of selected players with the Pittsburgh Steelers in 1993, although the battery was somewhat modified and expanded. Our group began baseline testing the entire roster of the Chicago Bears and members of the New York Jets in 1995 and 1996, using a battery largely overlapping the battery selected by Lovell et al. The National Hockey League adopted a league-wide baseline testing program using a somewhat different pencil-and paper battery in the late 1990s, although no data from this program have ever

**Table 2. Computerized Neurocognitive Tests Proposed for Baseline Batteries**

| Test | Description |
|---|---|
| ANAM[43] | This computerized battery originally arose from a Department of Defense program focused on determining the cognitive effects of countermeasures to chemical weapons, environmental stressors, and medications. Seven subtests are included, 6 or which have been employed in the investigation of concussions: Simple Reaction Time, Matching to Sample (visual working memory), Continuous Performance Test (sustained attention), Math Processing (processing speed and working memory), Spatial Processing (visual matching) and Sternberg Memory (verbal working memory). It takes 15 to 20 minutes to administer, depending on how many subtests are used, and employs a pseudo-randomization procedure to minimize practice effects on repeat testing. There is no overall index score. |
| CogSport[44] | This battery reportedly requires 15 to 20 minutes to complete. It contains measures of speed, accuracy, and consistency for responses within domains described as Psychomotor, Decision Making, Problem Solving, and Memory, although there is some inconsistency in the test descriptors used in the Web site versus publications. There appear to be 8 scores produced by the test and no overall index score. CogSport is one application from a group of computerized neurocognitive tests developed by the company CogStage Ltd. |
| HeadMinder CRI[21] | The Concussion Resolution Index, developed by HeadMinder, Inc. is a relatively new computerized test that utilizes a Web-based administration and scoring system. It can be administered from any computer with an Internet connnection and takes 20-25 minutes to complete. There are 6 subtests; one of these (Animal Decoding) is similar to the Digit Symbol/Symbol Digit tasks, in that subjects have to type in numbers, keyed to specific animals being presented. Symbol Scanning is a second processing speed subtest, similar to the WAIS-III Symbol Search subtest. There are also 2 reaction time subtests and 2 memory subtests. The latter 2 tests use a continuous-recognition type of format, in which subjects are presented with a series of pictures, some of which are repeated. They are instructed to press the spacebar whenever they recognize a previously presented picture. Three overall index scores are derived from the 6 subtests: Processing Speed, Simple Reaction Time, and Complex Reaction Time. There is no overall index score. |
| ImPACT[45] | This computerized test battery was created by investigators who were active in implementing pencil-and-paper neuropsychological testing in the National Football League and National Hockey League. It requires 20 to 25 minutes to complete. There are 6 subtests, each with multiple associated scores: Word Memory, Design Memory, X's and O's, Symbol Match, Color Match, and Three Letters. Five composite scores are reported: memory composite (verbal), memory composite (visual), visual motor speed composite, Reaction time composite, and Impulse control composite. Five composite scores are calculated for version 2 of the test, as opposed to 3 composite scores in version 1. There is no overall index score. A self-report symptom scale is also included in this program. |

been published. Most of the National Football League teams have adopted some type of baseline testing program over the last several years as well, although this is not a uniform endeavor, and the composition of the test batteries differs somewhat from team to team. To date, there has been no systematic exploration of the relative sensitivity of any of the constituent subtests of these batteries, nor has there been an attempt to create a uniform scaling system or composite battery score.

In addition, at least 4 computerized tests have been adapted for this purpose, 3 of which are commercially available (CogSport, HeadMinder CRI, and ImPACT) (Table 2). The potential benefits of computerized testing are that trained test administrators might not be needed, multiple subjects could potentially be tested simultaneously (if multiple computers were available), and reaction time data could be recorded.[46]

## Steps to Validating a Sport Concussion Neurocognitive Battery

As noted above, several steps are necessary to validate a neurocognitive battery for use in the management of sport-related concussion. They include the following: (1) establish test-retest reliability (stability); (2) establish sensitivity; (3) establish validity; (4) establish reliable change scores and an algorithm for classifying impairment; and (5) determine clinical utility (eg, detection of impairment in the absence of symptoms). Each of these 5 steps will be briefly explained. Athletic trainers or team physicians reviewing candidate batteries for use in their programs should ensure that each of these requirements has been met by any proposed battery before investing in it for routine clinical use.

**Reliability.** Different types of reliability are considered in the psychometric characteristics of a test, but in this context, we are most interested in establishing *test-retest reliability,* or the extent to which scores on the battery remain stable over time. This includes the exploration of any practice effects, or improvement in performance associated with repeat testing. In the management of sport-related concussion, one would ideally prefer a measure with high test-retest correlation across the different time intervals that are likely to be involved in the practical application of these measures, with minimal or no associated practice effects. Because the time interval between preseason baseline testing and the occurrence of a concussion may be several months in duration, merely demonstrating test-retest reliability over a very short interval (eg, days) is not adequate to establish reliability for this application. In addition, unless new baseline measures for each player are obtained on an annual basis, data regarding very long-term (ie, greater than 1 year) reliability should be established. Unless reliability is quite high, a test is unlikely to be useful for the purpose of individual decision making, and it has been suggested that a reliability of close to 0.9 is necessary for this purpose.[47] To illustrate this, intelligence quotient (IQ) scores have a normal mean of 100 and standard deviation of 15. With a reliability of 0.9, an individual would have to experience a decline of approximately 7 IQ points in order for one to conclude that this change represented evidence of an impairment (ie, a change of that magnitude would occur less than 10% of

the time by chance). With a reliability of only 0.7 (typical for many of the tests reviewed below), an individual would have to exhibit a decline of 20 IQ points to move out of that same confidence interval. This point is extremely critical for the purpose of establishing reliable change scores for the purpose of individual decision making.

**Sensitivity.** Establishing test sensitivity requires demonstrating that *the test can differentiate between clinical patients (in this case, athletes with concussion) and normal controls.* The most basic way of fulfilling this requirement is to find a statistical difference in a prospective study between the scores of players with concussion and appropriate matched controls on the test. This should be relatively easy to do for sport-related concussion, particularly if comparisons are made during the acute phase of recovery. Tests vary in terms of sensitivity, however, and virtually no head-to-head comparisons of different candidate tests have been performed that would allow a consumer to make a decision as to which test battery is most sensitive. In addition, a major difference exists between establishing sensitivity at the group level and establishing sensitivity for the purpose of individual clinical decision making. For example, in contrasting the group means (between players with concussion and those without concussion) on a neurocognitive test, relatively small differences may reach statistical significance, even if the group distributions are largely overlapping. Therefore, sensitivity in the strict sense may have been established, although sensitivity from a practical standpoint may still be inadequate (see ''Clinical Utility'' below). Sensitivity is often appropriately discussed in conjunction with specificity, which essentially refers to the rate of false positives: in this case, players who would be classified as impaired when they are actually normal. The unique methods employed in the baseline testing paradigm allow false-positive rates to be precisely determined and controlled, so this is not discussed separately but addressed in the ''Change Scores/Classification Rates'' section for each candidate set of tests. Theoretically, the best way to compare candidate batteries would be on the basis of correct classification rates, with false-positive rates held at equivalent levels for each battery, but these data are rarely reported, making such comparisons essentially impossible (see below).

**Validity.** Different types of validity exist, but the basic concept behind validity is *establishing that a test measures what it is purported to measure.* Validating a test of memory, for example, involves ensuring that the test actually does measure memory. In order to establish validity, most test developers explore (at a bare minimum) concurrent validity. In order to do this, they investigate the degree to which a test under development correlates with established measures of the ability in question. An experimental test of memory ought to correlate well with established measures of memory, for example (convergent validity), and correlate less well with measures of other types of cognitive functioning (divergent, or discriminant, validity). Another approach to establishing validity is to demonstrate that the test is sensitive to impairment in clinical populations with known deficits (eg, using amnesic patients who are demonstrably impaired on traditional tests of memory to validate a novel test of memory). This is less important in baseline testing than in standard clinical assessments, as baseline testing is primarily concerned with *detecting* neurocognitive impairment associated with concussion, rather than *characterizing* the nature of the impairment. This information is still useful, however, particularly in evaluating findings from empirical investigations (eg, determining which neurocognitive domains are most sensitive to the effects of concussion).

**Change Scores/Classification Rates.** No psychological test is perfectly stable over time, and there is always a certain degree of variability from one test session to the next. This test-retest variability is composed of a combination of true variance, or *real* normal fluctuations in cognitive functioning over time, and error variance, or changes in test scores attributable to flaws in the measurement technique employed. The overall variance in a test score can be quantified by using reliability data from test-retest studies in normal controls. This allows the derivation of confidence intervals around that particular score. These confidence intervals reflect the probability that a change in a given score is attributable to normal variation, as opposed to a statistically significant improvement or decline in performance. A number of approaches are available to calculate change scores, and the more sophisticated approaches can account for practice effects and the psychological phenomenon of regression to the mean, as well as error variance, in developing a Reliable Change Index, or RCI. This method was recently reviewed by Collie et al,[48] with a particular focus on the application of various RCI techniques to neurocognitive test data in concussion management. Determining the probability of score changes of different magnitudes over different time intervals is critical to the interpretation of neurocognitive retest data, and no NP test battery can be considered complete without providing this information. For example, presume that you are an athletic trainer and have a player who was concussed 5 days ago, is now asymptomatic, and undergoes repeat neurocognitive testing as a final step for clearance for return to play. You are told that the player's score on the battery is 5 points lower than it was during baseline testing 3 months ago. That particular piece of information by itself is of little value unless you know the probability associated with a change of that magnitude. It is possible, for example, that 60% of players could have a change score that of that magnitude or greater simply due to normal variation. If that were the case, the most probable explanation for your player's 5-point decline would be normal variability and not concussion-related deficit. It should also be noted that a single overall index score for performance is highly preferred in this context. Not only are summary index scores generally more reliable than subtest scores, but change-score probabilities are much more easily interpreted for a single score. Overall probability must be divided by the number of scores involved, because the more scores one analyzes, the higher the probability that one of them will be abnormal, or out of range, simply due to chance. For test batteries that provide users with multiple scores, probability estimates based on these multiple comparisons should be produced (eg, a Bonferroni-type correction). This is less ideal than having a single overall index score and may limit the sensitivity of the battery, particularly if the effect of concussion is to lower all scores slightly.

**Clinical Utility.** It is possible for a psychological test to be very sensitive and very reliable but still not be clinically useful. This would be true if, for example, the test was so long and difficult to administer that it would not be practical to use on a regular basis. For the purpose of concussion management, not only must a test satisfy all of the requirements listed above, but *the test should also be proven sensitive in detecting neurocognitive impairments once self-reported postconcussive symptoms have resolved.* Because no existing concussion-management guideline would allow a player to return to play while

symptomatic, neurocognitive testing would not alter clinical decision making while players are endorsing subjective symptoms. Finally, in order for a test to be clinically useful in this context, there should be a clear, probability-based rationale for identifying change scores (see above) as indicative of impairment and, thereby, providing an empirically supported algorithm for classifying a player as impaired on the basis of neurocognitive test performance. This simply involves setting specificity levels from empirically derived test-retest data from normal controls. The probability of obtaining change scores of different magnitudes due to normal variability can be directly determined this way, and cutoff scores can be provided.

### Criteria Met by Existing Tests

The section below outlines the degree to which existing neurocognitive tests used in the management of sport-related concussion have met the 5 criteria detailed above. The literature review for this purpose included MEDLINE and PsychLit searches from 1990 through 2004, a secondary search of all appropriate references from identified articles, and a review of data posted on the Web sites of the commercially available tests as of January 2004.

### Pencil-and-Paper Tests

**Reliability.** Most of the standard pencil-and-paper tests that have been published for clinical use include test-retest stability coefficients in the manual associated with the test. It is important to distinguish test-retest reliability from other measures of reliability (eg, internal consistency), as the former determines the utility of a test for individual decision making when interpreting change scores. It has been suggested that a .90 level of reliability is desirable for making decisions about individual change,[47] but individual tests rarely achieve this level of test-retest stability. For example, Barr[49] reported data from 12 NP tests administered to high school athletes at 60-day test-retest intervals, including data from the Hopkins Verbal Learning Test, Trail Making Test, Controlled Oral Word Association, and Wechsler Digit Span and Digit Symbol Tests. The stability coefficients ranged from .39 to .79, and the author concluded that caution needed to be exercised in interpreting change scores using these measures due to poor test-retest stability. In standard clinical practice, individual test scores are often combined into index scores, which are much more stable. For example, the test-retest stability for the WAIS-III Processing Speed Index ranges from .83 to .92 (depending on age),[30] and the test-retest stability for the Repeatable Battery for the Assessment of Neuropsychological Status total score ranges from .82 to .88.[36] Unfortunately, no attempt has been made to apply a systematic approach to combining multiple paper-and-pencil tests into a single index score for this purpose. A number of psychometric considerations are involved in this process; for a comprehensive recent review of the issues, see Frazier et al.[50]

**Sensitivity.** We reviewed the peer-reviewed literature for prospective controlled studies that included at least 10 concussed subjects. In 1996, Macciocchi et al[18] reported data from 183 collegiate football players who had sustained concussions, compared with 48 nonathlete student control subjects. Players were tested preseason, 24 hours postinjury, 5 days postinjury, and 10 days postinjury. The authors reported data from 3 tests (Paced Auditory Serial Addition Test, Trail Making Test, and Digit Symbol Test). There were 2 Paced Auditory Serial Addition Test scores and 2 Trail Making scores, for a total of 5 dependent neurocognitive measures. Both the control subjects and the concussed players exhibited significant improvement on these measures (as measured by the Hotelling $T^2$) at the 24-hour postinjury time point compared with preseason baseline testing. The magnitude of this improvement, however, was significantly less on 4 of the 5 variables for the concussed athletes than for the student controls. At the 5-day postinjury test session, this trend was reversed, and the players demonstrated greater gains from the previous test session, with no apparent group differences. Simple (raw) change scores were computed, but the variance associated with these was not reported. These data suggest some effect from concussion on 4 of 5 neurocognitive test scores at 24 hours postinjury but no effect at 5 days postinjury.

In the same year (1996), Maddocks and Saling[26] reported data from 10 professional Australian Rules Football players who sustained concussions, with a control group of 10 Australian Football League umpires matched for age and education. They were administered the Digit Symbol Test, Paced Auditory Serial Addition Test, and a choice reaction-time test (2 dependent variables), for a total of 4 dependent neurocognitive variables. Subjects were tested preseason and 5 days postinjury. No significant differences were noted between the groups at baseline, but for an unspecified reason, the postinjury data were analyzed via 4 separate analyses of covariance, using preseason performance as the covariate. No group differences were observed on the Paced Auditory Serial Addition Test or on 1 of the 2 dependent variables from the reaction time test (movement time). The analysis of covariance for the Digit Symbol Test comparison reached significance ($P = .04$), and the result for the other reaction time measure (decision time) was also significant ($P = .01$), indicating relatively worse performance by the concussed group. The authors did not control their overall alpha for group comparisons. The results suggest relatively subtle effects of concussion on certain processing speed measures at 5 days postinjury. One caveat to this study is that the concussions sustained seem to have been relatively severe, with 5 of 10 involving loss of consciousness. This is in contrast to larger, more recent prospective studies whose authors observed loss of consciousness in less than 10% of sport-related concussions,[27,51] suggesting that the sampling procedures may have selected for more severe concussion.

Hinton-Bayre et al[52] reported data in 1997 from 10 players concussed while playing professional rugby in Australia. This study involved a complicated design, with the administration of 2 baseline test sessions and a 1- to 2-week retest interval, but 3 of 10 concussed players were injured before undergoing a second baseline test session. Controls were also professional rugby players; as best can be determined from the methods, all controls underwent 2 baseline test sessions and underwent a third test session ''at various times midseason.'' There were 4 dependent variables: Digit Symbol Test, Speed of Comprehension, Symbol Digit Modalities Test, and Spot-the-Word Test. A significant practice effect was shown from the first to the second preseason baseline test session on 3 of the 4 tests (excluding Spot-the-Word). The postinjury test session occurred 24 to 48 hours postinjury. The authors carried out multiple comparisons, including comparing the postinjury scores with preinjury baseline session #2, preinjury baseline average score, and preinjury baseline maximum score. If parametric comparisons (paired *t* tests) were nonsignificant, nonparamet-

ric tests were calculated (Wilcoxon T). They reported worse performance in the injured group postinjury on at least 1 statistical measure for 3 of the 4 tests (no differences on Spot-the-Word), whereas the control group was essentially unchanged from baseline performance. The methods of this study were complicated by the fact that 30% of the concussed group did not obtain the benefit of a second baseline test session, reducing the comparability of the control group. In addition, overall alpha was not controlled for multiple comparisons, and no rationale was given for undertaking both parametric and nonparametric comparisons. In addition, no direct statistical comparison was made of the injured players to the controls.

Collins et al[53] reported data from a prospective study of college football players in the United States in 1999. Sixteen players who suffered concussions were compared with 10 controls (also collegiate football players). Eight NP tests were administered at baseline; within 24 hours postinjury; and on days 3, 5, and 7 postinjury. The only group comparison reported was a discriminant function analysis (presumably stepwise), attempting to discriminate injured players from controls using the immediate postinjury data only. This reportedly resulted in an 89.5% correct classification rate. No comparison was provided of the baseline performance of the 2 groups or of any other group comparison of performance at any other time point. There is, therefore, no way of determining the sensitivity of any of the constituent neurocognitive tests from this study, given the way the data were reported. Discriminant analyses can capitalize on chance differences, and it is possible that the reported level of discrimination could have been reached even without group differences on any of the 8 variables. In addition, it is not clear from this paper that the 2 groups performed at equivalent levels at baseline.

Echemendia et al[54] reported data in 2001 from a study of male and female athletes in various sports from a single university setting. A total of 29 athletes with concussion were compared with 20 controls matched for age, sex, ethnicity, and sport played. Test sessions occurred at baseline preseason; 2 hours postinjury; and 2, 7, and 30 days postinjury. Eight neurocognitive tests were administered, together with a symptom checklist, and a total of 23 separate dependent variables were examined. An abbreviated battery was given at the 2-hour postinjury time point, and only 12 dependent variables were obtained from that time point. The baseline performances of the 2 groups were compared using a multivariate analysis of covariance, with Scholastic Assessment Test score as a covariate, but all subsequent comparisons were performed using multiple analysis of variance only. The reason for this was not specified. The groups did not reportedly differ at baseline (per multivariate analysis of covariance), but significant group differences were identified at 2 hours and 2 days postinjury via multiple analysis of variance. No group differences were observed at 1 week. At the 2-day postinjury test point, 5 of the 23 dependent measures reached significance on univariate testing, including 3 measures from the Hopkins Verbal Learning Test, Digit Span: Digits Backward, and 1 measure from the Stroop Color Word Test. This appears to have been a well-designed study, although it is somewhat puzzling that the methods used for the baseline group comparison differed from those used to compare the 2 groups postinjury. The results are consistent with earlier studies, identifying significant differences on some percentage of the tests administered up to 2 days postinjury. These authors indicated in the discussion section of the paper that injured athletes did not differ from controls in subjective symptoms at 48 hours postinjury, although interestingly, the version of the symptom checklist they used was not specified and the data were not reported in the paper. This finding might suggest a unique role for NP data in identifying impairment in asymptomatic players. In more recent studies, however, symptoms as recorded by detail checklist have clearly been found to persist for longer than 48 hours,[27,55] again bringing into question the incremental utility of NP test data in clinical decision making. Moreover, the results of all test sessions after the immediate postinjury test session must be interpreted with caution, given the possibility that practice effects may differ between groups after that point (see below).

Guskiewicz et al[56] in 2001 reported a study of 36 concussed collegiate athletes compared with 36 controls using 4 neurocognitive tests (Hopkins Verbal Learning, Wechsler Digit Span, Stroop Color Word, and Trail Making), for a total of 7 dependent measures. Subjects were tested at baseline and then at 1, 3, and 7 days postinjury. The data were analyzed using individual, repeated-measures analyses of variance. Only 3 of the 7 dependent variables produced a significant group-by-time interaction effect, and only 2 of these followed a pattern consistent with recovery from concussion (Trail Making Part B and Digit Span: Digits Backward). The magnitude of the observed effects was small, and only 2 of 7 univariate comparisons on day 1 were significant (without controlling for overall alpha).

Field et al[55] recently (2003) reported data from 54 high school and collegiate athletes who underwent testing at baseline; within 24 hours postinjury; and on days 3, 5, and 7 postinjury. Five NP tests were used, with approximately half of the sample receiving a sixth test (Brief Visual-Spatial Memory Test-Revised). They reported data from only 2 memory tests given (Hopkins Verbal Learning Test and Brief Visual-Spatial Memory Test-Revised), with 2 scores from each memory test being analyzed. No baseline comparisons were reported between the 2 groups. The postinjury data were analyzed separately for the high school and college groups, using a multiple analysis of variance testing difference scores (raw score change from baseline). The subjects with concussion performed significantly worse than controls in the $<$24-hour test period. The college sample did not differ from the control group after that time. The concussed sample of high school players did continue to perform more poorly as measured by a significant multiple analysis of variance on days 3 and 7 but not (apparently) on day 5. The interpretation of these data is hampered somewhat by the fact that the dependent variables were selected from a much larger set of tests administered and that no attempt was made to control for (or even compare) baseline performance of the concussed and control groups.

Peterson et al[57] also recently reported the results of a study of collegiate athletes from various sports with concussion ($n = 28$) to athlete controls ($n = 18$) using 5 NP tests (Hopkins Verbal Learning, Trail Making, Symbol Digit Modalities, Digit Span, and Controlled Word Association). Eight variables from these tests were combined a priori into 6 conceptual domains. The baseline data from a pool of 350 participants was used to scale each individual variable to a standard score with a mean of 100 and standard deviation of 15. No transformations were reported for data with non-normal distributions, however, suggesting that the scaled score distributions would reflect the raw score distributions (ie, they might be skewed). Postinjury testing took place on days 1, 2, 3, and 10 after concussion. The

**Table 3. Effect Sizes for Pencil-and-Paper Tests**

| Test | Effect Size (range)* | Postinjury Test Interval (days) |
|---|---|---|
| Digit Symbol | 0.5–0.8 | 1–5 |
| Trail Making Part A | NS, 0.1–0.2† | 1–2 |
| Trail Making Part B | NS, <0.1–0.7 | 1–2 |
| Symbol Digit Modalities | NS to 0.9 | 1–2 |
| Stroop Color Word | NS, 0.6–0.7 | 1–2 |
| Controlled Word Association | NS to <0.1 | 2 |
| Digit Span | NS, 0.4–0.7 | 1–2 |
| Hopkins Verbal Learning: learning | NS, 0.5–2.0 | 1–2 |
| Hopkins Verbal Learning: delayed recall | 0.3–2.0 | 2 |
| Hopkins Verbal Learning: recognition | NS, 0.4–0.6 | 1–2 |
| Paced Auditory Serial Addition | NS, 0.3–0.6 | 1–5 |

*Effect sizes (Cohen d) may be not directly comparable due to differences in study design, and authors of some published studies did not report data in a format that allows for computation of effect size.
† NS indicates comparison between subjects with concussion and controls reported to be nonsignificant, without data available to calculate effect size.

data were analyzed using repeated-measures analysis of covariance, with baseline performance as the covariate. The only significant group-by-time interaction term occurred for the ''Speed of Information Processing'' domain, which consisted of the Trail Making Part B and Symbol Digit Modalities Tests. The groups were not found to differ at baseline on this domain score but did differ from each other at each postinjury test session, with the injured players performing worse. A standardized symptom checklist was also employed in this study, and injured players reported significantly higher symptom rates at each postinjury test date until the day 10 assessment.

Finally, McCrea et al[27] reported data from 94 concussed collegiate football players compared with 56 controls. Five neurocognitive tests were used (Hopkins Verbal Learning, Trail Making Part B, Symbol Digit Modalities, Stroop Color-Word Test, and Controlled Word Association), for a total of 7 dependent measures. Baseline testing and postinjury testing on days 2, 7, and 90 after concussion was obtained. The data were analyzed using a multivariate regression approach, controlling for baseline test performance. Direct statistical comparisons were not reported, however, as a descriptive approach to data presentation was employed. Some mild impairment was evident in the players with concussion at days 2 and 7 on the Stroop, Symbol Digit Modalities, and Controlled Word Association, but the data as presented make comparisons with previous studies difficult.

Overall, there is some evidence that standard pencil-and-paper tests are sensitive to the effects of concussion, at least within the first few days after concussion. However, the number of prospective studies is limited, and most suffer from methodologic flaws, typically involving a failure to control for multiple comparisons or to account for baseline performance, or both. In most of these studies, fewer than half of the dependent measures employed reached significance in postinjury group comparisons, and there is little consistency across studies regarding which variables are the most sensitive. Another potential confounding factor in the interpretation of data from many of these studies is that practice effects from the immediate postinjury test session may differ between concussed players and controls. That is, if concussed players are tested while obviously encephalopathic from the effects of the concussion, they are likely to benefit less from that test exposure than controls, who may benefit to a greater extent from that first postinjury test session. This factor complicates the interpretation of group data beyond the first postinjury test session. In some studies (eg, Peterson et al[57]), it is clear that the concussed players' performance dropped below controls at the first postinjury test session, then continued to improve in parallel with the control group as a result of ongoing practice effects. It is therefore difficult to conclude that the concussed group was continuing to exhibit a deficit indicative of brain dysfunction when they were improving at the same rate as controls on subsequent testing. Table 3 provides a summary of effect sizes for the different pencil-and-paper tests for the studies reviewed above, over test intervals ranging from 1 day to 5 days postinjury, although the differences in study design and reporting of results preclude direct comparisons or the computation of a meaningful meta-analysis.

**Validity.** For the most part, standardized NP pencil-and-paper tests have undergone extensive validation studies, including cross-validation with other NP tests and clinical validation studies involving patients with known patterns of impairment. For example, previous researchers[58] have demonstrated that scores on the Hopkins Verbal Learning Test correlate significantly with those obtained on other standard measures of verbal memory. Scores on this test are also sensitive to memory impairment in various clinical disorders involving compromise of memory functions.[59] It is beyond the purview of this paper to provide a comprehensive review of the validation studies involving traditional pencil-and-paper tests that have been adapted for the purpose of monitoring the effects of sport-related concussion. This is more of an issue when exploring the psychometric characteristics of novel computerized tests that have been developed for this purpose.

**Change Scores/Classification Rates.** The need to implement some rational method for the interpretation of change scores in this setting has been recognized for several years.[60] Barr[49] has provided reliable change scores for several individual pencil-and-paper NP tests from test-retest data in nonconcussed high school athletes with an 8-week retest interval. As noted above, however, the low reliability of these individual tests eventuates in rather large confidence intervals, even over a brief retest period. In 1999, Hinton-Bayre et al[61] reported the results of an RCI analysis on an expanded set of data from their earlier 1997 paper (see above) on concussion in Australian professional rugby players. In this study, 20 concussed players were evaluated using 3 dependent variables, with RCI intervals derived from a control group of 13 players. This was a rather complicated design (see above for details) involving a second baseline session for all controls but only a portion

of the concussed group. Using an alpha level of .05 (1 tailed), 80% of concussed players were classified as impaired within 1 to 3 days postinjury, but 23% of controls were also classified as impaired (false positives) at this point (overall correct classification of 79%). No authors to date have attempted to derive a composite score from paper-and-pencil NP tests or reported test-retest stability (or associated change scores) in a reasonably sized sample of nonconcussed athletes from clinically relevant intervals.

**Clinical Utility.** No researchers have demonstrated that pencil-and-paper NP tests can detect concussion once players are asymptomatic as measured by a standardized symptom checklist. This was suggested by Echemendia et al,[54] but they did not report which checklist they used, nor did they report data from their symptom checklist. In addition, they indicated that symptoms had returned to baseline levels within 48 hours, which is a more rapid resolution than has been reported in other studies.[27,55] Most of the standard pencil-and-paper test batteries used to date can be completed in less than 30 minutes but require one-to-one administration by a trained examiner. No consensus exists to date regarding which tests should be incorporated in a battery for this purpose, and no consistent approach to interpreting results has been promulgated.

## Computerized Tests

The section below reviews the available published data on the 4 computerized test batteries that have been reported useful in detecting the effects of sport-related concussion (Table 4). Three of these are commercially distributed; the Automated Neuropsychological Assessment Metrics (ANAM) (National Rehabilitation Hospital Assistive Technology and Neuroscience Center, Washington, DC) was developed by scientists working for the U.S. Government and is available free of charge. Sensitivity for each test was explored by reviewing data from published prospective controlled studies with a concussed number of at least 10, as was performed above for the pencil-and-paper tests.

## Automated Neuropsychological Assessment Metrics

**Reliability.** We could not identify any published data on measures of reliability for ANAM. Practice effects do occur on multiple measures within this battery.[62,63] No authors to date have explored the reliability of these measures across different computer platforms.

**Sensitivity.** As with the pencil-and-paper tests, we reviewed all published, prospective, controlled studies with 10 or more concussed subjects. Only 1 prospective controlled study has been published using ANAM to detect impairment after sport-related concussion.[63] This involved cadets at West Point, 64 of whom suffered boxing-related concussions, compared with 18 controls. All subjects were tested at baseline and at 4 postinjury intervals: 0–23 hours, 1–2 days, 3–7 days, and 8–14 days postinjury. Six dependent measures were included, and the data were analyzed using a mixed-model repeated-measures analysis of variance for each dependent variable separately. Only 1 of the 6 dependent variables yielded a significant group-by-time interaction (Spatial Processing subtest). On post hoc testing, group differences were identified on this subtest only for the first 2 postinjury test sessions (0–23 hours and 1–2 days postinjury). Apart from a significant group difference on 1 other subtest (Mathematical Processing) on the first postinjury test session, no other post hoc comparison (total number of comparisons = 24) reached significance. Overall alpha was not controlled. These data suggest that ANAM is generally lacking in sensitivity and that the utility for detecting individual impairment after concussion using this instrument is, therefore, questionable.

**Validity.** Two peer-reviewed articles[64,65] have explored the relationship between ANAM subtests and standard NP tests of cognitive processing speed, executive functions, and working memory. The results support the construct validity of the ANAM in the measurement of these cognitive domains.[64,65] A number of researchers[65–68] have demonstrated the sensitivity of ANAM to environmental, pharmacologic, and toxic stressors, as well as other diseases of the central nervous system, providing additional evidence of clinical validity in the application of this battery for the detection of diffuse mild encephalopathic conditions.

**Change Scores/Classification Rates.** No reports have been published to date regarding the derivation of a global score from the various ANAM subtests or any type of change scores.

**Clinical Utility.** No investigators have demonstrated that ANAM is sensitive to the effects of concussion once subjective symptoms have resolved. The length of the test is appropriate for this application.

## CogSport

**Reliability.** In a peer-reviewed article, 1 group[69] reported test-retest stability for CogSport over short time intervals (1 hour and 1 week). This was calculated using intraclass correlation coefficients, which are typically employed for subjective rating scales. The coefficients for 1-week retest stability for the 8 dependent variables (4 tests, each with measures of speed and accuracy) reported ranged from .31 to .82 (median = .6). This is relatively poor reliability, failing to meet suggested standards for individual decision making (see above). To our knowledge, no authors have reported reliability across different computer platforms.

**Sensitivity.** No published prospective controlled studies have demonstrated that the CogSport computerized test battery is sensitive to the effects of concussion.

**Validity.** In the same publication referenced in the section above,[69] the authors also reported data correlating the 8 dependent variables with scores on the Digit Symbol Test and the Trail Making Test Part B. They again employed intraclass correlation coefficients, with minimal to modest correlations between Trail Making and the speed measures from the CogSport battery (.23–.44) and modest to relatively strong correlations between the Digit Symbol and the CogSport speed measures (.42–.86). None of the CogSport accuracy measures were correlated with either traditional test. There was no attempt to explore divergent validity measures.

**Change Scores/Classification Rates.** There has been no attempt in any published work to derive a composite score from the CogSport battery, nor are there any published reports validating change-score methods with this battery. The authors of the battery have published on the statistical issues involved in deriving change scores for detecting the effects of concussion in general,[48] but to our knowledge, they have not provided an algorithm (or the data necessary to derive such an algorithm) for doing so with the CogSport battery. They have not presented data on the magnitude of practice effects for the

**Table 4. Current Status of Existing Tests with Respect to Necessary Criteria**

| Test | Reliability | Sensitivity | Validity | Change Scores | Clinical Utility |
|---|---|---|---|---|---|
| Paper-and-pencil | Reliability for most individual tests falls below the level desirable for individual decision making (.9), with stability coefficients typically ranging from .4 to .8. | Effect sizes from group studies are typically small, and many measues fail to reach statistical significance.<br>Group differences are usually detectable only within the first few days after concussion.<br>Studies rarely control for multiple comparisons. | Concurrent and clinical validity have typically been well-established for these tests in numerous studies. | No systematic approach to establishing a global score or change scores for multiple tests over clinically appropriate time periods has been reported.<br>Low reliability of individual tests results in large confidence intervals, diminishing sensitivity. | No study has been published showing that neurocognitive deficits can be identified after resolution of concussive symptoms.<br>No clear probability-based algorithm has been provided for identifying impairment. |
| ANAM | Reliability has not been formally reported, nor has it been studied/reported across different computer platforms. | Only one peer-reviewed study has been published that employed a prospective, controlled design. Only 1 of 6 variables yielded an overall group effect, without overall alpha control. Post hoc analyses also suggest limited sensitivity. | One peer-reviewed article has been published reporting concurrent validity data, with additional studies establishing clinical validity. | No attempt to establish a global score or derive any type of change score has been published to date. | No study showing that neurocognitive deficits can be identified after resolution of concussive symptoms has been published.<br>No clear probability-based algorithm has been provided for identifying impairment. |
| CogSport | One peer-reviewed paper reporting test-retest reliabilities over short time frames has been published. Stability coefficients were generally equivalent to individual pencil-and paper tests (.31 to .82). | There are no peer-reviewed, prospective, controlled studies demonstrating sensitivity to concussion. | One peer-reviewed paper demonstrated positive correlations with 2 independent measures of processing speed only. No exploration of clinical validity has been offered. | No attempt to establish a global score or derive any type of change score to date has been published. | No study showing that neurocognitive deficits can be identified after resolution of concussive symptoms has been published.<br>No clear probability-based algorithm has been provided for identifying impairment. |
| HeadMinder CRI | One peer-reviewed paper has been published with relatively short retest intervals (14 days) but a fairly large sample (n = 175) and adequate reliability (.68-.82), somewhat better than individual pencil-and-paper tests but below desirable levels for individual decision making. | There are no peer-reviewed prospective, controlled studies to date. Test sensitivity has been suggested by detection of impairment in some athletes based on normal test-retest changes of short duration from separately collected normative data. | One peer-reviewed paper has been published demonstrating expected concurrent validity with 7 standardized tests. | Carefully constructed exploration of reliable change-score method has been reported but only for short time intervals. Practice effect was identified on 1 of 3 indices. | No study showing that neurocognitive deficits can be identified after resolution of concussive symptoms has been published.<br>No clear probability-based algorithm has been provided for identifying impairment. |
| ImPACT | No peer-reviewed paper reporting reliability has been published. The test Web site reports reliability over short retest intervals. These range from .54 to .76 (version 1 of the test), below desirable levels for individual decision making. | Only one peer-reviewed, prospective, controlled study has been published. Only data from 1 of 3 composite scores were reported, and the control and concussed groups differed at baseline. No direct comparison of the groups was made postinjury. | No studies of concurrent or clinical validity have been published. | One study has been published using reliable change scores, derived from short retest intervals in controls: 41%–51% of athletes with concussion were identified as impaired <72 hours postconcussion (10% expected false-positive rate), but 54% of these players were symptomatic at the time. | No study has been published showing that neurocognitive deficits can be identified after resolution of concussive sysptoms.<br>No clear probability-based algorithm has been provided for identifying impairment. |

CogSport battery or on test-retest reliability for retest intervals greater than 1 week.

**Clinical Utility.** No authors have demonstrated that CogSport is sensitive to the effects of concussion once subjective symptoms have resolved. The length of the test appears to be appropriate.

### HeadMinder Concussion Resolution Index

**Reliability.** Two-week test-retest stability of the 3 index scores (see below) was reported to be .82 for processing speed (PS), .70 for sample reaction time (SRT) index, and .68 for complex reaction time (CRT).[70] These are somewhat below suggested levels for individual decision making but higher than for most individual pencil-and-paper tests.

Significant practice effects were observed only on the PS index for this retest interval. No longer-interval test-retest reliability data has been reported. To our knowledge, reliability across different computer platforms has not been explored or reported.

**Sensitivity.** No prospective, controlled studies have demonstrated that the HeadMinder CRI is sensitive to the effects of concussion. Sensitivity has been explored by using test-retest data from normal subjects (n = 175) tested 2 weeks apart and a subsample (n = 117) tested for a third time 1 to 2 days after the second test session.[70] The HeadMinder CRI has 6 subtests, several of which have both accuracy and speed measures (15 dependent variables in all). The data from these subtests were subjected to a factor analysis, and 4 factors were derived. The information from the factor analysis was used to help compose 3 index scores, which were employed for subsequent analyses. These include a PS index, an SRT index, and a CRT index. Two statistical models for deriving significant change scores were developed (RCI and regression-based change scores) from the normative retest data and applied to data from 26 athletes, primarily at the high school and college levels, who sustained a concussion during a single school year. The duration between baseline and postinjury testing for the concussed group was not reported. On average, concussed players underwent their initial postinjury testing less than 48 hours postinjury and a second postinjury testing approximately 6 days postinjury. The PS index did not appear to be sensitive to concussion at either time point, with classification rates similar to expected rates due to chance, regardless of the methods employed. The SRT classified 27% to 31% of concussed players (depending on the statistical model) as impaired on the first postinjury testing (5% false-positive rate expected), and the CRT classified 50% to 53% of concussed players as impaired on the first postinjury testing. Only 12% to 15% of concussed players were identified as impaired on the second postinjury test session. These data appear to have been separately analyzed for an earlier publication, in which error scores and subjective symptoms were also reported.[71] At the first assessment point, it seems that 89% of their subjects were endorsing subjective symptoms. Only 2 players were found to have impairment on neurocognitive testing in the absence of symptoms (not significantly different from chance). The interpretation of these data is complicated by the fact that they were not collected as part of a prospective controlled study, and the test-retest intervals between the control and injured subjects were obviously quite different. The data suggest that components of the HeadMinder CRI might in fact be sensitive to concussion, but this is difficult to determine from the existing data.

**Validity.** The authors referenced above[70] also reported concordant and divergent validity data for the 3 CRI index scores with 6 standard NP tests of working memory, processing speed, fine motor skill, and response inhibition. These correlational analyses generally support the construct validity of the CRI index scores. We are unaware of any additional clinical validity data.

**Change Scores/Classification Rates.** The authors have carefully explored 2 common methods for calculating change scores and applied these to an adequately sized normative sample. This occurred for a relatively short test-retest interval, however (maximum of 2 weeks), which is not adequate for routine clinical use, particularly because a practice effect was identified on 1 of the 3 index scores (see ''Reliability'' section above for details). No attempt has been made to derive a single composite score, which might increase reliability and eliminate the problem of controlling overall alpha.

**Clinical Utility.** No researchers have demonstrated that the HeadMinder CRI can detect the effects of concussion in a sample of players once subjective symptoms have resolved. The length of the test appears to be appropriate.

### ImPACT

**Reliability.** We were unable to identify any peer-reviewed paper reporting reliability data on ImPACT. The authors have posted reliability data for short retest intervals (maximum of 2 weeks) from a normative sample of 49 high school and collegiate athletes on their Web site. This was from version 1 of the test (the only version with sensitivity data from a published prospective controlled study). These were reported to be .54 for the Memory Composite, .76 for the Processing Speed Composite, and .63 for the Reaction Time Composite (for the 2-week interval). These stability coefficients are below suggested levels for individual decision making (see above) and generally commensurate with typical stability coefficients for individual paper-and-pencil tests.

**Sensitivity.** Only 1 peer-reviewed article involving a prospective controlled study with ImPACT has been published.[72] This involved 64 high school athletes who had suffered concussion, compared with 24 controls. Only data from the Memory Composite score were reported (3–5 composite scores make up the total, depending on the version of the test), and the 2 groups differed at baseline, with the concussion group performing below controls. Although the concussion group's Memory Composite score was reported to be significantly below their baseline at 1.5, 4, and 7 days postinjury, there were no direct postinjury comparisons between the control and concussed groups. The fact that the authors chose to report only a portion of the data available from this battery hampers interpretation of these data, as does the lack of comparable baseline performance between controls and injured players and the lack of any direct postinjury group comparisons.

**Validity.** We were unable to identify any reported data on concurrent or divergent validity studies with standardized NP tests or any reported clinical validity data with other patient groups.

**Change Scores/Classification Rates.** The authors have posted a preliminary RCI analysis from the reliability data described above on their Web site. As would be predicted given the low reliability of these scores, the 90% confidence in-

tervals for the composite scores are quite large. For example, a 90% confidence interval for the Memory Composite score would require a drop of nearly 13 points to reach a criterion of impaired. In reviewing the group data from the 1 controlled study that has been published on ImPACT,[72] the concussed group dropped only 8.3 points on average at the first postinjury test session (36 hours postinjury) on this measure; no other measures were reported. No attempt to derive a single global composite score to improve reliability has been reported. The authors also recently published an article examining reliable change scores on version 2.0 of the test.[73] These change-score intervals were derived from 56 young high school and college nonconcussed students, tested on average 6 days apart (range, 1–13 days), and applied to a sample of 41 high school and college athletes tested preseason and then retested within 72 hours of concussion. Using somewhat less conservative confidence intervals of 80%, the authors reported data for 4 of 5 composite scores (data for the Impulse Control Composite score were not reported). A significant practice effect was identified only on the Processing Speed Composite score. Between 41% and 51% of athletes were classified as impaired across the 4 composite scores (expected false-positive rate of 10%). A standardized concussion symptom scale was also administered, and 54% of concussed athletes were symptomatic on the basis of this scale at the time of the neurocognitive testing.

**Clinical Utility.** No studies have demonstrated that ImPACT is sensitive to the effects of concussion once subjective symptoms have resolved. The length of the test appears to be appropriate for this application.

## SUMMARY

Our objective was to focus on the use of NP test instruments in detecting and tracking concussion-related cognitive impairments as an aid in the management of athletes with concussion. The criteria necessary to justify the routine use of any such instrument were reviewed, along with the degree to which currently available instruments have been demonstrated to meet these criteria. Although some of these issues have been discussed in at least one prior review article,[60] a number of commercially available batteries have become available in the interim, and we undertook the present article as a comprehensive update on the state-of-the-art of NP testing for the management of sport-related concussion.

Unfortunately, no existing conventional or computerized NP batteries proposed for use in the assessment and management of sport-related concussion have met all of the criteria necessary to warrant routine clinical application. Therefore, important questions regarding the validity, reliability, and clinical utility of these instruments remain unanswered. As a result, test-retest data from any of these instruments are difficult to interpret, and any such interpretation must rely far more heavily upon clinical judgment than statistical algorithms. Given these facts, additional research is clearly necessary before NP testing can be considered a component of the routine standard of care in the management of sport-related concussion, particularly as the risks of premature return to play remain poorly defined.

However, NP testing is a reliable, objective method for evaluating the effects of central nervous system injury and disease, including mild TBI. A substantial amount of additional research will be required, however, in order for any of the proposed batteries to meet the necessary criteria for this purpose. This should include the following:

- Establishing test-retest reliability over time intervals that are practical for this clinical purpose. Because baseline testing is likely to precede postinjury testing by a period of weeks to months (or even years), test-retest reliability should be established for all applicable retest time periods.
- Demonstrating, through a prospective controlled study, that the battery is sensitive in detecting the effects of concussion.
- Establishing validity for any novel test battery, through standard psychometric procedures employed to determine which neurocognitive abilities a new NP test is measuring.
- Deriving reliable change scores, with a probability-based classification algorithm for deciding that a decline of a certain magnitude is attributable to the effects of concussion, rather than random test variance. In addition, for tests producing multiple scores, probability should be adjusted appropriately for the number of scores generated.
- Demonstrating that the proposed battery is capable of detecting cognitive impairment once subjective symptoms have resolved. This should occur through a controlled prospective study, tracking symptoms through the use of a detailed symptom checklist, with NP testing implemented once symptoms have resolved. Unless an NP battery is capable of detecting impairment after subjective symptom resolution, it cannot alter clinical decision making under any of the current management guidelines. Meeting this criterion would also satisfy the criterion of sensitivity.

Until additional research is completed in order to satisfy these criteria, NP testing for the purpose of managing sport-related concussion should be interpreted conservatively, given the limitations of the existing data, or be limited to research purposes. Athletic trainers are urged to exercise caution in the implementation of any NP testing protocol in their approach to the management of sport-related concussion until this method is more firmly established through the necessary empirical research. A number of authors have reported data indicating that subjective symptoms, documented through the use of a standardized symptom checklist, are evident for a period of time as long (or longer) postinjury than detectable NP impairments.[27,57,71,73] Given these data, the use of a standardized symptom checklist in addition to routine clinical examination is suggested as a reasonable approach to monitoring recovery from sport-related concussion, until the incremental utility of NP testing (or other methods) can be established.

## REFERENCES


1. Mueller FO. Catastrophic head injuries in high school and collegiate sports. *J Athl Train.* 2001;36:312–315.
2. Powell JW, Barber-Foss KD. Traumatic brain injury in high school athletes. *JAMA.* 1999;282:958–963.
3. Piland SG, Motl RW, Ferrara MS, Peterson CL. Evidence for the factorial and construct validity of a self-report concussion symptoms scale. *J Athl Train.* 2003;38:104–112.
4. LeBlanc KE. Concussion in sport: diagnosis, management, return to competition. *Compr Ther.* 1999;25:39–44.
5. Cantu RC. Head injuries in sports. *Br J Sports Med.* 1996;30:289–296.
6. Warren WL Jr, Bailes JE. On the field evaluation of athletic head injuries. *Clin Sports Med.* 1998;17:13–26.
7. Sturmi JE, Smith C, Lombardo JA. Mild brain trauma in sports: diagnosis and treatment guidelines. *Sports Med.* 1998;25:351–358.
8. Concussion Sport Group. Summary and agreement statement of the First

International Conference on Concussion in Sport, Vienna 2001. *Physician Sportsmed.* 2002;30(2):57–63.
9. Guskiewicz KM, McCrea M, Marshall SW, et al. Cumulative effects associated with recurrent concussion in collegiate football players: the NCAA Concussion Study. *JAMA.* 2003;290:2549–2555.
10. Cantu RC. Second-impact syndrome. *Clin Sports Med.* 1998;17:37–44.
11. McCrory PR, Berkovic SF. Second impact syndrome. *Neurology.* 1998; 50:677–683.
12. Kors EE, Terwindt GM, Vermeulen FL, et al. Delayed cerebral edema and fatal coma after minor head trauma: role of the CACNA1A calcium channel subunit gene and relationship with familial hemiplegic migraine. *Ann Neurol.* 2001;49:753–760.
13. McCrea M, Kelly JP, Kluge J, Ackley B, Randolph C. Standardized assessment of concussion in football players. *Neurology.* 1997;48:586–588.
14. Guskiewicz KM. Postural stability assessment following concussion: one piece of the puzzle. *Clin J Sport Med.* 2001;11:182–189.
15. Randolph C. Implementation of neuropsychological testing models for the high school, collegiate, and professional sport settings. *J Athl Train.* 2001; 36:288–296.
16. Lovell MR, Collins MW. Neuropsychological assessment of the college football player. *J Head Trauma Rehabil.* 1998;13:9–26.
17. Echemendia RJ, Julian LJ. Mild traumatic brain injury in sports: neuropsychology's contribution to a developing field. *Neuropsychol Rev.* 2001; 11:69–88.
18. Macciocchi SN, Barth JT, Alves W, Rimel RW, Jane JA. Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery.* 1996;39:510–514.
19. ImPACT. Available at: http://www.impacttest.com. Accessed February 2004.
20. CogSport. Available at: http://www.cogsport.com. Accessed February 2004.
21. *Concussion Resolution Index* [computer program]. New York, NY: HeadMinder, Inc; 1999.
22. Levin HS. A guide to clinical neuropsychological testing. *Arch Neurol.* 1994;51:854–859.
23. Dikmen S, Machamer J, Temkin N. Mild head injury: facts and artifacts. *J Clin Exp Neuropsychol.* 2001;23:729–738.
24. Ponsford J, Willmott C, Rothwell A, et al. Factors influencing outcome following mild traumatic brain injury in adults. *J Int Neuropsychol Soc.* 2000;6:568–579.
25. Barth JT, Alves WM, Ryan TV, et al. Mild head injury in sports: neuropsychological sequelae and recovery of function. In: Levin HS, Eisenberg HM, Benton AL, eds. *Mild Head Injury*. New York, NY: Oxford; 1989:257–275.
26. Maddocks DL, Saling MM. Neuropsychological deficits following concussion. *Brain Inj.* 1996;10:99–103.
27. McCrea M, Guskiewicz KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study. *JAMA.* 2003;290:2556–2563.
28. Shapiro AM, Benedict RH, Schretlen D, Brandt J. Construct and current validity of the Hopkins Verbal Learning Test-revised. *Clin Neuropsychol.* 1999;13:348–358.
29. Benedict RH. *Brief Visuospatial Memory Test-Revised.* Odessa, FL: Psychological Assessment Resources; 1997.
30. Wechsler D. *Wechsler Adult IntelligenceScale.* 3rd ed. San Antonio, TX: The Psychological Corp; 1997.
31. Smith A. *Symbol Digit Modalities Test.* Los Angeles, CA: Western Psychological Services; 1991.
32. Reitan RM, Wolfson D. *The Halstead-Reitan Neuropsychological Test Battery.* Tucson, AZ: Neuropsychology Press; 1985.
33. Benton AL, Hamsher K, Sivan AB. *Multilingual Aphasia Examination.* Iowa City, IA: AJA Assoc; 1983.
34. Golden JC. *Stroop Color and Word Test.* Chicago, IL: Stoelting Co; 1978.
35. Gronwall DM. Paced auditory serial-addition task: a measure of recovery from concussion. *Percept Mot Skills.* 1977;44:367–373.
36. Randolph C. *Repeatable Battery for the Assessment of Neuropsychological Status (RBANS).* San Antonio, TX: The Psychological Corp; 1998.
37. Cantu RC. Return to play guidelines after a head injury. *Clin Sports Med.* 1998;17:45–60.
38. Practice parameter: the management of concussion in sports (summary statement). Report of the Quality Standards Subcommittee. *Neurology.* 1997;48:581–585.
39. Kelly JP, Nichols JS, Filley CM, Lillehei KO, Rubinstein D, Kleinschmidt-DeMasters BK. Concussion in sports: guidelines for the prevention of catastrophic outcome. 1991;266:2867–2869.
40. McCrea M, Kelly JP, Randolph C, et al. Standardized assessment of concussion (SAC): on-site mental status evaluation of the athlete. *J Head Trauma Rehabil.* 1998;13:27–35.
41. McCrea M. Standardized mental status testing on the sideline after sport-related concussion. *J Athl Train.* 2001;36:274–279.
42. Nassiri JD, Daniel JC, Wilckens J, Land BC. The implementation and use of the Standardized Assessment of Concussion at the U.S. Naval Academy. *Mil Med.* 2002;167:873–876.
43. Reeves D, Thorne R, Winter S, Hegge F. *Cognitive Performance Assessment Battery (UTC-PAB).* San Diego, CA: Naval Aerospace Medical Research Laboratory and Walter Reed Army Institute of Research; 1989. Report 89-1.
44. Makdissi M, Collie A, Maruff P, et al. Computerised cognitive assessment of concussed Australian Rules footballers. *Br J Sports Med.* 2001;35:354–360.
45. Lovell MR, Collins MW. *ImPACT: Immediate Post-Concussion Assessment and Cognitive Testing.* Pittsburgh, PA: Neurohealth Systems, LLC; 1998.
46. Collie A, Maruff P. Computerised neuropsychological testing. *Br J Sports Med.* 2003;37:2–3.
47. Nunnally JC, Bernstein IH. *Psychometric Theory.* 3rd ed. New York, NY: McGraw-Hill, Inc; 1994.
48. Collie A, Maruff P, Makdissi M, McStephen M, Darby DG, McCrory P. Statistical procedures for determining the extent of cognitive change following concussion. *Br J Sports Med.* 2004;38:273–278.
49. Barr WB. Neuropsychological testing of high school athletes: preliminary norms and test-retest indices. *Arch Clin Neuropsychol.* 2003;18:91–101.
50. Frazier TW, Youngstrom EA, Chelune GJ, Naugle RI, Lineweaver TT. Increasing the reliability of ipsative interpretations in neuropsychology: a comparison of reliable components analysis and other factor analytic methods. *J Int Neuropsychol Soc.* 2004;10:578–589.
51. McCrea M, Kelly JP, Randolph C, Cisler R, Berger L. Immediate neurocognitive effects of concussion. *Neurosurgery.* 2002;50:1032–1042.
52. Hinton-Bayre AD, Geffen GM, McFarland KA. Mild head injury and speed of information processing: a prospective study of professional rugby league players. *J Clin Exp Neuropsychol.* 1997;19:275–289.
53. Collins MW, Grindel SH, Lovell MR, et al. Relationship between concussion and neuropsychological performance in college football players. *JAMA.* 1999;282:964–970.
54. Echemendia RL, Putukian M, Mackin RS, Julian LJ, Shoss N. Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. *Clin J Sports Med.* 2001;11:23–31.
55. Field M, Collins MW, Lovell MR, Maroon JC. Does age play a role in recovery from sports-related concussion? A comparison of high school and collegiate athletes. *J Pediatr.* 2003;142:546–553.
56. Guskiewicz KM, Ross SE, Marshall SW. Postural stability and neuropsychological deficits after concussion in collegiate athletes. *J Athl Train.* 2001;36:263–273.
57. Peterson CL, Ferrara MS, Mrazik M, Piland S, Elliot R. Evaluation of neuropsychological domain scores and postural stability following cerebral concussion in sports. *Clin J Sports Med.* 2003;13:230–237.
58. Lacritz LH, Cullum CM. The Hopkins Verbal Learning Test and CVLT: a preliminary comparison. *Arch Clin Neuropsychol.* 1998;13:623–628.
59. Brandt J, Corwin J, Krafft L. Is verbal recognition memory really different in Huntington's and Alzheimer's disease? *J Clin Exp Neuropsychol.* 1992;14:773–784.
60. Barr WB. Methodologic issues in neuropsychological testing. *J Athl Train.* 2001;36:297–302.
61. Hinton-Bayre AD, Geffen GM, Geffen LB, McFarland KA, Friis P. Concussion in contact sports: reliable change indices of impairment and recovery. *J Clin Exp Neuropsychol.* 1999;21:70–86.
62. Daniel JC, Olesniewicz MH, Reeves DL, et al. Repeated measures of cognitive processing efficiency in adolescent athletes: implications for

monitoring recovery from concussion. *Neuropsychiatry Neuropsychol Behav Neurol.* 1999;12:167–169.
63. Bleiberg J, Cernich AN, Cameron K, et al. Duration of cognitive impairment after sports concussion. *Neurosurgery.* 2004;54:1073–1080.
64. Bleiberg J, Kane RL, Reeves DL, Garmoe WS, Halpern E. Factor analysis of computerized and traditional tests used in mild brain injury research. *Clin Neuropsychol.* 2000;14:287–294.
65. Gottschalk LA, Bechtel RJ, Maguire GA, et al. Computerized measurement of cognitive impairment and associated neuropsychiatric dimensions. *Compr Psychiatry.* 2000;41:326–333.
66. Kabat MH, Kane RL, Jefferson AL, DiPino RK. Construct validity of selected Automated Neuropsychological Assessment Metrics (ANAM) battery measures. *Clin Neuropsychol.* 2001;15:498–507.
67. Rahill AA, Weiss B, Morrow PE, et al. Human performance during exposure to toluene. *Aviat Space Environ Med.* 1996;67:640–647.
68. Wilken JA, Kane R, Sullivan CL, et al. The utility of computerized neuropsychological assessment of cognitive dysfunction in patients with relapsing-remitting multiple sclerosis. *Mult Scler.* 2003;9:119–127.
69. Collie A, Maruff P, Makdissi M, McCrory P, McStephen M, Darby D. CogSport: reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clin J Sports Med.* 2003;13:28–32.
70. Erlanger DM, Feldman DJ, Kutner K, et al. Development and validation of a web-based neuropsychological test protocol for sports-related return-to-play decision-making. *Arch Clin Neuropsychol.* 2003;18:293–316.
71. Erlanger DM, Saliba E, Barth JT, Almquist J, Webright W, Freeman J. Monitoring resolution of postconcussion symptoms in athletes: preliminary results of a Web-based neuropsychological test protocol. *J Athl Train.* 2001;36:280–287.
72. Lovell MR, Collins MW, Iverson GL, et al. Recovery from mild concussion in high school athletes. *J Neurosurg.* 2003;98:296–301.
73. Iverson GL, Lovell MR, Collins MW. Interpreting change scores on ImPACT following sport concussion. *Clin Neuropsychol.* 2003;17:460–467.

## COMMENTARY

### Stephen N. Macciocchi, PhD, ABPP

*Editor's Note:* Stephen N. Macciocchi, PhD, ABPP, is the Director, Rehabilitation Psychology and Neuropsychology, Shepherd Center, Atlanta, GA.

The authors address a topic of importance for athletic trainers and other health care professionals involved in the management of cerebral concussions in athletes. A brief discussion of test development and validation is presented for readers who are not familiar with psychometrics. The authors review paper-and-pencil neuropsychological tests in a reasonably comprehensive manner, which allows readers to appreciate instruments currently available. The authors also provide a nice introduction to the relationship between test reliability and sensitivity but offer readers only one reference for reliability of commonly used paper-and-pencil tests.[1] The authors criticize existing research in several respects: most notably, investigators' historical failure to calculate reliable change indices (RCIs) for tests used in concussion research and investigators' failure to develop psychometrically coherent test composite indexes. Although concerns about paper-and-pencil test psychometrics are factually correct, the authors seem to expect much of some investigators, who were most likely examining the effects of concussion in an exploratory fashion rather than attempting to construct psychometrically sound instruments for making return-to-play decisions.

The authors also review currently used computerized tests. Some of these instruments were reported to have limited research support, whereas others have been more extensively examined. The authors summarize the strengths and weaknesses of the computerized test batteries based on their criteria (see Table 4). According to the authors, none of the instruments reviewed has met the psychometric criteria necessary to ''warrant routine clinical application.'' Test developers are likely to take umbrage at such assertions, but issues of standard error of measurement and RCI are very important for clinicians using neuropsychological tests. The authors also provide several reasonable recommendations regarding general test development but stipulate that test batteries must be ''capable of detecting cognitive impairment once subjective symptoms have resolved,'' a requirement that may be unrealistic given the variability in clinical response to concussive injury.

The authors' review of neuropsychological tests is valuable in a number of respects, which are directly or indirectly alluded to in the article. First, discussing psychometric considerations and test limitations should help to educate persons unfamiliar with these issues. Second, the authors have identified a number of primary research needs, which are important for developing instruments used in making return-to-play decisions as well as for investigators seeking a better understanding of the consequences of concussive injuries. Third, the authors have begun a process of critical evaluation, which can be difficult in our current, competitive scientific environment.

Although the authors should be applauded for their critical appraisal and recommendations, several issues not discussed by the authors merit mention. First, the authors fail to distinguish between studies seeking to validate a commercially developed battery[2–6] and exploratory studies with completely different goals.[7–10] Both types of studies deserve empirical scrutiny, just not using the same criteria. Second, the authors conclude that neuropsychological tests used in concussion assessment are unreliable and questionably valid and then opine that ''neuropsychological testing is a reliable, objective method for evaluating the effects of central nervous system injury and disease, including mild TBI.'' Despite the authors' attempt to casually extol the virtues of neuropsychological testing, the same concerns that apply to concussion testing are also concerns for neuropsychological testing in general. Unfortunately, the authors do not address what makes concussion assessment different from assessing other forms of ''central nervous system injury.'' Reliability and SEM are always a concern and vary depending on the test in question.[11] Detecting very mild, transitory deficits in select neurocognitive networks may make concussion testing somewhat different than other clinical assessments, but many of the problems discussed by the authors are operative in all neuropsychological examinations.

Third, based on the data presented, the authors recommend interpreting neuropsychological tests ''conservatively.'' This recommendation is prudent, but the authors fail to give readers any practical guidelines for neuropsychological test interpretation, especially important because these tests seem to be widely applied in concussion management. Fourth, the authors recommend relying on symptom checklists but fail to mention that, in some cases, athletes minimize or intentionally deny symptoms when they are, in fact, symptomatic. Finally, as previously mentioned, the authors conclude that cognitive deficits resolve before self-reported symptoms resolve, which is the most common recovery pattern identified by group studies but is by no means observed in all athletes.

From a clinical point of view, athletic trainers are faced with concussion management decisions that are rarely based on a single test or decision point. The athletic trainer must consider history, physical and neurologic examinations, imaging studies (if relevant), neuropsychological test findings, and self-reported symptoms, as well as the athlete's motivation and response to injury. In most cases, athletes do not evidence protracted cognitive and physical symptoms.[9] Nonetheless, sometimes players have no cognitive deficits but report significant postconcussive symptoms (eg, headaches, dizziness, and memory complaints). In other cases, players deny symptoms, but evidence impaired test performance, even by ''conservative'' interpretive standards. The authors recommend using symptom checklists as a primary decision-making tool in part because return-to-play algorithms require symptom resolution. Despite the management value of symptom checklists, an athlete's denial of symptoms may be credible or implausible, but assuming that self-reported symptom resolution is valid in every case is problematic. When setting thresholds for decision-making errors (false positive versus false negative), athletic trainers should consider the problems with RCI noted by the authors, but in all cases, athletic trainers should evaluate the entire set of clinical, historical, and test data available and not rely on any single indicator for return-to-play decisions. Although the problems discussed by the authors merit serious attention, the use of neuropsychological data may help clinical decision making in some cases but not in others. Given the stated need for additional research, completely avoiding the use of neuropsychological tests in clinical practice may have the effect of preventing exploration of the very concerns identified in the current article.

## REFERENCES


1. Barr WB. Neuropsychological testing of high school athletes: preliminary and test-retest indices. *Arch Clin Neuropsychol.* 2003;18:91–101.
2. Erlanger DM, Saliba E, Barth JT, Almquist J, Webright W, Freeman J. Monitoring resolution of postconcussion symptoms in athletes: preliminary results of a Web-based neuropsychological test protocol. *J Athl Train.* 2001;36:280–287.
3. Erlanger DM, Feldman DJ, Kutner K, et al. Development and validation of a web-based neuropsychological test protocol for sports-related return-to-play decision-making. *Arch Clin Neuropsychol.* 2003;18:293–316.
4. Iverson GL, Lovell MR, Collins MW. Interpreting change scores on ImPACT following sport concussion. *Clin Neuropsychol.* 2003;17:460–467.
5. Lovell MR, Collins MW, Iverson GL, et al. Recovery from mild concussion in high school athletes. *J Neurosurg.* 2003;98:296–301.
6. Kabat MH, Kane RL, Jefferson AL, DiPino RK. Construct validity of selected Automated Neuropsychological Assessment Metrics (ANAM) battery measures. *Clin Neuropsychol.* 2001;15:498–507.
7. Hinton-Bayre AD, Geffen GM, McFarland KA. Mild head injury and speed of information processing: a prospective study of professional rugby league players. *J Clin Exp Neuropsychol.* 1997;19:275–289.
8. Macciocchi SN, Barth JT, Alves W, Rimel RW, Jane JA. Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery.* 1996;39:510–514.
9. McCrea M, Guskiewicz, KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study. *JAMA.* 2003;290:2556–2563.
10. Ponsford J, Willmott C, Rothwell A, et al. Factors influencing outcome following mild traumatic brain injury in adults. *J Int Neuropsychol Soc.* 2000;6:568–579.
11. Macciocchi SN, Barth JT. Methodological concerns in traumatic brain injury. In: Lovell MR, Echemendia RJ, Barth JT, Collins, MW, eds. *Traumatic Brain Injury in Sports: An International Neuropsychological Perspective.* Lisse, The Netherlands: Swets and Zeitlinger; 2004:281–298.


## AUTHORS' RESPONSE

We appreciate Dr. Macciocchi's comments and the concerns he raises regarding our article. He suggests that we should have distinguished between exploratory studies of the use of neuropsychological testing in sport-related concussion and studies attempting to validate computerized batteries intended for commercial distribution. He also questions how concerns regarding the application of neuropsychological testing in the management of sport-related concussion can be distinguished from similar concerns in routine clinical neuropsychological assessment. He notes that we ''fail to mention that, in some cases, athletes minimize or intentionally deny symptoms'' and suggests that it might not be appropriate to rely upon symptom checklists as a result. He also suggests that requiring tests to be capable of detecting impairment in players once they are otherwise asymptomatic may be ''unrealistic'' and worries that ''completely avoiding the use of neuropsychological tests in clinical practice may have the effect of preventing exploration of the very concerns identified in the current article.''

As neuropsychologists, we have been concerned with these issues as well, and we have devoted a good deal of time and research into addressing these issues over the years. We have the following responses:

First, distinguishing between ''exploratory studies'' and commercially driven validation studies of neuropsychological tests. In our opinion, this is an unnecessary distinction. The issue at hand is whether or not *any* neuropsychological test battery can reliably identify ''impairment'' in concussed players. The ultimate psychometric requirements of any such battery are the same, regardless of the provenance of the battery. The fact that some of the exploratory studies were not specifically designed to allow for the derivation of change scores does not preclude a review of their findings with respect to sensitivity and reliability. We see no reason to adopt different standards for our review of these instruments.

Second, assessment of sport-related concussion versus clinical neuropsychological assessment. We do not think these are comparable endeavors, for the most part. There are a limited number of situations in which neuropsychologists must address change scores in clinical practice. Neuropsychologists are typically engaged in performing much more comprehensive evaluations than the testing that occurs in the sport setting, without the benefit of preinjury baseline testing. These clinical evaluations are carried out for various types of differential diagnostic and treatment planning purposes. With regard to the importance of change scores in clinical practice (eg, detecting postsurgical change, tracking decline in dementia), studies over the past decade or so have detailed the very issues we bring up here and provided change-score information for commonly used tests (eg, WAIS-3, RBANS). In contrast, the potential role for neurocognitive testing in the management of sport-related concussion is constrained simply to the *detection* of the residual effects of the injury, to ensure complete recovery before return to play. This boils down to a simple mathematical algorithm, and clinical judgment should not be required if the tests meet the necessary criteria for this purpose.

Third, the problem of athletes minimizing symptoms. This is a concern that is often raised to champion the use of objective measures. We have pointed out, however, that when subjective symptom reporting is systematically obtained in prospective studies, players typically report symptoms for at least

as long as impairment can be detected using neurocognitive tests (or other objective measures such as balance testing[1–4]). This does not preclude the possibility that any given athlete will deny or minimize symptoms, but it does suggest that a systematic inventory of subjective symptoms is an appropriate first step in monitoring concussion recovery. Symptom checklists are also the fastest, simplest, and most economical way of monitoring recovery. Any more time-consuming, complicated, or expensive method (eg, neurocognitive testing) should demonstrate that it provides demonstrable value-added information in terms of detecting impairment once players are symptom free.

Fourth, waiting until symptoms resolve before testing is ''unrealistic.'' We do not understand this concern. If a player is symptomatic as documented by a symptom checklist, no current guideline would permit return to play. Therefore, testing players while symptomatic can add nothing to clinical decision making and is likely to confound the interpretation of subsequent testing due to uncontrolled practice effects. A recent consensus paper included a statement to this effect as well, advocating that neurocognitive testing be deferred until players are symptom free.[5]

Fifth, avoiding the use of neuropsychological testing may prevent further study. We clearly called for additional prospective research to establish the use of neurocognitive testing in the management of sport-related concussion. As neuropsychologists, we would obviously like to be able to recommend this as routine clinical practice, but the data do not allow us to make this recommendation.

Finally, the real risks involved in ''premature'' return to play have never been clearly defined, and no assessment technique or management intervention has ever been demonstrated to result in risk modification. We do not believe it is appropriate for athletic teams to expend significant resources in order to implement an unproven method (neurocognitive testing) in an attempt to modify an unknown risk. Until these questions are answered, we believe that athletic trainers and team medical personnel are best advised to rely on their clinical judgment, augmented by the systematic use of symptom checklists and by consultation with neuropsychologists when appropriate.

## REFERENCES


1. McCrea M, Guskiewicz KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study. *JAMA.* 2003;290:2556–2563.
2. Nassiri JD, Daniel JC, Wilckens J, Land BC. The implementation and use of the Standardized Assessment of Concussion at the U.S. Naval Academy. *Mil Med.* 2002;167:873–876.
3. Daniel JC, Olesniewicz MH, Reeves DL, et al. Repeated measures of cognitive processing efficiency in adolescent athletes: implications for monitoring recovery from concussion. *Neuropsychiatry Neuropsychol Behav Neurol.* 1999;12:167–169.
4. Bleiberg J, Kane RL, Reeves DL, Garmoe WS, Halpern E. Factor analysis of computerized and traditional tests used in mild brain injury research. *Clin Neuropsychol.* 2000;14:287–294.
5. McCrory P, Johnston K, Meeuwisse W, et al. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. *Br J Sports Med.* 2005;39:196–204.