2015

Repeat Baseline Assessment in College-Age Athletes

Arthur C. Maerlender
University of Nebraska-Lincoln, amaerlender2@unl.edu

Dennis L. Molfese
University of Nebraska-Lincoln, dmolfese2@unl.edu

Published in *Developmental Neuropsychology* 40:2 (2015), pp. 69–73.
doi: 10.1080/87565641.2015.1014089

Repeat Baseline Assessment in College-Age Athletes

Arthur Maerlender and Dennis L. Molfese

Center for Brain, Biology and Behavior, University of Nebraska–Lincoln, Lincoln, Nebraska

Corresponding author — Arthur Maerlender, University of Nebraska–Lincoln, C86 East Stadium, PO Box 880156, Lincoln, NE 68588. E-mail: amaerlender2@unl.edu

Abstract

Repeat baseline testing scores from one collegiate Division I NCAA school were analyzed to determine the necessity of this practice. ImPACT tests were taken between 13 and 40 months apart (median 24 months; final N = 67). No significant difference in any test composite score was obtained; the number of tests exceeding chance levels of change was insignificant. The results do not support the recommendation for repeating baseline testing in college athletes; replication is recommended.

While the advent of computerized batteries in concussion management has improved efficiency, the administration of baseline neuropsychological tests remains a time consuming practice. ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing) has recommended that baseline be re-administered every two years, thus doubling the number of baseline tests administered every year in a university setting. While this is clearly a positive commercial application, the rationale from a scientific point of view is open to question. The answer hinges on the amount of cognitive change or development athletes experience between the approximate ages of 18 to 22 years. While practice effects are well known to be present in repeat test administrations, the issue of developmental change over and above practice effects has yet to be determined for the ImPACT test.

Despite the ease and efficiency of computerized baseline testing, it has been criticized on the grounds of poor administrative practices (Lichtenstein, Moser, & Schatz, 2014; Moser, Schatz, Neidzwski, & Ott, 2011; Moser, Schatz, & Lichtenstein, 2013; Vaughn, Gerst, Sady, Newman, & Gioia,, 2014; also see Erdal, 2012) and poor validity due to effort ("sandbagging"). To enhance baseline test validity, we employed three sets of validity criteria to this sample. ImPACT provides a set of criteria believed to reflect suboptimal performance. As with any effort measure, individual characteristics must be taken into account before making clinical decisions about validity of results. However, in research studies, these criteria are applied to create a more normal sample.

Schatz and Glatts (2013) conducted an experiment in which two groups of baseline test takers received different instructions about performing badly on the test. The control group was given standard test instructions to "do their best." One poor effort group was simply instructed to try to fail (naïve fail), and the other was instructed to do poorly but so as not to be detected by ImPACT validity criteria (coached fail). Results demonstrated that the coached group scored significantly lower than the control group but not as poorly as the naïve group and in a manner that indicated consistently poor effort. The following indices were believed to be a reflection of invalid effort on the ImPACT: (a) Visual Motor Speed composite score < 25, (b) Reaction Time composite score > 0.80, (c) Word Memory Correct Distractors (Immediate + Delayed) < 22, (d) Design Memory Correct Distractors (Immediate + Delayed)< 16.

In our clinical practice, we also employ a set of additional criteria. When Color Match Total Correct = 0, the Reaction Time score also = 0, thus artificially decreasing the average reaction time score that factors into the Reaction Time composite. Similarly, when Three-Letters Average Counted Correctly is less than 5, it signals that on at least one of five trials the respondent counted forward rather than backward, thus affecting the Visual Motor Speed score negatively. We also question results with Total Symptom scores above 15 as it may signify illness and thus contribute to suboptimal performance.

The present study compared the outcome of ImPACT testing at a large Division I university athletes who had taken ImPACT two times between 2010 and 2014. Cases were removed if either test results exceeded strict validity criteria, were within a retest interval range of 13 to 39 months, and any cases in which a concussion occurred within the test interval. Demographic factors of age, sex and previous concussions were analyzed as potential covariates.

Method

Participants and Procedures

Archived data from varsity athletes from a Division I university who had been tested with ImPACT as part of the Athletic Department concussion management program were analyzed. Athletes were tested in groups by athletic training staff in quiet computer labs within the Athletic Department complex, or in the Center for Brain, Biology and Behavior assessment lab. There was variability in testing procedures, with the majority of testing completed in large groups with two to five supervisors and non-standardized instruction as is fairly typical (Vaughn et al., 2014). The original sample of baseline results from 2010 to 2014 included 899 records. Invalid baselines were determined by ImPACT software ("Baseline ++"), the criteria published by Schatz and Glatts (2013), and local validity criteria (unpublished) and removed (83 cases). From that sample, 81 cases were identified with two baseline tests 13 to 40 months apart. Twenty-four pairs were outside the interval and were removed leaving a final sample of 67 baseline pairs that met inclusion criteria (47 male, 20 female).

Measures

All data was obtained from the ImPACT report. Age at time of each test were calculated based on date of birth and dates of testing; number of concussions and sex were reported on the ImPACT report at first test. Test intervals were calculated as the difference between dates on test 1 and test 2. ImPACT composite scores labeled Verbal Memory, Visual Memory, Visual Motor Speed, and Reaction Time were used in examining neurocognitive performance.

Analyses

Demographic variables of age and number of concussions were each entered into regression equations with the composite scores at each test-time as the dependent variables. To determine if sex of the participant would systematically affect the change scores obtained during retesting, we used the regression-based z-scores (RBz) of change in analysis of variance of sex by score (ANOVA sex by composite RBz). Calculating RBz in serial assessment is recommended by Heilbronner and colleagues (2010).

A Bonferonni corrected significance level of .008 was applied to the results. To analyze the significance of any change from test 1 to test 2 for the entire sample, we first computed paired-samples t-tests for each composite, and applied the Bonferonni correction for multiple comparisons. We then used the RBz change scores and determined the number of cases that exceeded a 1.96 z-score threshold in either direction. This score represents the 95th percentile confidence interval for a two-tailed test.

Results

Demographic Factors

The mean age at test 1 was 18.09 years for males and 18.05 for females; at test 2 the mean ages were 20.4 and 20.4, respectively. There were 58 cases with no concussions (87%), 7 cases with 1 previous concussion (10%), and 2 with 2 previous concussions (3%). Neither age nor concussions accounted for significant variance in any composite score based on linear equations, with age or concussion number as the independent variables and composite scores as the dependent variables.

Between sex differences in change scores were not significant, with no effect greater than .09 (partial eta squared).

Absolute Changes in Scores From Test 1 to Test 2

Given the lack of effect for sex in any composite at both time points, t-tests for each composite were calculated for the entire sample (test 1, test 2). Again, no pair survived the Bonferonni correction. T-test statistics by composite appear in Table 1.

Table 1. Paired Sample t-Tests Results by Composite (*df* = 66)

Pair	Mean Difference	Std. Deviation	*t*	Sig. (2-Tailed)
Verbal Memory	–2.9403	9.0518	–2.659	0.01
Visual Memory	–1.7164	10.9777	–1.28	0.205
Visual Motor Speed	–0.67433	4.76133	–1.159	0.251
Reaction Time	–0.00045	0.06964	–0.053	0.958

Number of Cases With Significant Change as Indexed by RBz Scores

The frequencies of scores that exceeded the expected amount of change were also calculated to see if a high number of individuals exceeded expected scores. ImPACT regression parameters for test retest reliabilities were used to calculate the z-scores that represent amount of change over and above expected practice effects (and regression to the mean). For Verbal Memory, 4 cases exceeded the ±1.96 score (6%); no Visual Memory composites exceeded this criterion (0%), three Visual Motor Speed scores (4%), and seven Reaction Time (10%) scores exceeded the criterion. Thus, the percentages exceeding the criterion were exceptionally small and within chance levels.

Discussion

In this carefully selected sample of athletes with two baseline tests between 1 and 3 years apart, there were no significant score differences from test 1 to test 2. Number of concussions, age, and sex of the athlete were not significant predictors in this sample. There were some improvements in test scores, and with only 10% or fewer of cases changed more than expected. On average, Reaction Time scores showed the most change. Of interest, females tended to be slower on the second test, while men were faster. A reverse pattern was noted on Visual Motor Speed, with males slowing down on the second test and females getting faster. Memory composites improved across the board.

These findings indicate that the amount of change in ImPACT composite scores when retested within 3 years was not significant and thus indicated good score stability. Furthermore, the number of athletes with change scores that exceeded expectations at retesting was not statistically meaningful. Thus, the practice of repeating baseline testing to gain additional useful information is not supported.

This study was conducted on one population of Division I college athletes. However, several limitations are noted. First, the majority of baseline testing was completed in groups larger than recently recommended (Lichtenstein et al., 2014; Moser et al., 2011, 2013; Vaughn et al, 2014). Yet even with less than optimal environments for baseline administrations, the results were informative and indicate unnecessary redundancy of repeat testing in this population. The consistent increase in scores from test 1 to test 2 was expected and provides a level of validity to the findings. Finally, this study has no bearing on the recommendation for retesting younger athletes every two years. Until proven otherwise, these recommendations appear to be appropriate from a developmental perspective.

References

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

Heilbronner, R. L., Sweet, J. J., Attix, D. K., Krull, K. R., Henry, G. K., & Hart, R. P. (2010). Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: The utility and challenges of repeat test administrations in clinical and forensic contexts. Clinical Neuropsychologist, 24, 1267–1278.

Lichtenstein, J., Moser, R., & Schatz, P. (2014). Age and test setting affect the prevalence of invalid baseline scores on neurocognitive tests. The American Journal of Sports Medicine, 42, 479–484.

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

Moser, R., Schatz, P., & Lichtenstein, J. (2013). The importance of proper administration and interpretation of neuropsychological baseline and postconcussion computerized testing. Applied Neuropsychology: Child. doi:10.1080/21622965.2013.791825

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

Vaughn, C., Gerst, E. H., Sady, M., Newman, J. B., & Gioia, G. A. (2014). The relation between testing environment and baseline performance in child and adolescent concussion management. The American Journal of Sports Medicine. doi:10.1177/0363546514531732