NeuroRehabilitation 22 (2007) 243–251
IOS Press

Retest reliability in adolescents of a computerized neuropsychological battery used to assess recovery from concussion

Sidney J. Segalowitz[a,*], Patrick Mahaney[c], Diane L. Santesso[a,b], Leslie MacGregor[a], Jane Dywan[a] and Barry Willer[c]
[a]*Psychology Department, Brock University, St. Catharines, Ontario, Canada*
[b]*Affective Neuroscience Laboratory, Psychology Department, Harvard University, Cambridge, MA, USA*
[c]*Department of Psychiatry, University at Buffalo, Buffalo, NY, USA*

Abstract. We examined in a group of 15-year-old adolescents the retest reliability over one week of 7 subscales of the Automated Neuropsychological Metrics (ANAM), a computerized battery based on standard neuropsychological test measures that is one of several such batteries available to assess concussion effects. Since the principle behind these computerized batteries is to assess athletes before injury and after injury to determine the level of deficit and whether the individual is safe to return to play, it is critical that such batteries have excellent retest reliability. Retest reliability of the ANAM was good, especially for the aggregate of throughput scores, reaching 0.87, but lower for individual subtests, especially for those measuring only speed of processing. Thus, the ANAM aggregated score appears to have robust reliability for cognitive measures involving memory and attention in 15-year-olds. Limitations related to assessing return-to-baseline after concussion in adolescents are discussed.

1. Introduction

Medical management of sports-related concussion has been a topic of considerable interest in the last decade. There are now standardized approaches to cognitive evaluation [20] and postural stability [13] that can be used on the field to assess the concussed athlete. There are a variety of symptom checklists that have emerged including the Sport Concussion Assessment Tool (SCAT) [21]. There has been a major reconsideration of management of the concussed patient in the emergency room including the advice given to family on signs and symptoms to watch for [2,11]. Perhaps the most dramatic change in concussion management has been the use of neuropsychological tests administered prior to injury and re-administered after concussion to test the athlete's return to pre-injury status [12, 25]. In this paper we will review the issues and success of this innovative use of neuropsychological testing and then assess one of the computer based versions of tests currently on the market for use with concussion assessment.

A recent review [28] of concussion and post concussion syndrome provided a model for distinguishing concussion from mild traumatic brain injury (mTBI) and post concussion syndrome (PCS). The model uses the most commonly accepted definition of mTBI and the one proposed by the American Congress of Rehabilitation Medicine and the Centers for Disease Control: loss of consciousness for no more than 30 minutes or amnesia as a result of a mechanical force to the head, and a Glasgow Coma Score (GCS) of 13 to 15 [3]. The model also uses the most commonly accepted definition of concussion as established by the American Academy of Neurology (AAN): a trauma induced alteration of mental status that may or may not involve loss of consciousness [15,16]. Although not

*Address for correspondence: Sid Segalowitz PhD, Department of Psychology, Brock University, 500 Glenridge Avenue, St. Catharines, Ontario, Canada. E-mail: sid.segalowitz@brocku.ca.

ISSN 1053-8135/07/$17.00

explicitly stated in the AAN definition, concussion is generally viewed as a transient state from which the individual will recover fully in a relatively short period of time [21]. In contrast, mTBI is viewed as a permanent alteration of brain function even though the individual with mTBI may appear asymptomatic. Post concussion syndrome was defined in the Willer and Leddy [28] model as persistent symptoms of concussion past the period when the individual should have recovered (3 weeks) and therefore qualifies as mTBI. Neuropsychological testing is often used to describe the impairment associated with mTBI and PCS and have done so with relative success [10,19,22–24,26]. In the current paper the focus is on the use of neuropsychological measures to assess concussion and concussion recovery.

Randolph, McCrea and Barr [25] provide an excellent review of the development and use of neuropsychological testing in athletics. Various batteries of tests were used with college athletes for research and later for individual diagnosis and return to play decisions [1,19]. The National Hockey and National Football Leagues in the US established a precedent when they agreed to use neuropsychological tests to establish a pre-injury baseline of cognitive performance to which players with concussion can be compared [18]. The expectation is that the athlete should return to baseline cognitive performance before returning to the field of play. The development of computerized testing programs as opposed to paper and pencil tests allowed for baseline testing of many athletes at once, without requiring a neuropsychologist on site to oversee the testing procedures [25].

There are a variety of other advantages to the use of computerized testing procedures. Cernich et al. [9] provide a historical review of the development of computer based testing and the authors point out a number of major advantages: (1) more rigorous standardization of administration; (2) increased accuracy of timing (for reaction time tests); (3) ease of administration; (4) ease of scoring, data storage and data access; and (5) randomly available alternate testing items. Alternate testing items are necessary to reduce the likelihood of practice effects with repeated testing. The Cernich et al. [9] review also points to a number of concerns regarding the use of computerized tests that warrant careful consideration when using such tests to assist in clinical judgments such as return to play decisions. They discuss the potential for error when using different computers and different operating systems at baseline and post concussion assessment. They also point out that the level of experience each athlete has with computers may influence results.

Randolph et al. [25] describe three commercially available computer based neuropsychological testing programs including ImPACT (University of Pittsburgh, USA), CogSport (CogState Ltd., Victoria, Australia) and Headminder Concussion Resolution Index (HeadMinder Inc., New York, USA). They also discuss one computer based assessment program that was developed by the US military called ANAM (Automated Neuropsychological Assessment Metrics). Each assessment instrument contains a battery of subtests assessing key cognitive constructs which are vulnerable to concussion: verbal memory, visuospatial memory, working memory, processing speed, and general reaction time. The Randolph et al. [25] review then describes the psychometric standards that must be met before any computer based assessment of cognitive function can be used for clinical purposes, such as return to play decisions. The most critical psychometric issue, according to Randolph et al. [25] is test-retest reliability. If a test does not have high test-retest reliability then the difference between a baseline score and a post concussion score may simply reflect the error of retesting rather any real difference in the performance of the athlete. For clinical decision making Randolph et al. [25] suggest a test-retest reliability of 0.9 is required. None of the computer-based testing programs met this criterion although some had not been rigorously tested.

The purpose of the current study is to examine the test-retest reliability of ANAM with adolescents. We selected ANAM for a variety of reasons. ANAM is not a commercial product. Most of the publications that describe the psychometric properties of any of the available computerized testing programs were written by the authors of the tests. When the authors are also share holders of the publisher of the program there is an inherent conflict of interest. This is not to suggest that authors with a commercial interest in the product would falsify data but they may have a tendency to not publish or perhaps not study critical psychometric properties. In 2005 when Randolph et al. [25] published their review, the authors reported there were no published reliability studies of ANAM or ImPACT and one published study each for CogSport and Headminder. In the two published studies of reliability, the reliability coefficients of individual sub-tests ranged from 0.31 to 0.82 [25].

ANAM is the result of approximately 30 years of computerized assessment test development [27]. It was developed for serial testing and precision measurement of cognitive functioning for the US military. Although ANAM was not developed specifically for concussion assessment, a sports medicine battery evolved and has undergone substantial psychometric evaluation [8]. Cernich et al. [8] collapsed the results from three separate studies in order to describe the psychometric properties of ANAM. One study involved the administration of ANAM 30 times over a four day period and demonstrated that individuals without concussion demonstrate fairly substantial practice effects [6]. The other two studies included high school and college athletes [7] and freshman military academy cadets [5]. In these latter two studies, ANAM demonstrated consistent correlations with traditional neuropsychological measures suggesting adequate concurrent validity. However, test-retest reliability for each subtest using the military sample ranged from a low of 0.38 to a high of 0.87 [8]. There was no test-retest reliability assessment completed on adolescents.

The purpose of the present study was to examine the test-retest reliability of ANAM within a relatively homogeneous population of adolescents (15 and 16 year olds). We deliberately elected to evaluate ANAM under ideal conditions where participants were assessed on the same computers, at the same time of day, and only seven days between administrations. We are aware that if ANAM is used to establish a baseline and then assess post concussion changes, the time from baseline assessment to post concussion assessment would be much more than one week. Further, none of the participants in this study had experienced concussion or had demonstrated any other known condition or illness that would influence cognitive performance. To be justified as a concussion assessment program, ANAM subtests should demonstrate very high reliability under ideal conditions because they are likely to have reduced reliability under the less than ideal conditions that generally characterize most sports medicine applications.

2. Methods

2.1. Participants

Participants were recruited from local high schools. Volunteers were solicited by the school principals. The announcement for participants indicated that the researchers were conducting a study of computer testing procedures and volunteers would be compensated $50 for their participation. Consent of a parent was required along with assent from the participant. The consent procedures were reviewed and approved by the research ethics committee of Brock University. Fifteen girls and 14 boys with the average age of 15.4 years (range = 15.0 to 16.8 years) were included in the study. No volunteers were refused participation. The participants were average to above average students and none had failed a grade or had academic difficulty. None had been diagnosed with a learning disability. All but two participants were right handed. Medical history was negative and none of the participants were on prescription medication. None of the participants had a history of concussion. All participants were quite familiar with computers and required no instruction on the use of the computer mouse.

2.2. Procedure

Participants came in twice at the same time of day (primarily afternoons) in a one-week interval. The sports medicine battery of ANAM [8] was administered on both occasions using the same stand alone desk top computer. The only persons present in the room at the time of the testing were the participant and a research assistant. The research assistant introduced the task and assisted with administration of ANAM only when there was a question. ANAM is, for the most part, self administered and the instructions on the screen are quite straightforward to follow. The ANAM sports medicine battery administered included the following subtests:

1. Code Substitution (CDS): This test uses a symbol-digit coding paradigm. Participants must scan a series of codes and match them to digits. As with all subtests, the participant is provided with several sample test items and given feedback on whether they are correct.
2. Code Substitution Delayed (CDD): This test presents the same symbols as CDS but the participant must remember the matching numbers from the earlier administration. The delay from CDS to CDD is approximately 10 minutes but depends on how quickly the participant proceeds through the intervening subtests.
3. Continuous Performance Test (CPT): This test is also called the Running Memory Continuing Performance Test [14]. The task is a continuous reaction time test using a ‘one back’ paradigm to assess working memory and sustained attention. Subjects are required to recall the last letter to appear on the screen and decide if the current letter displayed on the screen is the same as or different from the previous letter.

4. Mathematical Processing (MTH): The test requires the participant to perform basic arithmetic functions in order to determine if the three numbers presented in an equation with plus or minus signs are less than or greater than 5.
5. Match-to-Sample (MSP): The test presents the participant with a 4 $\times$ 4 red and white block design which disappears and the participant must identify which of two designs match the previous design.
6. Simple Reaction Time (SRT): In this test the participant is instructed to press the mouse key immediately upon presentation on the screen of a simple stimulus (an asterisk). The SRT measure occurs twice in the battery (beginning and end) so there is a SRT1 and a SRT2 measure. This test was designed as a pure reaction time assessment [27].

Each test provides an accuracy score (percent correct), an average time for correct responses (measured in milliseconds), and a throughput score (number of correct responses per minute). Test-retest reliability was calculated for each and for a composite of the scores.

Reliability was measured by the Intraclass Correlation Coefficient (*ICC*) and the Pearson correlation coefficient (r). While r reflects the degree to which participants' scores are ranked in the same order on second testing with similar magnitudes in individual differences, the *ICC* also takes into account any absolute changes over the two sessions. Such changes could represent increases in scores due to learning (practice effects) or automatization, or they could represent decreases due to distraction or simple boredom with the task. Thus, the *ICC* is required to inform us as to how clinically useful is any individual test score, i.e., how likely it is that we would obtain exactly the same value on a second testing. The r is useful for informing us as to how discriminating the test is across individuals in the context of group studies, even if the participants systematically improve or degrade their scores on a second testing. Thus, a high *ICC* value indicates that we can trust the particular score obtained while the r value indicates that we can trust the individual differences indicated by the score.

For purposes of research that compares a target group against a control group, the r reflects the reliability of interest. For purposes of assessing an individual, we need to know that when a test is administered we can trust its specific value and compare it against norms. This is especially an issue when we are administering neuropsychological tests repeatedly in order to document recovery. Standard psychological tests often should not be readministered within a year in order to reduce any learning or memory effect. Some problem solving tests cannot be usefully given twice ever because once the solution is obtained, the second testing is a measure of the person's ability to recall the solution rather than derive it. However, the purpose of the ANAM is to enable repeated testing without involving undue learning or memory effects or incurring large resource costs. Therefore, we need to know its *ICC* as well as r over this short testing period.

3. Results

3.1. Throughput measures

The number of correct responses per minute showed relatively strong retest reliability on the r values for measures that reflect some cognitive component other than simple reaction time (see Table 1). All the subtests achieved values similar to those in standard group research studies in cognitive psychology except perhaps for the SRT measures. The results achieved with this sample of adolescents were similar to the test-retest reliability correlations (r values) presented by Cernich et al. [8]. Cernich et al. reliability results are included in Table 1.

The *ICC* measures were somewhat lower, suggesting that there were simple test effects, which are reflected in the overall change in scores from the first test session to the second. As shown in Table 1, several of the tasks had significantly increased scores on the second session, and the total of these 5 scores (omitting the SRT measures) improved significantly, $t(27) = 5.8$, $p < 0.001$. For the individual scales, all the values reached statistical significance, illustrating that the retest is better than zero which is of course something we would expect, but only the nonSRT tasks reached comfortable levels for research purposes, that is, a reliability score greater than 0.6. No test on its own reached a high enough level for clinical utility. Note that the highest *ICC* reliability for a single test represents only about 50% of the variance across test sessions while the lowest reflects a 20% overlap across sessions. However, the sum of the 5 subscores reached near-clinical levels, with r and *ICC* values of 0.87, accounting for over 75% of the variance.

We sought further verification of the stability of the aggregated throughput scores by forming z-scores for

Table 1
Retest results for the ANAM throughput measures on each of the subtests. The aggregate measures indicate how improved the reliability is

	Pearson r	p-value	Adult Sample[1]	ICC	p-value	mean change[2]	t	p-value
CDD	0.67	<0.001		0.68	<0.001	1.46	0.755	0.457
CDS	0.81	<0.001		0.58	<0.001	10.06	7.771	<0.001
MSP	0.72	<0.001	0.66	0.72	<0.001	1.75	0.938	0.357
MTH	0.71	<0.001	0.87	0.61	<0.001	2.32	2.222	0.035
CPT	0.70	<0.001	0.58	0.65	<0.001	7.02	2.456	0.021
SRT	0.48	0.01	0.38	0.44	0.006	14.64	2.130	0.042
SRT2	0.50	0.007		0.47	0.004	−9.97	1.193	0.243
Average of tasks 1–5	0.87	<0.001		0.87	<0.001	4.52	5.833	<0.001
Average of all 7 z-scores	0.86	<0.001		0.86	<0.001			
Average of first 5 z-scores	0.88	<0.001		0.87	<0.001			

[1]The adult sample Pearson r correlations are taken from Cernich et al. [8] and are presented for comparison purposes.
[2]The mean change is the change in throughput from first to second testing.

Table 2
Retest results for the accuracy measures on each of the ANAM subtests. The simple reaction time (SRT) is omitted as this task did not have a choice in the response

	Pearson r	p-value	*ICC*	p-value
CDD	0.46	0.015	0.44	0.008
CDS	0.45	0.017	0.40	0.012
MSP	0.22	0.250	0.19	0.154
MTH	0.25	0.199	0.25	0.096
CPT	0.46	0.015	0.32	0.027

each of the test subscores. By doing this, we are able to equate them in an averaged aggregate measure knowing that they contribute equal weight to the average, producing an r and *ICC* of 0.86. This reliability is illustrated in Fig. 1 as not depending on extreme (outlier) scores, a common concern when testing a group of children where the variance across individuals may be high. The reliability is equally enhanced using z-scores from just the first 5 subtests, omitting the SRT tasks ($r = 0.877$, *ICC* = 0.874, $p < 0.001$).

3.2. Accuracy measures

Accuracy scores, independent of the response speed, showed weaker retest reliabilities (see Table 2). Retest values for these raw accuracy scores ranged from 0.22 to 0.46 for the r values, and 0.19 to 0.44 for the *ICC* values. The lowered reliabilities compared to those achieved for the throughput measures indicate that some involvement of speed of processing is necessary for acceptable reliabilities.

3.3. Reaction times

The reliability of the response times on their own is shown in Table 3. While some Pearson correlations were respectable for group studies, some were not, and the *ICC* values were especially limited.

Table 3
Retest coefficients for the mean response times on correct trials for each of the ANAM subtests

	Pearson r	p-value	*ICC*	p-value
CDD	0.74	<0.001	0.74	<0.001
CDS	0.83	<0.001	0.54	<0.001
MSP	0.59	0.001	0.59	<0.001
MTH	0.44	0.02	0.43	0.009
CPT	0.80	<0.001	0.65	<0.001
SRT	0.29	0.13	0.24	0.089
SRT2	0.46	0.013	0.38	0.02

Table 4
Intercorrelations among ANAM subtests at the first test session

	CDS	MSP	MTH	CPT	SRT	SRT2
CDD	0.653**	0.201	0.361	−0.012	−0.119	0.260
CDS		0.323	0.361	0.259	0.235	0.162
MSP			0.047	0.035	0.128	−0.108
MTH				0.601**	0.036	0.180
CPT					0.052	0.108
SRT						0.497**

**$p < 0.01$.

3.4. Intercorrelations among subtests

For reasons explained below, it is useful to know the intercorrelations among the subtests. As shown in Table 4 giving the intercorrelations upon first testing, the subtests appeared to be only moderately related, significance being reached only in two pairings.

4. Discussion

It is not surprising that interest in the neuropsychological assessment of concussion has increased given the attention to concussion generally. Kirkwood, Yeates and Wilson [17] reviewed the state of the science in the management of pediatric concussions and suggest that neuropsychological assessment has been

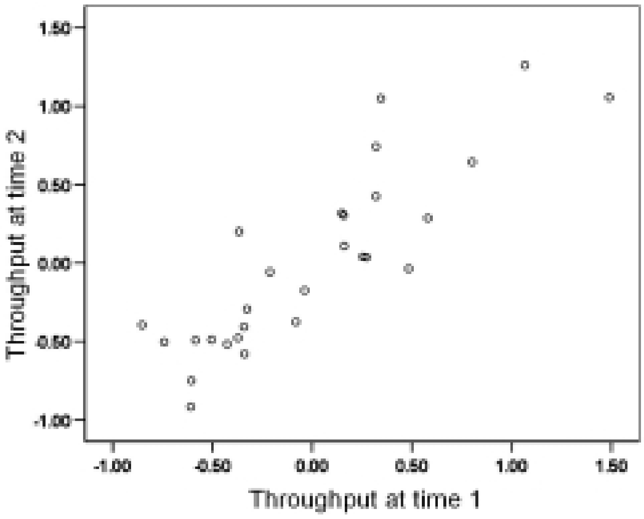

Fig. 1. Test-retest scatterplot of the aggregated throughput measures for all 7 subtests. The aggregate was formed by averaging the Z-scores from each of the subscales.

regarded as the best means of objectively identifying cognitive difficulties and making a differential diagnosis of concussion. The testing procedures and tests used have generally been thoroughly evaluated and tend to have excellent reliability. In fact, it is partly because of the reliability of these instruments that serial testing is not encouraged because they are vulnerable to practice effects. Further, neuropsychological testing has been impractical for assessment of sports related concussions because of the length of time it takes for a thorough assessment and the subsequent costs.

Brief testing batteries have been developed which made testing of large groups of athletes and persons with concussion feasible and a body of research ensued. Kirkwood et al. highlight some of the valuable findings that have come from this research. However, Belanger and Vanderploeg [4] recently published a meta analysis of neuropsychological test research on concussion and found that after one or two weeks neuropsychological tests were unable to differentiate those with concussion from those without concussion. In other words, while neuropsychological testing has been useful for research purposes it has not been useful for differentiating those with persistent symptoms (post concussion syndrome) from those who are fully recovered. It is possible, that neuropsychological tests used in brief batteries are less reliable than those used in more complete batteries [25].

The other development in neuropsychological testing that occurred as a result of research and practice with athletes is computerization of the brief batteries. As Kirkwood et al. [17] point out, computerization of batteries was thought to have many advantages over paper and pencil applications. The most important factor was the reduced cost and standardization of the testing procedures without the necessity of having a neuropsychologist in attendance to conduct the tests. The development of computerized testing programs made it feasible to test large numbers of athletes before the season (and therefore before the injury) in order to establish a baseline of cognitive performance. In theory, a player with a concussion should not return to play until their cognitive performance has returned to baseline. The use of neuropsychological tests for clinical decision making puts an expectation on the psychometric properties of the tests and procedures used.

The review by Randolph et al. [25] concluded that neuropsychological testing in the management of sports-related concussion is very useful for research but its use with individual athletes is limited by problems of untested or inadequate reliability. ANAM was one of the computerized testing programs discussed and at the time of the Randolph review there was limited information available on the psychometric properties of ANAM. Since that time, there have been a number of studies published which give us considerable insight into the factors assessed by ANAM and with our study there are now two analyses of test-retest reliability.

Of course, clinical assessment can only be as good as the tools used. While our data reflect retest reliability specifically on the ANAM, our results address more general issues in the clinical neuropsychology of concussion. Depending on their psychometric properties, neuropsychological test scores can be useful only for screening, only for group studies, or for clinical assessment of individuals. While standard clinical neuropsychological tests are designed and refined for clinical assessment, especially with respect to high reliability, short versions and computerized versions that are designed to be similar may differ considerably from their longer one-on-one counterparts. The reliability of component subtests of all standard measures never fare as well as aggregated measures. Our data suggest that the ANAM as a whole does well with respect to reliability.

4.1. *The significance of reliability values: Group versus clinical reliability*

ANAM subscales on the basis of the present data have sufficient reliability for studies involving screening and follow up, but for clinical applications an aggregated measure is needed. However, note that aggregating absolute scores gives higher weighting to those scales with larger scores. For these reasons, clinical utility would be enhanced if tables provided data from large samples of gender and age based cohorts showing conversion to z-scores.

We can fairly ask whether our reliability scores are exaggerated compared to the natural assessment context because of our care in reducing other sources of variation by testing participants individually and at the same time of day. The strength of the ANAM being in a computer format is that it can be administered without close supervision, and initial baseline measures would likely have larger error variance if taken in a group test session.

4.2. *Implications for the neuropsychology of concussion*

The construct validity of computerized tests is always a concern, not least because exposure and comfort with computers varies widely in our society, especially as a function of age. One of the few studies on construct validity for the ANAM indicates that it has modest relations with standard neuropsychological tests, although this was in a rather different population from ours [14]. We should also point out, however, that standard neuropsychological tests may have altered validity and retest reliabilities in a select sample reminding us that reliability and validity of test scores need to be rechecked when considering a specific population. For example, many standard neuropsychological assessment tools made for adults do not generalize their validity to children. For this reason, the application of ANAM to high school athletes needs to be evaluated for validity and reliability on that group [1].

4.3. *So what does ANAM performance reflect?*

The ANAM subtests are constructed to tap into the same processes as standard neuropsychological measures. However, as we have seen, the retest reliability of the accuracy or the response time scores is quite modest. The consistency arises when one considers the throughput score, a reflection of efficiency rather than pure accuracy of speed. In fact, despite the fact that all the ANAM measures are timed tests and participants feel the pressure to perform quickly (perhaps partly due to the computer-context), we can confidently conclude from our results that the scores do not simply reflect response speed. First, intercorrelations among scales are not large: Only the CDD-CDS and MTH-CPT pairings were consistently significant (besides the two SRT halves being related to each other, of course). Also, the fact that the SRT scores were the *least* reliable suggests that the other scales with higher reliability are tapping into some other processes, presumably cognitive efficiency measures.

The fact that the raw (non-throughput) scores did not achieve impressive reliability values while the throughout scores did suggests that there is some aspect of timing that is critical to performance on the ANAM. We can conceptualize this as some form of information processing efficiency and not simply speed of processing.

Trait or state? We tried to schedule for everyone the two test sessions at the same time of day in order to minimize variance due to circadian arousal cycles, time since last meal, and fatigue from daily activities. What this does, of course, is emphasize the trait characteristics of the ANAM in our measure of reliability. However, this means that we do not yet really know the robustness of the ANAM across normal daily routines. It may be that variation in ANAM performance is affected by these sorts of factors to the same extent as individual differences in skills. This has two implications. First, the ANAM should be given at a fixed time of day to all participants if the researcher wants to maintain the high levels of reliability demonstrated here. Second, the clinician should be careful about interpreting return-to-baseline data if gains or losses can be attributed to fatigue or arousal factors. Of course, this concern can be alleviated by another study specifically examining the effects of circadian contributions to variance, for example, by carrying out a study such as the present one varying the timing of the two test sessions. If the reliability of the ANAM does not reduce, then we know that it taps into some very consistent trait characteristics.

4.4. Implications for understanding recovery from concussion

We know that response speed is reduced after concussion, but so is information processing efficiency, independent of speed *per se*. Such factors are critical for learning and normal independent living skills. In the context of adolescent injury and return to normal participation in group functions, perhaps the most stressful for parents and teachers is the question of returning to the normal level of athletic and other risky activities. Returning to the individual's normal baseline speed of processing may be one prerequisite for "return to play" for the injured adolescent, but return to their baseline efficiency of information processing is surely a prerequisite for avoiding further injury. By extension, one must be careful about interpreting other test scores that rely heavily on response time as reflecting adequate recovery after concussion.

The ANAM does not measure and therefore does not reflect many of the other important cognitive information processing skills. For example, we cannot know from ANAM scores whether an individual has returned after a concussion to a baseline level of problem solving, school learning skills, complex analytic thought, or affect regulation. However, one can imagine that this is not an issue for situations where the school official or parent must decide when previous tasks should be taken up again. The ANAM has, so far, proved itself to have some of the psychometric properties necessary for this purpose.

Acknowledgments

This research was supported by a grant from the Ontario Ministry of Health Promotion (to BW and SJS) and from the Natural Sciences and Engineering Research Council of Canada (to SJS). We would like to thank James Desjardins and Sonia Sanichara Kahn for their help in data collection. Finally, we wish to thank the District School Board of Niagara and the Niagara Catholic District School Board for allowing access to students through appropriate ethical procedures.

References

[1] J.T. Barth, J.R. Freeman and J.E. Winters, Management of sports-related concussions, *Dent Clin North Am* **44** (2000), 67–83.

[2] J. Bazarian, M. Hartman and E. Delahunta, Minor head injury: predicting follow-up after discharge from the Emergency Department, *Brain Inj* **14** (2000), 285–294.

[3] J.J. Bazarian, B. Blyth and L. Cimpello, Bench to bedside: evidence for brain injury after concussion–looking beyond the computed tomography scan, *Acad Emerg Med* **13** (2006), 199–214.

[4] H.G. Belanger and R.D. Vanderploeg, The neuropsychological impact of sports-related concussion: a meta-analysis, *J Int Neuropsychol Soc* **11** (2005), 345–357.

[5] J. Bleiberg, A.N. Cernich, K. Cameron, W. Sun, K. Peck, P.J. Ecklund, D. Reeves, J. Uhorchak, M.B. Sparling and D.L. Warden, Duration of cognitive impairment after sports concussion, *Neurosurgery* **54** (2004), 1073–1078; discussion 1078–1080.

[6] J. Bleiberg, W.S. Garmoe, E.L. Halpern, D.L. Reeves and J.D. Nadler, Consistency of within-day and across-day performance after mild brain injury, *Neuropsychiatry, Neuropsychology, and Behavioral Neurology* **10** (1997), 247–253.

[7] J. Bleiberg, R.L. Kane, D.L. Reeves, W.S. Garmoe and E. Halpern, Factor analysis of computerized and traditional tests used in mild brain injury research, *The Clinical Neuropsychologist* **14** (2000), 287–294.

[8] A. Cernich, D. Reeves, W. Sun and J. Bleiberg, Automated Neuropsychological Assessment Metrics sports medicine battery, *Arch Clin Neuropsychol* **22(**Suppl 1) (2007), 101–114.

[9] A.N. Cernich, D.M. Brennana, L.M. Barker and J. Bleiberg, Sources of error in computerized neuropsychological assessment, *Arch Clin Neuropsychol* **22(**Suppl 1) (2007), 39–48.

[10] D.M. Erlanger, K.C. Kutner, J.T. Barth and R. Barnes, Neuropsychology of sports-related head injury: Dementia Pugilistica to Post Concussion Syndrome, *The Clinical Neuropsychologist* **13** (1999), 193–209.

[11] M. Fung, B. Willer, D. Moreland and J.J. Leddy, A proposal for an evidenced-based emergency department discharge form for mild traumatic brain injury, *Brain Inj* **20** (2006), 889–894.

[12] S.H. Grindel, M.R. Lovell and M.W. Collins, The assessment of sport-related concussion: the evidence behind neuropsychological testing and management, *Clin J Sport Med* **11** (2001), 134–143.

[13] K.M. Guskiewicz, Postural stability assessment following concussion: One piece of the puzzle, *Clin J Sport Med* **11** (2001), 182–189.

[14] M.H. Kabat, R.L. Kane, A.L. Jefferson and R.K. DiPino, Construct validity of selected Automated Neuropsychological Assessment Metrics (ANAM) battery measures, *The Clinical Neuropsychologist* **15** (2001), 498–507.

[15] J.P. Kelly and J.H. Rosenberg, The development of guidelines for the management of concussion in sports, *The Journal of Head Trauma Rehabilitation* **13** (1998), 53–65.

[16] J.P. Kelly and J.H. Rosenberg, Diagnosis and management of concussion in sports, *Neurology* **48** (1997), 575–580.

[17] M.W. Kirkwood, K.O. Yeates and P.E. Wilson, Pediatric sport-related concussion: a review of the clinical management of an oft-neglected population, *Pediatrics* **117** (2006), 1359–1371.

[18] M. Lovell, M. Collins and J. Bradley, Return to play following sports-related concussion, *Clin Sports Med* **23** (2004), 421–441, ix.

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

[20] M. McCrea, J.P. Kelly, C. Randolph, J. Kluge, E. Bartolic, G. Finn and B. Baxter, Standardized assessment of concussion (SAC): on-site mental status evaluation of the athlete, *Journal of Head Trauma and Rehabilitation* **13** (1998), 27–35.

[21] P. McCrory, K. Johnston, W. Meeuwisse, M. Aubry, R. Cantu, J. Dvorak, T. Graf-Baumann, J. Kelly, M. Lovell and P. Schamasch, Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004, *Br J Sports Med* **39** (2005), 196–204.

[22] S.R. Millis, M. Rosenthal, T.A. Novack, M. Sherer, T.G. Nick, J.S. Kreutzer, W.M. High, Jr. and J.H. Ricker, Long-term neuropsychological outcome after traumatic brain injury, *The Journal of Head Trauma Rehabilitation* **16** (2001), 343–355.

[23] W. Mittenberg and S. Strauman, Diagnosis of mild head injury and the postconcussion syndrome, *The Journal of Head Trauma Rehabilitation* **15** (2000), 783–791.

[24] G. Mooney, J. Speed and S. Sheppard, Factors related to recovery after mild traumatic brain injury, *Brain Inj* **19** (2005), 975–987.

[25] C. Randolph, M. McCrea and W.B. Barr, Is neuropsychological testing useful in the management of sport-related concussion? *J Athl Train* **40** (2005), 139–152.

[26] P.M. Rees, Contemporary issues in mild traumatic brain injury, *Arch Phys Med Rehabil* **84** (2003), 1885–1894.

[27] D.L. Reeves, K.P. Winter, J. Bleiberg and R.L. Kane, ANAM Genogram: Historical perspectives, description, and current endeavors, *Arch Clin Neuropsychol* **22**(Suppl 1) (2007), 15–37.

[28] B. Willer and J.J. Leddy, Management of concussion and post-concussion syndrome, *Current Treatment Options in Neurology* **8** (2006), 415–426.

[29] H. Yaghi, Pre-university students' attitudes towards computers: An international perspective, *J Educat Computing Research* **16** (1997), 237–249.