BRIEF COMMUNICATION

Reading Based IQ Estimates and Actual Premorbid Cognitive Performance: Discrepancies in a College Athlete Sample

Amanda R. Rabinowitz,[1] AND Peter A. Arnett[1]

[1]Psychology Department, The Pennsylvania State University, University Park, Pennsylvania

Abstract

The present study sought to evaluate the Wechsler Test of Adult Reading (WTAR) Full-Scale IQ (FSIQ) estimate as an index of premorbid ability in a sample 574 of healthy college athletes participating in a sports concussion management program. We compared baseline neuropsychological test performance with the WTAR FSIQ estimate obtained at baseline. Results revealed that the discrepancy between actual neuropsychological test scores and the WTAR FSIQ estimate was greatest for those with estimated FSIQs greater than 107. The clinical implication of this finding was evaluated in the 51 participants who went on to sustain a concussion. For individuals with higher IQ estimates, the WTAR estimate obtained post-concussion suggested greater post-concussion decline than that indicated by comparison with actual baseline neuropsychological performance. (*JINS*, 2012, *18*, 139–143)

Keywords: Intelligence, Cerebral concussion, Neuropsychological tests, Traumatic brain injury, Sports injuries, Intelligence tests

INTRODUCTION

Comparing neuropsychological test performance with a reading-based IQ-estimate is one method that neuropsychologists use to infer impairments in cognitive functioning. This approach has been supported by research demonstrating that estimates of intellectual functioning do indeed predict neuropsychological test performance across a range of ability levels (Diaz-Asper, Schretlen, & Pearlson, 2004). However, merely demonstrating a relationship between premorbid IQ estimates and cognitive abilities (e.g., attention, memory, etc.) does not directly speak to the *accuracy* of these instruments for predicting performance in these other neurocognitive domains. For example, Schretlen, Buffington, Meyer, and Pearlson (2005) demonstrated that, although correlations between National Adult Reading Test-Revised (NART-R) and concurrent Verbal and Full-Scale IQ (FSIQ) were robust, the NART-R's correlations with concurrent functioning in other cognitive domains were significantly weaker (Schretlen et al., 2005).

Discrepancies between predicted and actual performance could result in classifying individuals as impaired, when they are not in-fact experiencing cognitive decline, or vice versa. In fact, a study of base rate data for IQ-Memory Score discrepancies across IQ-strata demonstrated that memory performance was likely to exceed FSIQ in those with FSIQs below average, whereas the opposite was true in those with FSIQs in the High Average or Superior range. The authors suggest that clinicians who are unaware of this significant interaction between IQ and IQ-Memory Score discrepancy may be at risk of interpreting a normative performance pattern as indicative of memory decline (Hawkins & Tulsky, 2001).

The present study sought to replicate and extend these findings, by evaluating the Wechsler Test of Adult Reading (WTAR) as an index of premorbid ability in healthy college athletes participating in a sports-concussion management program. Participating athletes complete a comprehensive neuropsychological battery, including the WTAR, as a baseline measurement of their cognitive functioning. This provides the opportunity to compare the WTAR estimate with actual premorbid cognitive performance across a variety of domains—and specifically to quantify any possible discrepancy

Correspondence and reprint requests to: Amanda R. Rabinowitz, Department of Psychology, The Pennsylvania State University, 422 Moore Building, University Park, PA 16802-3106. E-mail: arr200@psu.edu

between actual and predicted performance. The present study sought to examine IQ-performance discrepancies across a range of cognitive domains, and to compare these discrepancies across ability levels. To evaluate the clinical implications of IQ-performance discrepancies, post-concussion test results were compared with baseline neuropsychological performance and a post-injury WTAR FSIQ estimate.

METHOD

Participants

The sample consisted of 574 college athletes (430 males and 144 females) participating in a concussion management program at a large state university. Athletes represented eight varsity athletic programs: football (31.4%), men's soccer (11.3%), women's soccer (10.6%), men's wrestling (2.1%), women's lacrosse (9.4%), men's lacrosse (14.3%), women's basketball (4.9%), and men's basketball (6.3%); and one non-varsity sports program, that is, men's ice hockey (9.4%). The average age of players at baseline was 18.5. The majority of the sample identified as Caucasian American (75%), 18% identified as African American, and 1% identified as Asian American. All participants were native English speakers. Thirty-eight percent of participants reported a history of concussion before study-inclusion. Fifty-one of these athletes (40 males and 11 females) went on to sustain a concussion during their study participation, and were tested then as well.

Procedure

Participants were administered a comprehensive battery of neuropsychological tests at baseline, before the start of their participation in team activities. Tests were administered by graduate students and undergraduate research assistants, under the supervision of a licensed psychologist and clinical neuropsychologist (P.A.). Athletes who went on to sustain a concussion were tested approximately 48 hr post-injury, using alternative forms of the test-battery. Concussions were identified at the time of injury by athletic trainers or coaching staff, and athletes were subsequently referred by one of the team physicians for testing. All injuries included in the present study met criteria for at least grade II concussions according to the American Academy of Neurology (AAN) Guidelines. Data were obtained in compliance with the standards of the university's Institutional Review Board.

Measures

The test battery consisted of several measures that assess cognitive functioning, including: the Hopkins Verbal Learning Test-Revised (HVLT-R; Benedict, Schretlen, Groninger, & Brandt, 1998), the Brief Visuospatial Memory Test-Revised (BVMT-R; Benedict, 1997), the Symbol-Digit Modalities Test (SDMT; Smith, 1982), the Digit Span Test (Wechsler, 1997), the PSU Cancellation Task, and the Stroop Color-Word Test (SCWT; Trenerry, Crosson, DeBoe, & Leber, 1989), and the

Wechsler Test of Adult Reading (WTAR; The Psychological Corporation, 2001). For athletes who underwent multiple evaluations, alternate forms of the HVLT-R, BVMT-R, SDMT, and PSU Cancellation Task were used (see Benedict et al., 1998; Benedict, 1997; Smith, 1982, for alternate form reliability information). With the exception of the WTAR, tests have demonstrated sensitivity to traumatic brain injury (Bailey, Echemendia, & Arnett, 2005; Bohnen, Twijnstra, & Jolles, 1992; Bruce & Echemendia, 2003; Ponsford & Kinsella, 1992). The Immediate Post-Concussion Assessment and Cognitive Testing computerized battery (ImPACT; Lovell, Collins, Podell, Powell, & Maroon, 2000) was administered along with paper-and-pencil tests at both time-points. Although complete validity data for the use of this test battery in sports-related concussion are yet unpublished, it assesses cognitive domains typically affected following such injury.

The ImPACT (Lovell et al., 2000) is a computerized test battery that was designed as a time-effective standardized method for collecting data to assist in concussion assessment and management. Six tests are included, designed to target attention, memory, processing speed, and reaction time. From these tests, five composite scores are derived: Verbal Memory, Visual Memory, Visuomotor Speed, Reaction Time, and Impulse Control. Studies in high school and college athletes have demonstrated that ImPACT performance is correlated with performance on similar paper-and-pencil neuropsychological tests (Iverson, Lovell, & Collins, 2005), and is sensitive to the acute effects of concussion (Schatz, Pardini, Lovell, Collins, & Podell, 2006).

The WTAR is a test of reading recognition that was designed for premorbid IQ estimation. The WTAR was developed and co-normed with the WAIS-III in both the US and the United Kingdom using the same large, representative sample of normally functioning adults. Using the normative data from the co-norming sample, WTAR scores can be converted to Full Scale IQ (FSIQ) estimates. The WTAR has strong correlations (.70–.80) with WAIS-III Full Scale IQ scores for a wide age range of WTAR scores, and WTAR performance is relatively resistant to the effects of traumatic brain injury (The Psychological Corporation, 2001). WTAR scores are highly correlated with other accepted premorbid measures—the American National Adult Reading Test (.90), National Adult Reading Test (.78), and the Wide Range Achievement Test – Revised Reading Test (.73). Test–retest reliability for the instrument is .92 for the 16–29 age group (The Psychological Corporation, 2001).

Approach to Data Analysis

To assess the discrepancy between WTAR FSIQ estimates and actual premorbid cognitive performance, all neuropsychological test indices were transformed into standard score (SS) units, using the athletes at baseline as a reference—hence, putting all test indices on the same metric as the FSIQ score. Each SS-transformed test index was subtracted from the WTAR FSIQ estimate to create an IQ-performance discrepancy score. Positive discrepancy scores indicate that the WTAR FSIQ score was higher than actual performance on the neuropsychological

indices (i.e., the WTAR FSIQ estimate *over-estimates* actual performance); negative scores indicate that the FSIQ score was *lower* than these indices (i.e., the WTAR FSIQ estimate *under-estimates* actual performance). These discrepancy scores were used in subsequent analyses. To examine the interaction between WTAR derived FSIQ and IQ-performance discrepancy, athletes were divided into high, mid-range, and low IQ groups, based on the first and third quartiles of the distribution of WTAR FSIQ estimates in the baseline sample.

For the purpose of the post-concussion analysis, an FSIQ estimate derived from the post-injury WTAR was used. As with baseline test-scores, all neuropsychological test indices were transformed into SS units using athletes at baseline as a reference. Post-concussion scores were each subtracted from the WTAR FSIQ estimate to create difference scores, representing post-concussion cognitive decline on a given task, as determined using the FSIQ estimate as the comparison standard. These scores will be referred to subsequently as WTAR-decline scores. The intra-individual mean of WTAR-decline scores across tasks was calculated to create a mean WTAR-decline score for each individual. To create baseline-decline scores, post-concussion SS-transformed scores were subtracted from the baseline SS-transformed score for each test index. The intra-individual mean of baseline-decline scores across tasks was calculated to create a mean baseline-decline score for each individual. The number of scores to decline 15 SS points or greater was calculated to create indices of the total number of scores for which the athlete exhibited clinically significant decline; separate indices were calculated using WTAR- and baseline-decline scores.

RESULTS

Baseline Analysis

The reliability of IQ-performance discrepancies scores for individual test-indices ranged from .45 to .88; reliability coefficients were higher than .60 for 11 of 14 test-indices. Athletes were divided into high, mid-range, and low groups, based on the distribution of WTAR FSIQ estimates in the baseline sample—162 athletes had FSIQ scores that fell in the low quartile group (range = 81–100; M = 96), 149 had estimates falling in the high quartile group (range = 107–117; M = 109.5), and 217 had estimates in the mid-range (range = 100–107; M = 104). Multivariate analysis of variance (MANOVA) revealed a significant multivariate effect of IQ group on IQ-performance discrepancies for individual neuropsychological test indices (Pillai's Trace $F = 5.12$; $p < .001$). Statistical significance of univariate tests was established using the Hochberg method correcting for multiple comparisons (Hochberg & Rom, 1995). Univariate results also supported a significant effect of IQ group on discrepancy scores; tests were significant for 11 of 14 neuropsychological test indices. These 11 significant effects all exhibited the same pattern: the WTAR FSIQ score was higher (over-estimation) than actual neuropsychological test performance for members of the high IQ group and, to a lesser extent, the mid-range IQ group (See Table 1).

Post-concussion Analysis

To evaluate the stability of the WTAR in response to injury, test–retest reliability and reliable change indices for the WTAR FSIQ estimate in the present sample were consulted (Jacobson & Truax, 1991). The correlation between baseline and post-concussion WTAR FSIQ estimate suggested good reliability ($R_{xx} = .87$), and only one of 51 participants exhibited any reliable change (a decline, using the 95% confidence interval). To evaluate the effect of comparison standard (post-injury WTAR *vs.* baseline neuropsychological test performance), a repeated measures general linear model (GLM) was constructed using method (WTAR-decline *vs.* baseline-decline) as a within subjects factor, and the post-injury

Table 1. Discrepancy between baseline test performance and WTAR FSIQ index across IQ groups

Test index	Low IQ	Mid-range IQ	High IQ	*F*	*p* value	Adj. α
ImPACT Verbal Memory	−1.15	4.10	5.82	8.9	.000	.006*
ImPACT Visual Memory	0.51	2.40	6.70	6.7	.001	.008*
ImPACT Visual Motor Speed	1.74	3.35	5.02	1.9	.155	.050
ImPACT Impulse Control	−1.49	2.88	8.42	15.7	.000	.006*
ImPACT Reaction Time	−0.61	3.90	6.63	8.7	.000	.005*
BVMT Total Recall	0.36	2.88	6.68	7.1	.001	.007*
BVMT Delayed Recall	−0.67	3.03	7.40	11.0	.000	.005*
HVLT Total Recall	0.58	3.29	6.10	5.4	.005	.013*
HVLT Delayed Recall	−0.73	3.80	6.39	8.6	.000	.004*
SDMT	0.53	2.82	4.85	3.2	.044	.025
Digit Span	2.26	1.42	5.24	3.3	.037	.017
Stroop-W	−1.33	4.01	6.10	11.9	.000	.004*
Stroop-CW	0.41	3.09	6.10	5.4	.005	.010*
PSU Cancellation Task	−3.84	3.66	8.83	25.2	.000	.004*

Note. *The *p* value is significant according to the Hochberg & Rom (1995) method correcting for multiple comparisons. All scores are in standard score units. The Low IQ group is comprised of participants with WTAR FSIQ estimates <100. The Mid-Range IQ group contains participants with WTAR FSIQ estimates between 100 and 107. The High IQ group contains individuals with estimates >107. Positive values indicate overestimation of actual neuropsychological test performance by WTAR FSIQ, whereas negative values indicate underestimation.
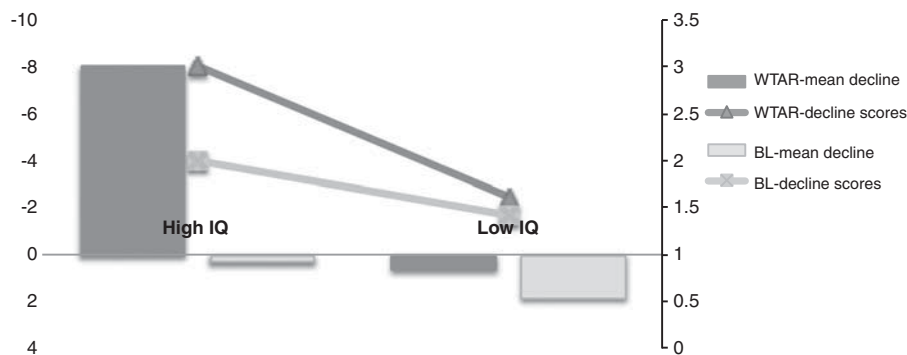
Fig. 1. Repeated measures GLM results for *method* by WTAR FSIQ interaction on suggested post-concussion decline, operationalized as both mean-decline, and the number of declined scores. *Method* is a two-level within subjects factor, WTAR-decline versus Baseline-decline. WTAR FSIQ score is a continuous random factor. To illustrate the nature of the interaction, participants were divided into "High IQ" (those participants with FSIQ estimates falling above the 3rd quartile) and "Low IQ" (those participants with FSIQ estimates falling below the 1st quartile) groups, based on the distribution of WTAR FSIQ scores in athletes assessed post-concussion. Bars represent mean post-concussion decline in SS points (axis on left). The dark gray bar represents mean decline using WTAR FSIQ score as a comparison standard; the light gray bar represents mean decline using baseline test-performance as a comparison standard. Lines represent the mean number of scores to decline by 15 SS points or greater (axis on right). The dark gray line represents the mean number of declined scores, using the WTAR FSIQ score as a comparison standard; the light gray line represents the mean number of declined scores, using baseline test-performance as a comparison standard.

WTAR FSIQ estimate as a covariate. GLM results revealed a significant main effect of method on the magnitude of suggested post-concussion decline ($F = 6.7$; $p < .05$), and a significant method by FSIQ estimate interaction ($F = 8.2$; $p < .01$). To test the possible clinical significance of this finding, a separate repeated measures GLM was constructed evaluating the total number of post-concussion scores to decline by 15 SS points or more. Results revealed a significant main effect of method ($F = 4.1$; $p < .05$), and a significant method by FSIQ estimate interaction ($F = 4.8$; $p < .05$). There was no main effect of FSIQ estimate on post-concussion decline ($F = 1.3$; $p = .26$), or the total number of declined scores post concussion ($F = .01$; $p = .92$).

The nature of these significant interactions is illustrated in Figure 1—for those with higher FSIQ estimates, the WTAR comparison standard suggested greater post-concussion decline, and a greater number of declined scores than when baseline test results were used as the comparison standard.

DISCUSSION

The primary purpose of the present study was to evaluate the validity of a reading based premorbid IQ estimator as an indicator of premorbid cognitive performance across ability levels. Athletes' performance on a neuropsychological test battery at baseline was compared with their predicted performance based on the WTAR FSIQ estimate. Results suggest that, although the WTAR FSIQ estimate may be a reasonable estimate of premorbid performance for college athletes with FSIQ estimates below 100, in individuals with FSIQ estimates above 107, WTAR estimates were significantly higher than athletes' performance on neuropsychological tests (i.e., the FSIQ estimate over-estimated actual performance)

by an average of 6.3 SS points. This result is consistent with the findings reported by Hawkins and Tulsky (2001), who reported that significant IQ-Memory Score discrepancies were normative in individuals with High Average and Superior FSIQ scores.

To evaluate the possible clinical implications of these discrepancies, post-concussion decline from post-injury WTAR FSIQ was compared with decline from actual baseline test-performance. Results suggest that the choice of comparison standard—baseline performance or WTAR FSIQ estimate—did not make a difference for individuals with lower IQ estimates. However, for individuals with higher IQ estimates, choice of comparison standard had a significant impact on the magnitude of suggested concussion-related cognitive decline, with the WTAR FSIQ estimate suggesting greater post-concussion decline than that indicated by comparison with actual baseline performance.

These findings suggest that choice of comparison standard—the WTAR FSIQ estimate, or actual premorbid performance—could influence clinical decision making for individuals with higher FSIQ estimates. It is possible that this effect is in part due to the influence of practice effects—all of the athletes that were assessed post-concussion had previous exposure to alternate forms of the same tests at baseline. However, the results of the baseline analyses suggest that the WTAR FSIQ estimate overestimated premorbid ability in these individuals, which likely contributed to the discrepancy between WTAR- and baseline-decline scores.

There are limitations of the present study that bear noting. The dependent variables used in the current study are difference-scores, and there are psychometric limitations of such scores that may threaten the validity of these variables. Specifically, the difference between two variables will be less

reliable than either of its component scores, particularly when component scores are correlated with one another (Williams & Zimmerman, 1977). As reported, reliability of IQ-performance discrepancy scores was reasonably high, and so this issue may not have been problematic in the present study. Additionally, the present sample is not representative of large normative samples, or many patients who seek neuropsychological assessment. Healthy college athletes represent a restricted range of intellectual functioning (WTAR FSIQ M = 103; *SD* = 6). All participants in the present study had IQ estimates broadly in the average range (between Low and High Average according to WAIS descriptors), as would mostly be expected in a group of healthy college students. Hence, the extent to which these findings may generalize to other populations remains an empirical question. Lastly, it is important to note that the present findings do not address the validity of using an actual measure of premorbid IQ (i.e., the WAIS-IV) as a comparison standard.

Despite these limitations, these findings are consistent with other work (Dodrill, 1997; Hawkins & Tulsky, 2001; Schretlen et al., 2005) suggesting that a premorbid IQ estimate may be a misleading benchmark for determining whether a given neuropsychological test score indicates impaired functioning. Neuropsychologists should estimate premorbid IQ in their patients with good reason—this information provides a context for test-interpretation. However, the present results suggest that premorbid estimates should be used as a comparison standard with caution, particularly for individuals with higher IQ estimates.

ACKNOWLEDGMENTS

REFERENCES

Bailey, C.M., Echemendia, R.J., & Arnett, P.A. (2005). *Visual memory predicts concussion status independent of verbal memory and post-mild traumatic brain injury symptoms in college athletes*. Paper presented at the 33rd annual meeting of the International Neuropsychological Society.

Benedict, R.H. (1997). *Brief Visuospatial Memory Test – Revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Benedict, R.H., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test-Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist*, *12*(1), 43–55.

Bohnen, N., Twijnstra, A., & Jolles, J. (1992). Performance in the Stroop color word test in relationship to the persistence of symptoms following mild head injury. *Acta Neurologica Scancdinavica*, *85*, 116–121.

Bruce, J.M., & Echemendia, R.J. (2003). Delayed-onset deficits in verbal encoding strategies among patients with mild traumatic brain injury. *Neuropsychology*, *17*(4), 622–629.

Diaz-Asper, C.M., Schretlen, D.J., & Pearlson, G.D. (2004). How well does IQ predict neuropsychological test performance in normal adults? *Journal of the International Neuropsychological Society*, *10*(1), 82–90.

Dodrill, C.B. (1997). Myths of neuropsychology. *The Clinical Neuropsychologist*, *11*(1), 1–17.

Hawkins, K.A., & Tulsky, D.S. (2001). The influence of IQ stratification on WAIS-III/WMS-III FSIQ-General Memory Index discrepancy base-rates in the standardization sample. *Journal of the International Neuropsychological Society*, *7*(7), 875–880.

Hochberg, R., & Rom, D. (1995). Extensions of multiple testing procedures based on Simes' test. *Journal of Statistical Planning and Inference*, *48*(2), 141–152.

Iverson, G.L., Lovell, M.R., & Collins, M.W. (2005). Validity of ImPACT for measuring processing speed following sports-related concussion. *Journal of Clinical and Experimental Neuropsychology*, *27*(6), 683–689.

Jacobson, N.S., & 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.

Lovell, M.R., Collins, M.W., Podell, K., Powell, J., & Maroon, J. (2000). *ImPACT: Immediate post-concussion assessment and cognitive testing. Computer software*. Pittsburgh, PA: Neuro-Health Systems, LLC.

Ponsford, J., & Kinsella, G. (1992). Attentional deficits following closed-head injury. *Journal of Clinical and Experimental Neuropsychology*, *14*(5), 822–838.

Schatz, P., Pardini, J.E., Lovell, M.R., Collins, M.W., & Podell, K. (2006). Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. *Archives of Clinical Neuropsychology*, *21*(1), 91–99.

Schretlen, D.J., Buffington, A.L., Meyer, S.M., & Pearlson, G.D. (2005). The use of word-reading to estimate premorbid ability in cognitive domains other than intelligence. *Journal of the International Neuropsychological Society*, *11*(6), 784–787.

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

The Psychological Corporation. (2001). *The Wechsler Test of Adult Reading (WTAR):Test manual*. San Antonio, TX: The Psychological Corporation.

Trenerry, M.R., Crosson, B., DeBoe, J., & Leber, W.R. (1989). *Stroop neuropsychological screening test*. Odessa, FL: Psychological Assessment Resources.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-III (WAIS-III)*. New York: Psychological Corporation.

Williams, R.H., & Zimmerman, D.W. (1977). The reliability of difference scores when errors are correlated. *Educational and Psychological Measurement*, *37*, 679–689.