CHAPTER 3

NEUROPSYCHOLOGICAL ASSESSMENT OF SPORTS-RELATED CONCUSSION: MEASURING CLINICALLY SIGNIFICANT CHANGE

Aaron M. Rosenbaum[1]; Peter A. Arnett[1]; Christopher M. Bailey[1]; and Ruben J. Echemendia[2]

[1]*Department of Psychology, 566 Moore Building, The Pennsylvania State University; University Park, paa6@psu.edu*

[2]*Psychological and Neurobehavioral Associates, Inc. State College, PA 16801.*

Abstract: In recent years there has been a dramatic increase in the use of neuropsychological tests to evaluate the effects of concussion in competitive athletes and assist in return to play decisions. In this chapter, we focus on one factor that can limit the sensitivity of neuropsychological tests to concussion--practice effects. The data we present suggests that the HVLT-R, Trails B, Stroop 2, and SDMT are most susceptible to practice effects upon repeated administration. Nonetheless, we show that even for these tests, a majority of control athletes do not show significant practice effects after several administrations when the reliability of the measures and regression to the mean are controlled for. Still, the fact that a significant minority of athletes show practice effects on these tests should serve as a note of caution for interpreting these commonly used clinical neuropsychological tests post-concussion. In contrast to these test indices, the Stroop 1 and Trails A showed little evidence for practice effects even when administered several times. Because the Stroop 1 also showed evidence for sensitivity to concussion, it emerged as perhaps the best test in terms of combined resistance to practice effects and concussion sensitivity. In terms of return to play decisions, because we found that a negligible number of controls displayed evidence for reliable decline from baseline on all six test indices, the data we present in this chapter strongly suggest that when concussed athletes continue to show performance reliably below baseline performance at one-week post-concussion on any of the noted test indices, great caution should be exercised in recommending return to play. Additionally, any athlete who is still reliably below baseline on two of the test indices at one-week post-concussion should not return to play because residual persisting cognitive effects from the concussion are highly likely. Future work can extend this research by using larger samples, better matching on overall cognitive ability.

Keywords: Concussion; Neuropsychology; Cognitive tests; Mild traumatic brain injury.

1. EPIDEMIOLOGY AND SYMPTOMS OF CONCUSSION

Concussion is a relatively common occurrence in sports. Concussion in sport was recently defined in a summary statement at the Prague Conference (McCrory et al., 2005) as "...a complex pathophysiological process affecting the brain, induced by traumatic biomechanical forces (p. 196)." Concussion has also been previously defined by Kelly and Rosenberg (1997) as "a temporary alteration in consciousness not necessarily with loss of consciousness (p. 2867)." In their study of over 17,000 athletes Guskiewicz and colleagues (2000) found that 5.1% of athletes involved in college football and soccer had sustained a concussion throughout their careers. Other studies indicate that 250,000-300,000 hospitalizations per year are due to concussions (Gerberich et al., 1983; National Institutes of Health Consensus Development Panel on Rehabilitation of Persons with Mild Traumatic Brain Injury, 1999; Thurman et al., 1998). These figures most likely represent an underestimate considering that many athletes who sustain low grade concussions either do not recognize that they have sustained a concussion (National Institutes of Health Consensus Development Panel on Rehabilitation of Persons with Mild Traumatic Brain Injury, 1999) or fail to report the concussion for fear of being removed from competition (Echemendia and Cantu, 2004).

Concussions typically involve a combination of physical (e.g., headache, dizziness, etc.), cognitive (e.g., attention & concentration difficulties), and affective (depression, anxiety, etc.) symptoms (Alves et al., 1993; Berlanger et al., 2005; Cantu, 2001; Gouvier et al., 1992; Guskiewicz et al., 1997; Guskiewicz, et al., 2000; Kelly and Rosenberg, 1997; Zasler, 1999). Empirical studies suggest that when concussed athletes are held out of competition, their symptoms tend to resolve within 7-10 days (Echemendia and Julian, 2001; Lovell et al., 1999). However, if athletes, especially young athletes, sustain a second injury before the physiological or overt symptomatology has resolved, there is a risk for Second Impact Syndrome (SIS), a condition where there is an irreversible increase in pressure in the brain which leads to death (Cantu, 1998). Thus, it is essential to monitor overt symptomatology as well as neuropsychological functions in concussed athletes.

2. MEASURING CHANGE I NEUROPSYCHOLOGICAL TESTING

Practice effects on neuropsychological measures are well documented (Chelune et al., 1993; Iverson and Gaetz, 2004) and occur because exposure to test procedures or stimuli facilitates improved performance on subsequent testing. The improved performance may be due to procedural practice effects because of familiarity with the test procedure or content practice effects that occur because test stimuli are remembered from one test administration to the next. Practice effects occur in both traditional "paper and pencil" tests and computerized test batteries. The use of alternate forms helps to mitigate content practice effects but not procedural practice effects. These practice effects can be problematic in test interpretation because improvement due solely to practice effects may be confused with improved neurocognitive functioning. Practice effects are typically measured using a non-injured group of participants tested at least twice. Practice effects vary with the number and time interval of testing, with those tests occurring in close proximity having the greatest practice effects. It is also worthy to note that practice effects have shown an asymptote and may reach a ceiling after two administrations. There are relatively few studies that have reported on practice effects in competitive athletes with concussion. Competitive athletes tend to be young (<35 years of age), motivated to perform well due to a desire to return to competition, and are susceptible to brain injury. Such susceptibility necessitates neuropsychological testing to monitor injuries and prevent further injuries from occurring before the athlete fully recovers (Echemendia and Julian, 2001). To provide greater precision in identifying real cognitive change post-concussion, it would be valuable if more information concerning practice effects on commonly used neuropsychological instruments were available for competitive athletes. The present chapter will present data to address practice effects in a group of elite college athletes who are at risk for or who have sustained a concussion. This chapter expands upon and refines previously presented data from the Penn State Concussion project (Mackin et al., 1997).

Conventional concussion monitoring and management programs have evolved using the baseline test paradigm pioneered by Barth and his colleagues at the University of Virginia (Barth et al., 1989). Using this paradigm, athletes involved with sports that are at risk for concussion are usually tested when they first join a team. If athletes sustain a concussion, they are tested serially post-injury using the same battery of tests with alternate forms, if available. Post-injury test data are compared to baseline data in order to identify any decrements in performance. If significant declines are found, the athlete is generally not returned to play until baseline levels of functioning are achieved. There is wide variability in the number and timing of post-injury testing. Some programs require a fixed testing interval such as 2-7 days post-concussion, while others advocate that players should not be tested until asymptomatic (McCrea et al., 2005). Generally, it is recommended that players should be returned to competition when neuropsychological test performance returns to baseline and overt symptoms are no longer reported (Echemendia and Cantu, 2004).

Because of the relatively high prevalence of concussion and the potentially serious consequences of premature return to competition, neuropsychological measures have provided useful information to assist in the monitoring of concussion and in return to play decisions. Furthermore, because of frequent post-injury retest intervals, it is also important to consider the likely influence of practice effects. Several methods have been developed to evaluate the magnitude of practice effects and clinically significant change. These include effect size (Dikmen et al., 1999), reliable change index (RCI; (Jacobson and Truax, 1991)), and a reliable change index adjusted for practice effects (RCI practice; (Chelune et al., 1993)). There have only been a few studies where the RCI method has been applied to the neuropsychological performance of athletes with or without concussion (Barr and McCrea, 2001; Hinton-Bayre et al., 1999; Iverson et al., 2003; Iverson and Gaetz, 2004). More typically, statistically significant change has been examined through the use of conventional statistical analyses. These analyses are not intended to measure clinically significant change in individuals; instead, these methods provide general group information about mean changes. The use of RCI and RCI practice help to accurately capture "error variance" in test scores and thereby produce a more clinically useful method for identifying when a clinically significant change has occurred. A central goal for this chapter is to provide a comprehensive analysis of several commonly used neuropsychological measures with concussed athletes using three of the most commonly employed measures of change. In particular, the Hopkins Verbal Learning Test – Revised (HVLT-R; Brandt and Benedict, 2001), Stroop tests (Trennery et al., 1989), Trailmaking tests A and B (Reitan, 1958), and the Symbol Digit Modalities Test (SDMT; (Smith, 1982)) will be reviewed based on data obtained from the Penn State Concussion project (Echemendia, 1997). What follows is a review of some data addressing the practice effects in and reliability of these measures. The Appendix of this chapter includes a description of each way of measuring change discussed, and formulas for calculating them.

Although our focus in this chapter is practice effects in non-injured control athletes who are tested at time intervals similar to concussed athletes, we also chose to examine the possibility of practice effects in injured athletes. At first glance, the use of the term "practice effects" in injured athletes appears to be a misnomer, because the change in score for injured athletes contains both practice effects and true change because of neurocognitive improvement. However, we chose to examine possible practice effects in injured athletes as well, especially by one-week post concussion, because if we observe improvement of injured athletes that goes beyond their original baseline, then this cannot simply be due to cognitive recovery. Some practice effect must be present in such a scenario. Regardless, for ease of exposition in this chapter, we will refer to both practice effects as they are observed in non-injured athletes and practice effects plus cognitive recovery observed in injured athletes simply as "practice effects."

2.1. Measuring Change on the Trailmaking Tests

Several studies have examined practice effects on Trails A and B in non-athlete populations (Basso et al., 1999; Bornstein et al., 1987; Craddick and Stern, 1963; desRosiers and Kavanagh, 1987; Dikmen et al., 1999; Dikmen et al., 1983; Dye, 1979; Matarazzo et al., 1974; Mitrushina and Satz, 1991). However, most of these studies have examined practice effects at only one retest interval (Basso et al., 1999; Bornstein et al., 1987; desRosiers and Kavanagh, 1987; Dikmen et al., 1999; Matarazzo et al., 1974). Two studies on non-athletes have examined practice effects at two or more retest intervals (Craddick and Stern, 1963; Mitrushina and Satz, 1991). Results from both studies revealed statistically significant practice effects on Trails A. On Trails B Mitrushina and Satz (1991) compared mean differences between time points and found statistically significant practice effects at two, one-year intervals. Craddick and Stern (1963) used the same approach and found statistically significant practice effects on Trails B after a third one-month interval. In the studies using only one retest interval, significant practice effects were found on both Trails A and B (desRosiers and Kavanagh, 1987; Dye, 1979). In contrast, Dikmen and colleagues (Dikmen et al., 1999) used the effect size method and Basso and colleagues (1999) used the RCI practice method and found no evidence of significant change on either Trailmaking test.

Practice effects on Trails A and B have also been examined in athletes. In a study conducted by Macciocchi and colleagues (2001), MANOVA was used to examine differences among 24 concussed athletes on Trails A and B, the SDMT, and several other neuropsychological tests. Athletes were tested at 24 hours, five days, and ten days post-injury following a first concussion. These investigators found that even these concussed athletes showed significant improvement from pre-season baseline scores to five days post-injury on Trails A, however, they did not show significant improvement at 24 hours post-injury. On Trails B, these athletes showed a large improvement in performance from baseline to 24 hours post-injury that persisted to five days post-injury. For both Trails A and B, there was no further improvement in performance from five to 10 days post-injury.

Macciocchi and colleagues (1996) reported on practice effects for Trails A and B in 183 athletes with one concussion and matched non-injured, non-athlete student controls. Practice effects were reported for both athletes and controls at 24 hours, five days, and ten days post-injury. Conventional statistical analyses were used to compare mean differences between injured athletes and controls. These investigators found that both injured athletes and their non-injured matched controls showed improvements in performance at each testing interval that increased as the number of testing intervals increased on both Trails A and B. No RCI or RCI Practice comparisons were reported.

Guskiewicz and colleagues (2001) also examined practice effects for Trails A and B. These researchers used repeated measures ANOVAs to examine practice effects relative to baseline at 24 hours, three days, and five days post-injury in a sample of 36 injured athletes and 36 non-injured matched control athletes. In contrast to Macciocchi and colleagues' studies, they found that injured athletes showed worse performance relative to controls compared with baseline at all time points post-injury on Trails B. Injured athletes also displayed a significant decline from their baseline scores at 24 hours post-injury. The groups were not differentiated on Trails A at any post-injury time point, but examination of mean scores for the groups reveals that the concussed group showed evidence of notable improvement from baseline to day five post-injury. The control sample showed effectively static performance from baseline to the 24 hour time point, and then notable improvements at three and five days on Trails A.

Iverson and Gaetz (2004) compared 126 non-concussed collegiate football players at pre-season and then post-season on Trails A and B. They found that, overall, the athletes displayed significant practice effects between time points for both measures. Applying the RCI methodology to this same sample and using 90% confidence intervals (as we do in our sample below), these investigators found that only about 5% of both football and soccer players improved from pre-season to post-season on Trails A. These values were similar for football players when the RCI Practice method was applied, whereas practice effects for the soccer players were reduced to about 3%. On Trails B using the RCI methodology, about 8% of both groups of players improved from pre-season to post-season. When the RCI Practice methodology was applied, this value was reduced to about 5-6%. Several researchers have reported test-retest reliability coefficient for Trails A and B. The estimates for Trails A have ranged from .46-.79 and from .44-.90 for Trails B (Basso et al., 1999; desRosiers and Kavanagh, 1987; Matarazzo et al., 1974; Mitrushina and Satz, 1991).

2.2. Measuring Change on the Stroop Tests

Several studies have examined practice effects on the Stroop tests. Because there are several different versions of the test, however, the conclusions that can be drawn from the results cannot necessarily be applied to the version of the test that we have used in our research. Nonetheless, the results of these studies are fairly consistent. Statistically significant practice effects have been demonstrated using a computerized Stroop test (Davidson, 2003), an abbreviated (40 item) version (Houx et al., 2002), the Stroop (1935) version (Dikmen et al., 1999), the Golden (1978) version (Connor et al., 1988), and in the normative sample of the Trenerry (1989) version. With the exception of Dikmen and colleagues' (1999) study, which used the RCI practice method, statistical calculations comparing mean differences between administrations were used for each of the other studies to identify practice effects. Furthermore, these studies used between one and five retest intervals, but the intervals were not comparable to those typically used in neuropsychological testing for concussed athletes.

Hinton-Bayre and Geffen (2004) describe a study of 13 concussed and 13 control Australian rules football players that used the Stroop, among other tests, at baseline and then two days post-concussion. On the Stroop A, where examinees were simply required to read color words on a page, concussed athletes displayed significantly slower performance at two days post-concussion compared with baseline. No significant change was noted for the control group. In contrast, on Stroop B, the color-word version of the test, no change post-concussion was observed in the concussed group, but the controls displayed a significant practice effect.

The test-retest reliability estimates of Stroop tests in previous research range from .79 to .84 on the color only version (Franzen et al., 1987; Dikmen et al., 1999) and from .77 to .90 on the color-word trial (Houx et al., 2002; Franzen et al., 1987; Trenerry et al., 1989).

2.3. Measuring Change on the Symbol Digit Modalities Test (SDMT)

A few studies have examined the nature and extent of practice effects on the SDMT. The data collected from the original normative sample (Smith, 1982) only included the mean scores of the participants at two testing intervals; no statistical analyses were reported to identify whether the difference between the mean values was statistically significant. Uchiyama and colleagues (Uchiyama et al., 1994) compared mean values at different retest intervals. The results of their analyses showed that the differences between the baseline scores and the scores at the retest intervals did not reach statistical significance.

In the Hinton-Bayre and Geffen (2004) study of Australian rules football players described above, they also used the SDMT. They found that the concussed group declined significantly in performance from baseline, but the performance of the control athletes was comparable to baseline. So, like the Stroop 1, the SDMT showed no evidence of a practice effect in the control group between two testing points.

In the study described earlier using 24 concussed collegiate athletes tested at baseline and then at post-concussion intervals of 24 hours five days and 10 days, Macciocchi and colleagues (2001) also examined the SDMT. Athletes who had sustained one concussion displayed notable improvement in performance from baseline at five days post-concussion, but especially by 10 days. Test-retest reliability coefficients for the SDMT have been reported. Uchiyama and colleagues (1994) reported a .79 coefficient and Smith (1982) reported a coefficient of .80.

2.4. Measuring Change on the Hopkins Verbal Learning Test – Revised (HVLT-R)

In Guskiewicz and colleagues' (2001) study described earlier, repeated measures ANOVAs were conducted to examine practice effects at 24 hours, three days, and five days post-injury in a sample of injured athletes and non-injured control athletes on the HVLT-R. They reported a significant group by day interaction for the HVLT-R, but examination of the group means of the concussed and control groups reveals very little change from baseline to any of the post-testing intervals. For example, the largest raw word increase for total immediate memory across the three HVLT-R learning trials was less than two words (from day 3 post-injury to day 5 post-injury time points in controls). Most of the other changes for both groups were approximately one word or less, suggesting that practice effects are not very significant at a clinical level. The use of alternate forms may have significantly attenuated practice effects in this case.

The HVLT-R-Revised manual (Brandt and Benedict, 2001) indicated that the test-retest reliability coefficient for this measure is .74.

2.5. Summary and Conclusions Regarding Measuring Change with the Trailmaking Tests, the Stroop test, the SDMT, and the HVLT-R

As presented above, some data have been published regarding practice effects on the SDMT, HVLT-R, Stroop tests, and Trails A and B in samples of injured and control athletes. Also, some investigators have considered the impact of practice effects on the interpretation of data after serial neuropsychological testing in athletes (e.g., (Bohnen et al., 1992; Echemendia and Julian, 2001)); however, most studies have compared mean differences between athletes and control groups rather than using methods that provide information about the magnitude of change (e.g., effect size) or about clinical significance (e.g., RCI and RCI practice) and the number of participants exhibiting significant practice effects. A few studies have examined RCI and RCI practice, but to our knowledge there are no published data in the sports neuropsychology research literature comparing the two methods. Although some studies have compared methods for the measurement of change using computerized measures (Erlanger et al., 2003) we have not found published research that has focused on examining practice effects on each of the four highlighted commonly used paper-and-pencil clinical neuropsychological tests presented in this paper. That said, McCrea and colleagues' (2005) recently published a study on collegiate athletes using a standardized regression-based method to evaluate change in a group of concussed and control athletes on some of the same tests we examine in this chapter and at the same time intervals post-concussion; however, the focus was not on practice effects, per se, but change following concussion. The data we outline below addresses some of the limitations in this literature.

3. METHOD

3.1. Participants

The data we present were derived from a subset of participants from the Penn State Concussion project (Echemendia, 1997). This project has been in progress since 1997 and extensive data have been obtained from athletes involved in high-risk collegiate sports. Participants were selected on the basis of their completion of the neuropsychological tests at baseline and at each of the post-concussion intervals: 2 hours, 48 hours, and one week. The sample included 60 concussed athletes and 28 control athletes. The mean age at baseline was 18.8 (1.1) years for the concussed athletes and 18.6 (0.9) years for controls. SAT scores were a mean of 833 (441) for concussed athletes and 1017 (344) for controls. Table 1 displays categorical demographic information about the participants. Most of the athletes in both the control and injured groups were Caucasian males participating in football and men's ice hockey.

Table 1. Demographic Information

Ethnicity		*Caucasian*	*African-American*
	Injured	53 (88%)	7 (12%)
	Control	27 (96%)	1 (4%)
Sex		Male	Female

	Injured	51 (85%)	9 (15%)
	Control	27 (96%)	1 (4%)
Sport	Injured	Football	17 (28%)
		Men's Soccer	2 (3%)
		Women's Soccer	4 (7%)
		Men's Ice Hockey	23 (38%)
		Men's Basketball	3 (5%)
		Women's Basketball	1 (2%)
		Men's Lacrosse	2 (3%)
		Women's Lacrosse	3 (5%)
		Men's Rugby	2 (3%)
		Women's Rugby	1 (2%)
		Wrestling	2 (3%)
	Control	Football	1 (4%)
		Men's Ice Hockey	20 (71%)
		Men's Soccer	1 (4%)
		Men's Basketball	2 (7%)
		Women's Basketball	1 (4%)
		Men's Swimming	3 (11%)

3.2. Measures.

The Symbol Digit Modalities Test (SDMT; (Smith, 1982)) involves matching numbers to figures and measures complex scanning, visual tracking, and information processing speed. The visual form of the task was used and the total number correct in 90 seconds was the dependent variable.

The Hopkins Verbal Learning Test – Revised (HVLT-R; Brandt and Benedict, 2001) measures verbal learning and memory. It involves the presentation of a list of 12 words across three trials followed by an immediate recall test after each trial and a delayed recall trial. Total recall across the three trials was used as the dependent variable.

The Stroop Color-Word test, as previously mentioned, has several forms. However, across all forms, this test examines attention, processing speed, and response inhibition (Stroop, 1935). We used Trenerry's (1989) version. This version includes the presentation of a list of 112 words in four columns with two trials: a word-only trial (Stroop 1) where the examinee reads only the words and ignores the colors, and the color-word trial (Stroop 2) where the participant reads the color that the word is printed in and ignores the word. For both Stroop 1 and Stroop 2, number correct per second was used as the dependent variable.

The Trailmaking Test (Reitan, 1958) examines simple and complex visual sequencing through the presentation of two trials: Trails A involves connecting a series of numbers, and Trails B requires connecting numbers to letters in an alternating sequence. For each task, total time was used as the dependent variable.

3.3. Procedure

The baseline tests were administered before athletes began their participation in athletics at Penn State. Post-concussion testing was conducted at 2 hours, 48-hours, and one-week post-injury. In the case of the controls, testing was yoked to an athlete in the sample at each of these time points. The tests were given as part of a larger battery of neuropsychological tests explained in more detail elsewhere (Echemendia et al., 2001). An attempt was made to match control athletes on gender, age, ethnicity, and SAT total score; control athletes were well-matched on these variables, with the exception of SAT scores. This issue is addressed below. The full test battery selected for this study consisted of SDMT, Trails A and B, Stroop and HVLT-R, which were administered at baseline, 48-hours, and one-week post-concussion. Because of logistical concerns the two-hour test battery was composed of a subset of tests, which included the Stroop and HVLT-R.

4. RESULTS

4.1. Preliminary Analyses

Independent samples t-tests were conducted to compare the age and total SAT score of the injured athletes and controls. No statistically significant difference between the groups was found on age, $t(79)=0.83$, $p > .10$.

However, when comparing the groups on total SAT score, controls had significantly higher scores than concussed athletes, $t(67.88)=-2.06$, $p < .05$. As a result of this significant difference, follow-up analyses were conducted comparing the groups at baseline on each of the neuropsychological measures. T-tests revealed no statistically significant differences between injured and control athletes on any of the tests. These follow-up analyses show that the group differences in SAT scores were not associated with group differences on the neuropsychological tests at baseline.

Chi-square analyses were conducted to identify differences between the groups in ethnicity or sex. Using Fisher's Exact tests, no significant differences were found for ethnicity or sex, Fisher's Exact $p > .10$ for both.

4.2. General Descriptive Information

Table 2 displays the means and standard deviations of the scores for both the concussed and control athletes on each measure. Perusal of these mean values shows that the concussed athletes displayed declines in performance from baseline to 48 hours on the HVLT-R and Stroop 1, and surprisingly, the control athletes showed a slight decline on the Stroop 1 at the same interval. On the remaining measures, slight improvements in performance were exhibited by both groups from baseline to 48 hours and then from 48 hours to one week. It is important to recognize that practice effects were not only exhibited by the control group, but were also found in the concussed group. It is also important to highlight that these are simply observations based upon a visual inspection of the data, not statistical analyses. Quantitative approaches to evaluating changes in the data follow below. Also, readers should be aware that, because the Trails A and B tests are timed, improvements on the tests are indicated by a decrease in the value in the table, the time required to complete the tests.

Table 2. General Descriptive Data for Each Measure According to Injured or Control Status

Measure	*Means*		*Baseline*	*n*	*48 hr*	*n*	*1 wk*	*n*
Hopkins Recall	Injured	mean	26.47	36	25.75	36	28.11	36
Total Learning		sd	3.07		4.58		4.7	
	Control	mean	28.22	18	30.5	18	31.72	18
		sd	3.75		3.73		3.39	
SDMT Total	Injured	mean	60.68	59	62.98	59	70.81	59
Correct		sd	9.42		10.13		13.81	
	Control	mean	62.15	27	65.67	27	75.67	27
		sd	7.28		8.99		14.71	
Stroop 1 Correct	Injured	mean	2.22	59	2	58	2.11	61
Per Time		sd	.38		0.48		0.52	

	Control	mean	2.23	27	2.19	28	2.18	28
		sd	.47		0.55		0.53	
Stroop 2 Correct Per Time	Injured	mean	1.03	59	1.12	58	1.2	60
		sd	.19		0.25		0.26	
	Control	mean	1.03	27	1.12	28	1.19	28
		sd	.17		0.24		0.27	
Trails A Time	Injured	mean	23.27	59	20.7	59	17.2	59
		sd	6.52		6.06		4.62	
	Control	mean	23.11	28	19.26	28	16.95	28
		sd	5.79		6.08		3.82	
Trails B Time	Injured	mean	52.62	58	49.05	58	39.53	58
		sd	14.29		15.51		11.19	
	Control	mean	46.89	28	45.38	28	40.21	28
		sd	11.9		13.05		12.4	

4.3. Test-Retest Reliability

Test-retest reliability scores were obtained using the first two scores obtained from each control athlete for each test. For the Stroop tests and HVLT-R, test scores were used from the baseline and two-hour assessment, and for the SDMT and Trails A and B, test scores from the baseline and 48-hour assessments were used. All of the test-retest reliability calculations are reported in Table 3. As can be seen from the table, the sample sizes varied among tests. This was due to the lack of complete data for all of the control participants. Participants were excluded in these analyses if they had not been administered the test at either baseline or the first post-concussion interval. Also, they were excluded if their performance at the first retest interval was two standard deviations or greater from the mean at baseline.

With the exception of the results of the HVLT-R, each of the test-retest reliability figures was within generally acceptable limits (near or above .70; Sattler, 2001, p. 102). Because the HVLT-R figure was well below the acceptable level ($r=.15$), the test-retest reliability figure from the HVLT-R manual was used to calculate the RCIs to follow, however, the sample was dissimilar to that of the current study because the participants were not athletes (Brandt and Benedict, 2001).

Further examination of the HVLT-R data for controls was illuminating. Although six controls displayed reliable increases in scores from baseline to the 2-hour post-concussion time point, three decreased (though not reliably) and two others displayed no change in score between the two time points. Thus there was a great deal of variability in this small ($n = 16$) sample of controls, which helps to explain the absence of a higher correlation. In one of the few other studies reporting test-retest reliability coefficients on the

HVLT-R in athletes, Barr (2003) reported a coefficient of .54 in 48 high school athletes tested approximately two months apart. Although this value is still below optimal standards for test-retest reliability coefficients, it is still much higher than what we found in our collegiate athletes. It may be that the shorter test-retest time interval in Barr's study can account for the higher coefficient. Bruce and Echemendia (2004) reported that control athletes displayed more semantic clustering on the HVLT-R than concussed athletes. It may be that the variability in performance in our control group was due to some of the group learning to employ this semantic clustering strategy at the 2-hour post-concussion time point and thus improving their performance and others not changing in this strategy use. Some may also not have improved due to poor motivation at the second time point, an issue discussed in more detail in another chapter in this volume (Bailey and Arnett, in press).

Table 3. Test-Retest Correlations Comparing Baseline Test Results to Post-Concussion Test Results

Test	*n*	*Pearson r*
Trails A 26	.64	
Trails B	25	.74
HVLT-R 16	.15	
SDMT	25	.70
Stroop 1 21	.69	
Stroop 2 22	.72	

4.4. Absolute Values of Practice Effects

Absolute values - the mean differences between test intervals - are displayed in Table 4. As shown, for both the injured and control athletes, practice effects tended to increase in size by one-week post-injury. Also, the majority of the observed practice effects appeared to be greater for the control athletes, a finding consistent with that reported by Echemendia et al. (2001).

Table 4. Absolute Values (Mean Differences) of Practice Effects[1]

Measure	*Means*	*Baseline-48 hr*	*Baseline-1 wk*

Hopkins Recall	Injured	.72	1.64
Total Learning	Control	2.28	3.5
SDMT Total	Injured	2.3	10.13
Correct	Control	3.52	13.52
Stroop 1 # Correct	Injured	-.22	-.11
Per Second[2]	Control	-.04	-.05
Stroop 2 # Correct	Injured	.09	.17
Per Second[2]	Control	.09	.16
Trails A Time[2]	Injured	-2.57	-6.07
	Control	-3.85	-6.16
Trails B Time[2]	Injured	-3.57	-13.09
	Control	-1.51	-6.68

[1] Calculations are completed by subtracting the baseline score from the retest score (Retest-baseline=mean difference)
[2] Trails A and B are timed tests, and decreases indicate an improvement in performance, whereas the negative values for the Stroop 1 test indicates a decline in performance.

4.5. Effect Sizes

Effect size is a commonly used measure to identify the magnitude of the difference between means. Cohen and Cohen (1983) describe this method, which has been used to identify clinically significant change in neuropsychological performance (Dikmen et al., 1999). Suggested cutoff scores for effect sizes are as follows: Small—greater than .2; medium—greater than .5; large—greater than .7.

Table 5 displays the effect size measurements for both injured and control athletes. In the case of controls, effect sizes went from medium to large from baseline to 48-hours and baseline to one-week post-injury on the HVLT-R, Stroop 2, and Trails A, from small to large on the SDMT, and from minimal to medium on Trails B. No practice effects were evident for controls on the Stroop 1 at either time point. Thus, with the exception of this latter test, all effect sizes were at least medium and usually large by one-week post-injury. For the concussed athletes, effect sizes went from either small or medium at 48-hours to large at one-week post-injury on the SDMT, Stroop 2, and Trails A and B. The effect size went from negligible to medium on the HVLT-R. Small but negative effect sizes (reflecting declines in performance) were evident on the Stroop 1 at both time points in the concussed athletes. To sum up, by one-week post-injury, regardless of injury status, the magnitude of the change in scores was either medium or large except for the Stroop 1.

Table 5. Effect Sizes for Injured and Control Athletes

Measure		*Retest Interval*	*Effect Size*
Hopkins Recall Total Learning	Injured	48 hr	-0.19
		1 wk	0.44[2]
	Control	48 hr	0.61[2]
		1 wk	0.93[3]
SDMT Total Correct	Injured	48 hr	0.32[2]
		1 wk	1.39[3]
	Control	48 hr	0.48[1]
		1 wk	1.86[3]
Stroop 1 Correct Per Second	Injured	48 hr	-0.45[1]
		1 wk	-0.21[1]
	Control	48 hr	-0.11
		1 wk	-0.08
Stroop 2 Correct Per Second	Injured	48 hr	0.54[2]
		1 wk	1.04[3]
	Control	48 hr	0.59[2]
		1 wk	0.98[3]
Trails A Time	Injured	48 hr	0.44[1]
		1 wk	1.05[3]
	Control	48 hr	0.66[2]
		1 wk	1.06[3]
Trails B Time	Injured	48 hr	0.30[1]
		1 wk	1.10[3]
	Control	48 hr	0.13
		1 wk	0.56[2]

[1]Small effect size
[2]Medium effect size
[3]Large effect size

Note. Positive effect size values reflect improvements in performance by the group and negative effect size values reflect declines in performance

---

4.6. RCI Calculations

The RCI is commonly used to identify clinically significant change. The index is often used to identify changes in psychotherapy at pre-treatment and post-treatment (Jacobson and Truax, 1991), but has also been used by several researchers to identify clinically significant change on neuropsychological tests. In this study, the 90% confidence interval (change from baseline needed for clinically significant change) was used for both RCI methods and cutoff scores for change. Although the confidence intervals are no different for the RCI and RCI practice methods, the RCI practice method accounts for practice effects by subtracting the mean practice effect from the calculated difference between baseline and retest scores. For both the RCI and RCI Practice calculations that follow, the possibility of regression to the mean was evaluated per Speer's (Speer, 1992) guidelines by correlating the difference between baseline and the first retest interval raw scores with the baseline raw score (See Appendix A). Regression to the mean was present only on the Trails A and B tests. As a result, the true adjusted scores were used only on the Trails A and B tests in the calculation of the RCI and RCI Practice results.

Table 6 shows the results of the RCI calculations. As illustrated, the majority of participants in both groups showed no reliable change at both 48-hours and one-week post-injury. Nonetheless, practice effects (improvements in performance) were apparent for the HVLT-R, SDMT, Stroop 2, and Trails B for some of the controls at 48 hours with additional control participants showing practice effects at one week. There was no notable occurrence of practice effects for the controls on Stroop 1 or Trails A at either retest interval; in fact, nearly all (93% at both time points) of controls displayed no change from baseline on the Stroop 1, and 96% and 93% of controls displayed no evidence of practice effects at the two time points, respectively. Similar to controls, a reasonably large percentage of injured athletes exhibited clinically significant practice effects on the SDMT, Stroop 2, and Trails B at 48 hours, and also like controls, additional participants showed practice effects by one-week post-injury on these tasks. Additionally, approximately one-third more concussed athletes displayed clinically significant improvement on Trails B at one-week post-injury compared with the controls. Clinically significant decreases were exhibited by a notably larger percentage of concussed athletes compared with controls at 48-hours post-injury on the HVLT-R, the SDMT, and the Stroop 1; these discrepancies largely washed out by one-week post-injury.

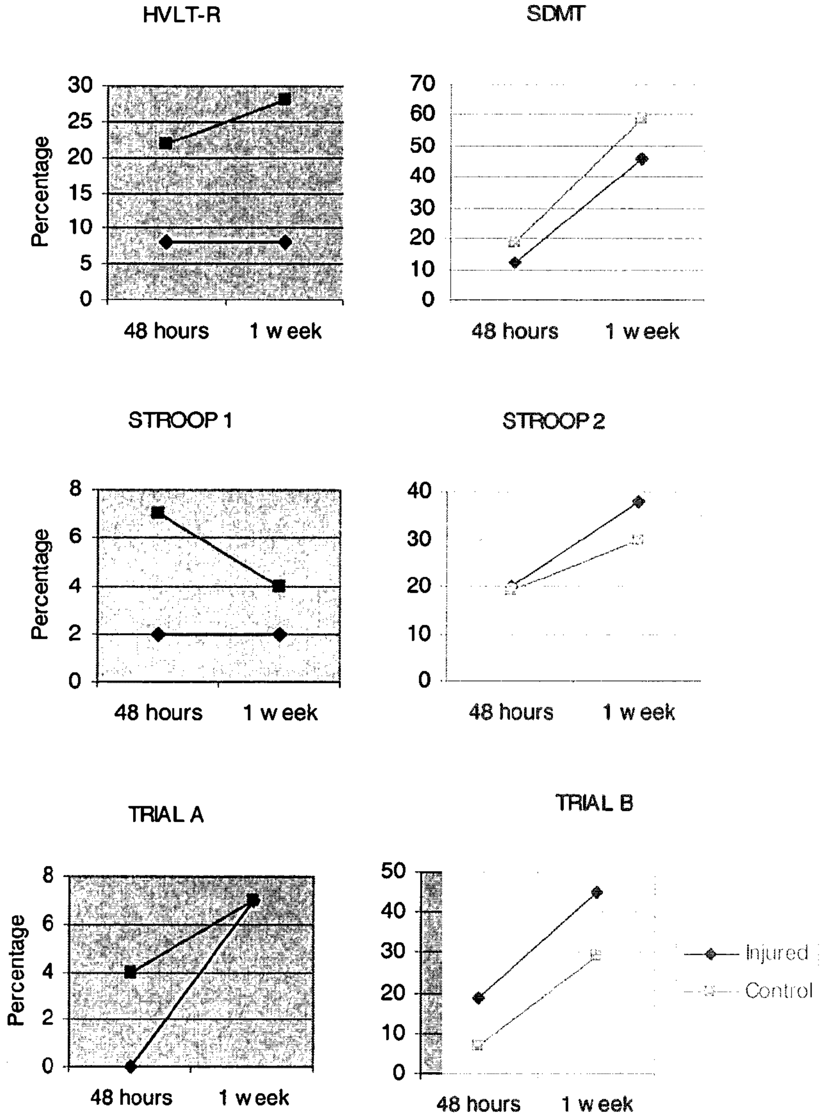

Fig. 1. Percentage of Participants Showing Practice Effects By Group, Test, and Time Point Using RCI Calculations.

Note: HVLT-R = Hopkins Verbal Learning Test –Revised, SDMT = Symbol Digit Modalities Test.

Chi-Square tests were conducted on the two groups at 48-hours and one-week post-injury and then followed up with Fisher's Exact tests to compare only the increased and decreased participants in each group. The only Chi-

Square and Fisher's Exact tests that exceeded traditional levels of statistical significance ($p < .05$) occurred for the Stroop 1 at 48 hours. As shown in Table 6, these effects were primarily due to the fact that significantly more concussed athletes displayed declines in performance compared with the control group (X^2 (2, N = 83) = 6.17, $p < .05$; Fisher's Exact (n = 12) $p < .05$). It is also noteworthy that less than 4% (3/83) of participants (1 concussed and 2 control subjects) showed any evidence of a practice effect (improved performance) at 48 hours. This trend persisted to one-week post-injury where only 2% (2/83) of participants (1 concussed and 1 control subject) displayed any evidence of a practice effect. Although no evidence of differential performance between groups was noted on Trails A, it is nonetheless worth highlighting that there was minimal evidence of practice effects on this test as well. Only 1% (1/87) of participants showed practice effects at 48 hours, and then only 7% (6/87) displayed practice effects at one-week post-injury. If consideration of controls can be thought of as the purest way to measure practice effects, by the one-week post-injury time point, all other tests revealed evidence of significant practice effects with 28% of controls improving reliably on the HVLT-R, 59% on the SDMT, 30% on Stroop 2, and 29% on Trails B. Figure 1 illustrates these results.

Table 6. RCI Calculations of Increases, Decreases, and No Change

		RCI Calculations					
		48 hours			1 week		
		same	increase	decrease	same	increase	decr.
HVLT-R	Injured	29/36 (81%)	3/36 (8%)	4/36 (11%)	31/36 (86%)	3/36 (8%)	2/36 (6%)
	Control	14/18 (78%)	4/18 (22%)	0/18 (0%)	13/18 (72%)	5/18 (28%)	0/18 (0%)
SDMT	Injured	46/59 (78%)	7/59 (12%)	6/59 (10%)	31/59 (53%)	27/59 (46%)	1/59 (2%)
	Control	21/27 (78%)	5/27 (19%)	1/27 (4%)	11/27 (41%)	16/27 (59%)	0/27 (0%)
Stroop 1	Injured	46/56 (82%)	1/56 (2%)	9/56 (16%)	53/59 (90%)	1/59 (2%)	5/59 (8%)
	Control	25/27 (93%)	2/27 (7%)	0/27 (0%)	25/27 (93%)	1/27 (4%)	1/27 (4%)
Stroop 2	Injured	42/56 (75%)	11/56 (20%)	3/56 (5%)	35/58 (60%)	22/58 (38%)	1/58 (2%)
	Control	22/27 (81%)	5/27 (19%)	0/27 (0%)	18/27 (67%)	8/27 (30%)	1/27 (4%)
Trails A	Injured	50/59	0/59	9/59	50/59	4/59	5/59

		(85%)	(0%)	(15%)	(85%)	(7%)	(8%)
	Control	24/28 (86%)	1/28 (4%)	3/28 (11%)	25/28 (89%)	2/28 (7%)	1/28 (4%)
Trails B	Injured	41/58 (71%)	11/58 (19%)	6/58 (10%)	31/58 (53%)	26/58 (45%)	1/58 (2%)
	Control	23/28 (82%)	2/28 (7%)	3/28 (11%)	18/28 (64%)	8/28 (29%)	2/28 (7%)

4.7. RCI Practice Calculations

The RCI for practice effects is a formula that is used to detect statistically significant changes that occur on measures that have been administered at least twice to the same individual. This procedure is intended to identify performance at retest that is indicative of a significant improvement or deterioration in functioning that surpasses the mean practice effect and is not due to test error. As noted, a few studies have used this method to identify reliable changes on neuropsychological measures (Chelune et al., 1993; Dikmen et al., Temkin et al., 1999). As with the RCI method, the standard error was calculated based on the standard deviation of the controls at baseline. Clinically significant increases and decreases were defined in the same manner as the RCI method using the 90% confidence interval.

As seen in Table 7, a majority of concussed and control athletes displayed no evidence for reliable change in scores after accounting for unreliable changes and expected practice effects at either 48-hours or one-week post-injury. Still, more controls compared with concussed athletes exhibited reliable improvements in performance beyond expected practice effects at 48 hours (17% to 3%) and one week (17% to 0%) on the HVLT-R, and more concussed athletes displayed declines in performance relative to controls at 48 hours. These differences were reflected in statistical trends for the Fisher's Exact test comparing the increased and decreased concussed and control groups at 48 hours ($n = 19$, $p < .10$) and one week ($n = 11$, $p = .06$). Additionally, the overall Chi-Square at one-week post-concussion was statistically significant, X^2 $(2, N = 54) = 6.44$, $p < .05$. At one week, a larger percentage of controls compared with concussed athletes (30% versus 10%) displayed evidence of reliable improvements beyond practice effects on the SDMT, a difference reflected in the Fisher's Exact Test for the increased and decreased groups ($n = 39$, $p < .10$), and the Chi-Square test, X^2 $(2, N = 86) = 5.19$, $p < .10$. Although the groups were not notably different, 10% or more of concussed athletes displayed reliable improvements beyond practice effects at 48 hours on Trails B, and at one-week post-injury on the Stroop 2. Compared with controls, notably more concussed athletes displayed improvements beyond practice effects on the Stroop 2 at 48 hours (14% versus 4%) and on Trails B (21% versus 7%) at one-week post-injury. As far as reliable declines in performance at 48 hours, more concussed athletes displayed declines beyond expected practice effects compared with controls on the HVLT-R (33% versus 17%), SDMT (19% versus 4%), Stroop 1 (11% versus 0%), Stroop 2 (14% versus 0%), and Trails A (42% versus 25%). By one-week post-injury, however, there were fewer participants in each group who showed evidence of reliable decline from baseline on these same tests after accounting for practice effects. Although, a notably larger percentage of concussed athletes were significantly below baseline at 48-hours post-injury compared with controls, it appears that these differences were largely eliminated by one-week post-injury. These relative changes are consistent with the sports concussion literature that has shown that a large majority of concussed athletes return to baseline cognitive functioning by 7-10 days post-injury (Berlanger et al., 2005; Echemendia et al., 2001; Lovell et al., 1999). A final highlight in the data is that on the Stroop 2 at 48 hours, more concussed athletes changed in their performance compared with controls, X^2 $(2, N = 83) = 7.16$, $p < .05$, with 14% increasing and 14% declining in performance compared with only 4% and 0% of controls, respectively.

Table 7. RCI Practice Calculations of Increases, Decreases, and No Change

		RCI Practice Calculations					
		48 hours			1 week		
		same	increase	decrease	same	increase	decr.
HVLT-R	injured	23/36 (64%)	1/36 (3%)	12/36 (33%)	30/36 (83%)	0/36 (0%)	6/36 (17%)
	control	12/18 (67%)	3/18 (17%)	3/18 (17%)	13/18 (72%)	3/18 (17%)	2/18 (11%)
SDMT	injured	43/59 (73%)	5/59 (8%)	11/59 (19%)	35/59 (59%)	6/59 (10%)	18/59 (31%)
	control	23/27 (85%)	3/27 (11%)	1/27 (4%)	12/27 (44%)	8/27 (30%)	7/27 (26%)
Stroop 1	injured	50/56 (89%)	1/56 (2%)	5/56 (9%)	53/59 (90%)	1/59 (2%)	5/59 (8%)
	control	25/27 (93%)	2/27 (7%)	0/27 (0%)	25/27 (93%)	1/27 (4%)	1/27 (4%)
Stroop 2	injured	40/56 (71%)	8/56 (14%)	8/56 (14%)	45/58 (78%)	7/58 (12%)	6/58 (10%)
	control	26/27 (96%)	1/27 (4%)	0/27 (0%)	20/27 (74%)	4/27 (15%)	3/27 (11%)
Trails A	injured	34/59 (58%)	0/59 (0%)	25/59 (42%)	38/59 (64%)	0/59 (0%)	21/59 (36%)
	control	21/28 (75%)	0/28 (0%)	7/28 (25%)	20/28 (71%)	0/28 (0%)	8/28 (29%)

Trails B	injured	42/58 (72%)	10/58 (17%)	6/58 (10%)	44/58 (76%)	12/58 (21%)	2/58 (3%)
	control	20/28 (71%)	3/28 (11%)	5/28 (18%)	23/28 (82%)	2/28 (7%)	3/28 (11%)

5. DISCUSSION

Neuropsychological testing is now commonly used in post-concussion assessments in competitive sports. The results of such testing are often critical in determining return to play decisions. The state of the art in neuropsychological assessment of sports-related concussion involves conducting baseline testing when athletes begin their tenure with a particular sports program. If they then experience a concussion at some point during competition, they are typically re-tested with alternate forms of the same tests they received at baseline to determine whether they have experienced a decline in performance. Any significant decline in performance is thought to reflect the cognitive manifestation of the mild traumatic brain injury the athlete has suffered. Although such an approach seems straightforward, a number of factors can interfere with accurate assessment in such situations. One critical factor identified in this chapter is practice effects. Because many of the clinical neuropsychological tests used to measure the effects of sports-related concussion are susceptible to practice effects, accurate assessment of sports-related concussion can be challenging.

The study presented in this chapter examined several issues related to practice effects on neuropsychological tests in an elite group of collegiate athletes. It is important to underscore the fact that the athletes in this study were given the tests multiple times in one week, but this is typical in sports-related concussion neuropsychological testing. One purpose of the study was to calculate and report the absolute value of practice effects on four (six test indices) commonly used neuropsychological tests that are thought to be sensitive to concussion in athletes. The data indicated that, with the exception of the Stroop 1, practice effects were apparent in both concussed and control athletes on most tasks. At 48 hours, small to medium effect sizes for change from baseline were evident in both concussed and control athletes on the SDMT, Stroop 2, and Trails A. The controls also displayed a medium effect size on the HVLT-R and the concussed athletes showed a small effect size for Trails B. By one-week post-injury, regardless of injury status, again with the exception of the Stroop 1, medium but mostly large effect sizes for practice effects from baseline performance were found. The effect size relative to baseline on the Stroop 1 for controls was negligible, suggesting no practice effect, and concussed athletes displayed a small effect size reflecting a mild decline from baseline at both time points.

The results for the RCI calculations provided more information about change at the individual athlete level, but the results essentially mirrored the effect size results. When considering only the athlete controls, by one-week post-injury all tests except the Stroop 1 revealed evidence for significant practice effects; 28% of controls improved reliably on the HVLT-R, 59% on the SDMT, 30% on Stroop 2, and 29% on Trails B. For the Stroop 1 by contrast, 93% of controls displayed no evidence for reliable change at either time point; only 7% showed practice effects at 48 hours and only 4% at one-week post-injury. Importantly, the Stroop 1 was the only task to reveal statistically significant effects when non-parametric tests were applied to control and concussed groups displaying increases and decreases in performance at 48 hours. These effects were primarily due to the fact that significantly more concussed athletes displayed declines in performance compared with the control group. Although no evidence of differential performance between groups was noted on the Trails A test at either time point, it is nonetheless worth highlighting that there was minimal evidence of practice effects on this test as well. Only 1% of participants showed practice effects at 48 hours, and then only 7% displayed practice effects at one-week post-injury.

The effect size and RCI results for the Stroop 1 are similar to the Stroop A results presented by Hinton-Bayre and Geffen (2004) in the study described earlier. Their Stroop A task was like our Stroop 1 where examinees simply needed to read color words on a page as quickly as possible. From baseline to two days post-concussion, they found no evidence for practice effects in control athletes. In contrast, these investigators found that concussed athletes displayed significantly slower performance at two days post-concussion compared with baseline, results that mirrored our 48-hour post-concussion findings for Stroop 1. Our data extend these findings by demonstrating them not only in terms of mean group comparisons, but also using RCI methodology. Our finding that Trails A revealed minimal practice effects even on repeated administration mirrored Iverson and Gaetz's (2004) findings of a similar nature. However, our results contrast with those of Guskiewicz and colleagues (2001) and Macciocchi and colleagues (1996) who reported some evidence for practice effects on Trails A.

The results for the RCI Practice calculations were also illuminating. The magnitude of practice effects was attenuated for most tests, a finding that was not surprising given that the RCI Practice calculations are designed to control for such practice effects. In contrast to the RCI without practice calculations, where 28% or more of athlete controls improved reliably on the HVLT-R, SDMT, Stroop 2, and Trails B by the one-week post-concussion time point, only the SDMT revealed practice effects at this magnitude (30%) for the RCI Practice calculations. Otherwise practice effects were above 10% for controls only on the HVLT-R (17%) and the Stroop 2 (15%) at one week. Similar to the effect size and RCI calculations, practice effects for the Stroop 1 were negligible at both 48 hours and one week for both controls and concussed athletes. Also like the RCI calculations, practice effects for Trails A were negligible at both time points in both groups for the RCI Practice calculations. As far as reliable declines in performance at 48 hours detected using the RCI Practice calculation, notably larger percentages of concussed athletes displayed declines beyond expected practice effects compared with controls on five of the six test indices (the only exception was Trails B). By one-week post-injury, however, similar percentages of participants in each group showed evidence of reliable decline from baseline on these same indices. Thus, although a notably larger percentage of concussed athletes were significantly below baseline at 48-hours post-injury compared with controls, these differences resolved by one-week post-injury. Again, such relative changes mirror much of the sports concussion literature where a large majority of concussed athletes have been shown to return to baseline cognitive functioning by 7-10 days post-injury (Berlanger et al., 2005; Echemendia et al., 2001; Lovell et al., 1999).

Another purpose of this chapter was to present test-retest reliability calculations for the four neuropsychological instruments in a sample of young athletes. The test-retest reliability coefficients for all of the tests except the HVLT-R were within the range of those reported in previous studies of non-athletes. The reliability coefficient for the HVLT-R was well below the range of acceptability. As noted earlier, this low reliability coefficient may have been due to the particularly low sample size of the HVLT-R in our sample (n=16) compared with the other measures, a high degree of variability in change in the controls in this sample that may have been due to some participants increasing their use of semantic clustering of words between test-retest points, a longer test-retest interval than has been reported in one of the few other studies examining test-retest reliability of the HVLT-R in athletes (Barr, 2003), and possibly motivational differences between time points. Despite the problems with the low test-retest reliability of the HVLT-R in our sample, it appears that the test-retest reliability obtained from a sample of young athletes is comparable to that of other samples of individuals for the other test indices.

5.1. Clinical Considerations

As shown with the data we presented in this chapter, practice effects are extremely common when athletes are administered the same tests (or alternate forms of the same tests) multiple times over a relatively short period of time. Even concussed athletes are susceptible to such effects, especially by one-week post-injury. Specifically, our effect size data showed that concussed athletes were significantly improved from their baseline test performance by one-week post-injury. That is, overall, these athletes who had suffered a concussion significant enough to result in their removal from play for one week or more were actually performing better than they were at baseline by one-week post-concussion. Of course it would not make sense to suggest that the cognitive functioning of these recently concussed athletes was actually improved over where it was at baseline. It seems most likely that they performed better than baseline due to practice effects involved with the tests. Nonetheless, it is still important to note that injured athletes were not able to benefit from practice at the 48-hour mark post-concussion when compared with controls, which brings up a significant clinical consideration. Practice effects, by definition, reflect a learning phenomenon. The overwhelming evidence in the sports concussion literature suggests that concussed athletes have difficulty learning and consolidating new learning. The present data suggest that injured athletes do not benefit as much from prior exposure to tests as their matched controls at the 48-hour retest. In other words, they do not learn as readily as their non-concussed counterparts. If this pattern is replicated in other samples, it suggests that the absence of a practice effect at 48-hours post-concussion provides significant clinical data that the athletes' neurocognitive functioning is below expectation, even if they are functioning at baseline levels. In contrast, this pattern does not appear at one-week post-injury where the concussed athletes appeared to have "caught up" with the control group. Therefore, practice effects appear to have differential clinical utility depending on the time elapsed from injury.

Taken together, these data argue that the use of RCI with correction for practice effects may be the most scientifically valid approach to the interpretation of neuropsychological test scores following concussive injury, with the recognition that RCI practice scores will vary depending on time from injury and the number of previous test exposures. It is also important to underscore that reliable declines from baseline on the clinical neuropsychological tests reviewed were rare for athlete controls, especially at the one-week post-concussion time point (0% to 7% across the six test indices using the RCI calculation). As such, it is reasonable to assume that a concussed athlete who still displays a reliable decline from baseline (using the RCI calculation with a 90% confidence interval) at one-week post-concussion is still experiencing the cognitive effects of that concussion and likely should not be returned to play. Although, to our knowledge no published algorithm has been developed to provide guidelines for how many tests reliably below baseline post-concussion warrant making the recommendation that an athlete not return to play, consideration of our athlete control data at one week suggests that performance below baseline on even one test index might make one proceed cautiously. Again, given the rarity of reliable declines from baseline in our athlete controls on the clinical tests reviewed, if an athlete one-week post-concussion were to perform reliably below baseline on two or more test indices, it would seem reasonable to suggest that the athlete not return to play until further time had passed for more recovery to take place.

Regarding tests that were most resistant to practice effects, the Stroop 1 appeared to be the best. Regardless of the measure of change used—effect size, RCI, or RCI Practice—the Stroop 1 appeared very resistant to practice effects. Trails A was also very resistant to practice effects in both athletes and controls, at least when RCI or RCI Practice calculations were used. Importantly, the Stroop 1 also showed some differential sensitivity to concussion at 48 hours, with more injured athletes performing reliably below baseline compared with control athletes. Thus the Stroop 1, at least from the perspective of practice effects, appears to be an excellent measure. It was resistant to practice effects even after four administrations, three of which occurred over a one-week period (2-hour, 48-hour, and one-week post-concussion periods), but still showed some sensitivity to concussion. It is also reassuring that Hinton-Bayre and Geffen (2004) reported comparable findings for a similar Stroop task examining mean changes over time. Trail A was also resistant to practice effects, but was only shown to be differentially sensitive to concussion at 48 hours using the RCI Practice formula.

Why were the Stroop 1 and Trails A tests so resistant to practice effects? One reason might be the relatively automatic nature of each test. In the case of the Stroop 1, examinees are simply required to read columns of color words as quickly as possible. Reading simple words, for most individuals, is a highly automated task because it is engaged in so frequently and has been practiced over many years. In the case of Trails A, examinees simply need to connect numbered circles (1-25) in sequence. Tasks that involve simple numbered sequences are also highly automated. Both Trails A and Stroop 1 require simple processing speed. As has been shown with computerized tests, simple processing speed does not generate significant practice effects when compared to more effortful or complex tasks. With these considerations in mind, it may be that both the Stroop 1 and Trails A were resistant to practice effects because, in essence, they are well-practiced speed-dependent tasks providing less room for the influence of practice effects. All of the other tasks are relatively novel and complex, so they are more likely to improve with practice because their novelty is decreased with each new presentation and as a result examinees become more proficient at them. Many of these clinical neuropsychological tasks are commonly chosen because their novelty makes them sensitive to the effects of concussion, but this very novelty also makes them most susceptible to practice effects. With these considerations in mind, it might be ideal to develop more extensive batteries that include tasks like the Stroop 1 and Trails A that are resistant to practice effects but have at least some sensitivity to concussion.

5.2. Limitations

As noted earlier, our athlete controls had significantly higher overall SAT scores compared with the concussed group. Given, however, that the groups did not differ significantly on their actual baseline test scores for the different neuropsychological test indices, such a difference most likely did not have much of an influence on our results. Still, efforts should be made in future research to match control and concussed athletes on SAT scores or other measures of global cognitive functioning. The small sample size used in the current study is certainly a limitation. The control group was especially small, and given that the data from the control group were used to calculate many of the indices and were used heavily in the interpretation of the results, larger samples of non-injured control athletes should be used in future studies examining practice effects. Furthermore, the sample size of the injured athletes was not especially large, and a larger sample would increase the power of the results. Nonetheless, the sample sizes of the groups used in this study are comparable to or surpass those used in most sports neuropsychology studies conducted in the past.

CONCLUSIONS

In recent years there has been a dramatic increase in the use of neuropsychological tests to evaluate the effects of concussion in competitive athletes and assist in return to play decisions. In this chapter, we have focused on one factor that can limit the sensitivity of neuropsychological tests to concussion--practice effects. The data we present suggests that the HVLT-R, Trails B, Stroop 2, and SDMT are most susceptible to practice effects upon repeated administration. Nonetheless, we show that even for these tests, a majority of control athletes do not show evidence for significant practice effects after several administrations of the tests when the reliability of the measures and regression to the mean are controlled for. Still, the fact that a significant minority of athletes show practice effects on these tests should serve as a note of caution for interpreting these commonly used clinical neuropsychological tests post-concussion. In contrast, the Stroop 1 and Trails A showed little evidence for practice effects even when administered several times. Because the Stroop 1 also showed evidence for sensitivity to concussion, it emerged as perhaps the best test in terms of combined resistance to practice effects and concussion sensitivity. Sports neuropsychologists would do well to try and identify other such tests for use in this arena.

In terms of return to play decisions, because we found that a negligible number of controls (0 to 7%) displayed evidence for reliable decline from baseline on all six test indices (using the RCI formula), the data we present in this chapter strongly suggest that when concussed athletes continue to show performance reliably below baseline at one-week post-concussion on any of the noted test indices, great caution should be exercised in recommending return to play. Additionally, although no widely accepted algorithm currently exists for making return to play decisions, given our data, any athlete who is still reliably below baseline on two of the test indices at one-week post-concussion should not return to play because residual persisting cognitive effects from the concussion are highly likely. Future work can extend this research by using larger samples, better matching on overall cognitive ability, and evaluating additional commonly used neuropsychological tests and emerging computerized batteries. This exciting field of sports neuropsychology presents many future opportunities for exploration that will surely lead to improved understanding of the nature of sports-related concussion, and ultimately better and more informed care for athletes who suffer such injury.

REFERENCES

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

Kelly, J. P., Rosenberg, J.H. (1997). Diagnosis and management of concussion in sports *Neurology, 48*(3), 575-580.

Guskiewicz, K. M., Weaver, N.L., Padua, D.A., and Garrett, W.E. Jr. (2000). Epidemiology of concussion in collegiate and high school football players. *American Journal of Sports Medicine, 28*(5), 643-650.

Gerberich, S. G., Priest, J. D., Boen, J. R., Straub, C. P., and Maxwell, R. E. . (1983). Concussion incidences and severity in secondary school varsity football players. *American Journal of Public Health, 73*(12), 1370-1375.

National Institutes of Health Consensus Development Panel on Rehabilitation of Persons with Mild Traumatic Brain Injury (1999). Rehabilitation of persons with mild traumatic brain injury. *Journal of the American Medical Association, 282*(10), 974-982.

Thurman, D. J., Branche, C. M., & Sniezek, J. E. (1998). The epidemiology of sports related traumatic brain injuries in the United States: Recent developments. *Journal of Head Trauma Rehabilitation, 13*(2), 1-8.

Echemendia, R. J., and Cantu, R.C. (2004). Return to Play Following Cerebral Brain Injury. In M. R. Lovell, Collins, M.W., Echemendia, R.J., and Barth, J.T. (Ed.), *Traumatic Brain Injury in Sports* (Vol. 1, pp. 479-498). New York: Taylor and Francis.

Alves, W., Macciocchi, S. N., and Barth, J. T. (1993). Postconcussive symptoms after uncomplicated mild head injury. *Journal of Head Trauma Rehabilitation, 8*(3), 48-59.

Berlanger, H. G., Curtiss, G., Demery, J.A., Lebowitz, B.K., and Vanderploeg, R.D. . (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society, 11*, 215-227.

Cantu, R. C. (2001). Postraumatic retrograde and anterograde amnesia: Pathophysiology and implications in grading and safe return to play. *Journal of Athletic Training, 36*(3), 244-248.

Gouvier, W. D., Cubic, B., Jones, G., Brantley, P., and Cutlip, Q. (1992). Postconcussion symptoms and daily stress in normal and head-injured college populations. *Archives of Clinical Neuropsychology, 7*, 193-211.

Guskiewicz, K. M., Riemann, B. L., Perrin, D. H., and Nashner, L. M. (1997). Alternative approaches to the assessment of mild head injury in athletes. *Medical Science and Sports Exercise, 29*(Suppl. 7), 5213-5221.

Zasler, N. (1999). Medical aspects. In S. A. a. M. Raskin, C. A. (Ed.), *Neuropsychological management of mild traumatic brain injury* (pp. 23-38). New York: Oxford University Press.

Echemendia, R. J., and Julian, L.J. (2001). Mild traumatic brain injury in sports: Neuropsychology's contribution to a developing field *Neuropsychology Review, 11*(2), 69-88.

Lovell, M. R., Iverson, G. L., Collins, M. W., McKeag, D., and Maroon, J. C. (1999). Does loss of consciousness predict neuropsychological decrements after concussion? *Clinical Journal of Sports Medicine, 9*(4), 193-198.

Cantu, R. C. (1998). Second-impact syndrome. *Clinics in Sports Medicine: Neurologic Athletic Head and Neck Injuries, 17*(1), 37-44.

Chelune, G. J., Naugle, R.I., Luders, H., Sedlak, J., and Awad, I.A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7*(1), 41-52.

Iverson, G. L. and Gaetz, M. (2004). Practical Considerations for Interpreting Change Following Brain Injury. In M. R. Lovell, Collins, M.W., Echemendia, R.J., and Barth, J.T. (Ed.), *Traumatic Brain Injury in Sports* (Vol. 1, pp. 323-356). New York: Taylor and Francis.

Mackin, R. S., Sabsevitz, D.S., Julian, L., Junco, R., Dwyer, M. and Echemendia, R.J. (1997). *Stability Coefficients and Practice Effects of Neuropsychological Tests in College Athletes: Preliminary Findings.* Paper presented at the Sports Related and Nervous System Injuries Orlando, Florida.

Barth, J. T., Alves, W.M., Ryan, T.V., Macciocchi, S.N., Rimel, R.W., Jane, J.A., and Nelson, W.E. (1989). Mild head injury in sports: Neuropsychological sequelae and recovery of function. In H. E. H. Levin, and A. Benton (Ed.), *Mild Head Injury* (pp. 257-275). New York, NY: Oxford University Press.

McCrea, M., Barr, W.B., Guskiewicz, K., Randolph, C.R., Marshall, S.W., Cantu, R., Onate J.A., and Kelly, J.P. (2005). Standard regression-based methods for measuring recovery after sport-related concussion. *Journal of the International Neuropsychological Society, 11*, 58-69.

Dikmen, S. S., Heaton, R.K., Grant, I., and Temkin, N.R. (1999). Test-retest reliability and practice effects of Expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society, 5*(4), 346-356.

Jacobson, N. S., and Truax., P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19.

Barr, W. B., and McCrea, M. (2001). Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion *Journal of the International Neuropsychological Society, 7*(6), 693-702.

Hinton-Bayre, A. D., Geffen, G.M., Geffen, L.B., McFarland, K.A., and Friis, P. (1999). Concussion in contact sports: Reliable change indices of impairment and recovery *Journal of Clinical and Experimental Neuropsychology, 21*(1), 70-86.

Iverson, G. L., Lovell, M.R., and Collins, M.W. (2003). Interpreting Change on ImPACT Following Sport Concussion. *Clinical Neuropsychologist, 17*(4), 460-467.

Brandt, J., and Benedict, R.H.B. (2001). *Hopkins Verbal Learning Test-Revised professional manual* Odessa, FL: Psychological Assessment Resources.

Trennery, M. R., Crosson, B., DeBoe, J., and Leber, W.R. (1989). *Stroop Neuropsychological Screening Test Manual.* Odessa, FL: Psychological Assessment Resources.

Reitan, R. M. (1958). Validity of the Trail-Making Test. *Perceptual and Motor Skills, 8*, 271-276.

Smith, A. (1982). *Symbol Digit Modalities Test Manual (Revised).* Los Angeles: Western Psychological Services.

Echemendia, R. J. (1997). Neuropsychological assessment of the college athlete: The Penn State concussion program. *Neuropsychology, 10*(2), 189-193.

Basso, M. R., Bornstein, R.A., and Lang, J.M. (1999). Practice effects on commonly used measures of executive function across twelve months. *Clinical Neuropsychologist, 13*(3), 283-292.

Bornstein, R. A., Baker, G.B., and Douglass, A.B. (1987). Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. *The Journal of Nervous and Mental Disease, 175*(4), 229-232.

desRosiers, G. a. K., D. (1987). Cognitive assessment in closed head injury: Stability, validity, and parallel forms for two neuropsychological measures of recovery. *The International Journal of Clinical Neuropsychology, 9*(4), 162-173.

Matarazzo, J. D., Wiens, A.N., Matarazzo, R.G., and Goldstein, S.G. (1974). Psychometric and clinical test-retest reliability of the Halstead Impairment Index in a sample of healthy, young, normal men. *The Journal of Nervous and Mental Disease, 158*(1), 37-49.

Craddick, R.A., and Stern, M.R. (1963). Practice effects on the trail making test. *Perceptual and Motor Skills, 17*, 651-653.

Mitrushina, M., and Satz, P. (1991). Effects of repeated administration of a neuropsychological battery in the elderly. *Journal of Clinical Psychology, 47*(6), 790-801.

Dikmen, S. S., Reitan, R.M., and Temkin, N.R. (1983). Neuropsychological recovery in head injury. *Archives of Neurology, 40*(6), 333-338.

Dye, O. A. (1979). Effects of practice on trail making test performance. *Perceptual and Motor Skills, 48*, 296.

Macciocchi, S. N., Barth, J.T., Littlefield, L., and Cantu, R.C. (2001). Multiple concussions and neuropsychological functioning in collegiate football players. *Journal of Athletic Training, 36*(3), 303-306.

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

Guskiewicz, K. M., Ross, S.E., and Marshall, S.W. (2001). Postural stability and neuropsychological deficits after concussion in collegiate athletes. *Journal of Athletic Training, 36*, 263-273.

Davidson, D. J., Zacks, R.T., and Williams, C.C. (2003). Stroop interference, practice, and aging. *Aging Neuropsychology and Cognition, 10*(2), 85-98.

Houx, P. J., Shepherd, J., Blauw, G-J., Murphy, M.B., Ford, I., Bollen, E.L., Buckley, B., Stott, D.J., Jukema, W., Hyland, M., Gaw, A., Norrie, J., Kamper, A.M., Perry, I.J., MacFarlane, P.W., Edo Meinders, A., Sweeney, B.J., Packard, C.J., Twomey, C., Cobbe, S.M., and Westendorp, R.G. (2002). Testing cognitive function in elderly populations: The PROSPER study. *Journal of Neurology, Neurosurgery, and Psychiatry, 73*(4), 385-389.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Golden, C.G. (1978). A manual for clinical and experimental users: Stroop color and word test. Stoelting: Wood Dale, IL.

Connor, A., Franzen, M., and Sharp, B. (1988). Effects of practice and differential instructions on Stroop performance. *The International Journal of Clinical Neuropsychology, 10*(1), 4.

Hinton-Bayre, A. D., and Geffen, G. (2004). Australian rules football and rugby league. In M. R. Lovell, Collins, M.W., Echemendia, R.J., and Barth, J.T. (Ed.), *Traumatic Brain Injury in Sports*. New York: Taylor and Francis.

Franzen, M. D., Tishelman, A.C., Sharp, B.H., and Friedman, A.G. (1987). Test-retest reliability of the Stroop Word Color Test across two intervals. *Archives of Clinical Neuropsychology, 2*, 265-272.

Uchiyama, C. L., D'Elia, L.F., Dellinger, A.M., Selnes, O.A., Becker, J.T., Wesch, J.E., Chen, B.B., Satz, P., van Gorp, W., and Miller, E.N. (1994). Longitudinal comparison of alternate versions of the Symbol Digit Modalities Test: Issues of form comparability and moderating demographic variables. *Clinical Neuropsychologist, 8*(2), 209-218.

Bohnen, N., Twijnstra, A, and Jolles, J. (1992). Post-traumatic and emotional symptoms in different subgroups of patients with mild head injury. *Brain Injury, 6*(6), 481-487.

Erlanger, D., Feldman, D., Kutner, K., Kaushik, T., Kroger, H., Festa, J., Barth, J., Freeman, J., and Broshek, D. (2003). Development and validation of a web-based neuropsychological test protocol for sports-related *return to play* decision-making. *Archives of Clinical Neuropsychology, 18*, 293-316.

Echemendia, R. J., Putukian, M., Mackin, R.S., Julian, L.J., and Schoss, N. (2001). Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. *Clinical Journal of Sports Medicine, 11*, 23-31.

Sattler, J. M. (2001). *Assessment of Children: Cognitive Applications, 4th Edition.* La Mesa, California: Jerome M. Sattler, Publisher, Inc.

Barr, W. B. (2003). Neuropsychological testing of high school athletes: Preliminary norms and test-retest indices. *Archives of Clinical Neuropsychology, 18*, 91-101.

Bruce J.M., a. E., R.J. (2004). Concussion history predicts self-reported symptoms before and following a concussive event. *Neurology, 63*, 1516-1518.

Bailey, C. M., and Arnett, P.A. (in press). Motivation and the assessment of sports-related concussion. In S. S. a. W. Sebastianelli (Ed.), *Foundations of Sports-Related Concussion*. Springer-Verlag.

Cohen, J., and Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences, 2nd Edition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Chelune, G. J., Naugle, R.I., Luders, H., Sedlak, J., and Awad, I.A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7*(1), 41-52.

Temkin, N. R., Heaton, R.K., Grant, I, and Dikmen, S.S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society, 5*(4), 357-369.

APPENDIX A

Effect Size Equation. The equation used to calculate effect size is as follows:

$$ES = \frac{Y_{OBS\ retest} - Y_{OBS\ baseline}}{SD_{baseline}}$$

$Y_{OBS\ retest}$ is the retest raw score of interest (e.g., 48 hour or 1 week post-injury), and $Y_{OBS\ baseline}$ is the raw score at baseline. $SD_{baseline}$ is the standard deviation of the athlete controls at baseline. Following Cohen (1988), the cutoff scores for effect sizes are as follows: small effect size—greater than .2; medium effect size—greater than .5; large effect size—greater than .7.

RCI Equation. The formula for the RCI (Jacobsen & Truax, 1991) is as follows:

$$RCI = \frac{(X_2 - X_1)}{S_{diff}}$$

X_2-X_1 is the change that a participant exhibits at test time 1 and test time 2, where X1 is the participant's score at time 1 (baseline) and X_2 is the participant's score at time 2. S_{diff} is an estimate of the standard error of the difference between baseline and the retest scores. This figure is calculated using the following formulas:

$$S_{diff} = \sqrt{2\,(S_E)^2}$$

$$S_E = S_1 \sqrt{1 - r_{xx}}$$

S_1 is the standard deviation of the scores of the normative sample at testing time 1. r_{xx} is the test-retest reliability coefficient. The reliability coefficient is calculated by obtaining the correlation between scores of a relevant normative sample at time 1 and time 2. In order to calculate the RCI, individuals' scores at baseline and at retest intervals (e.g., 48 hours, and one week) are used. When calculating the standard error (S_E), the standard deviation of the control baseline scores was used.

The result that is obtained from the RCI calculation is based on the difference between the raw scores at baseline and at the retest interval. The RCI is used to convert this difference divided by the standard difference into a standard score. Reliable change is apparent when the RCI is ± 1.64. Thus, after calculating the individual's RCI, the RCI is compared to ± 1.64. If a score is above 1.64, the individual's score is indicative of a clinically significant increase, and a score below -1.64 is indicative of a clinically significant decrease.

RCI Practice Equation. The formula for the RCI for practice effects (Chelune, et al., 1993) is as follows:

$$RCI\ practice = \frac{(X_2 - X_1) - X_{change\ normative}}{S_{diff}}$$

RCI practice is the Reliable Change Index adjusted for practice effects. X_2-X_1 is the change that a participant exhibits between test time 1 and test time 2. $X_{change\ normative}$ is the mean change between time 1 and time 2 for the normative sample. This figure is calculated by subtracting the test score at time 1 from the test score at time 2 for each member of the normative sample and obtaining the mean difference. S_{diff} is an estimate of the standard deviation of the difference between scores at baseline and those at the retest interval. This figure is calculated using the following formulas:

$$S_{diff} = \sqrt{2\,(S_E)^2}$$

$$S_E = S_1 \sqrt{1 - r_{xx}}$$

S_1 is the standard deviation of the scores of the normative sample at testing time 1. r_{xx} is the test-retest reliability coefficient. The reliability coefficient is calculated by obtaining the correlation between scores at test 1 to the score at test 2 for a relevant normative sample.

As with the RCI method, the standard error (S_E) was calculated using the standard deviation of the control athletes at baseline.

The value that is obtained from the RCI for practice calculation is based on the difference between the raw scores at baseline and at the retest interval. The difference between the raw scores is then placed into an equation that subtracts the practice effect from the difference between raw scores. The RCI for practice is used to convert this difference divided by the standard difference into a standard score. Clinical significance may be evident when the RCI is ± 1.64 (90% confidence interval). After calculating the individual's RCI, the RCI is compared to ± 1.64. If a score is above 1.64, the individual's score is indicative of a clinically significant increase, and a score below -1.64 is indicative of a clinically significant decrease. More concretely, only 5% of a normal population would be expected to exceed this cutoff in either direction, so the likelihood of such a score being abnormal, and thus clinically meaningful, is greater. Scores at or below the 5% level of the athlete controls are more likely in the concussed, in the case of the study we present.

Regression to the Mean Equation. For both the RCI and RCI practice, calculations were done to identify regression to the mean. Per Speer's(Speer, 1992) guidelines, regression to the mean was identified by correlating the difference between baseline and the first retest interval raw score with the baseline raw score. The equation is as follows:

$$r = (Y_{OBS\ 1st\ retest} - Y_{OBS\ baseline})(Y_{OBS\ baseline})$$

$Y_{OBS\ baseline}$ is the baseline raw score and $Y_{OBS\ 1st\ retest}$ is the raw score of the first retest interval. Regression to the mean is identified when this correlation is significant. When this occurred, estimated baseline scores adjusted for regression to the mean (mean true adjusted score) were calculated. The formula for calculating the mean true adjusted score is as follows:

$$T_1 = r_{xx}(Y_{obs} - Y_{baseline}) + Y_{baseline}$$

T_1 is the true adjusted score at baseline. r_{xx} is the test-retest reliability score comparing the test score at baseline to the first retest interval. Y_{obs} is the observed score for the participant at baseline. $Y_{baseline}$ is the mean observed score for the sample of control athletes at baseline. In the case that regression to the mean is identified, the true adjusted scores are used to replace the raw baseline scores which are in turn used to calculate the RCI and RCI practice.