The Pennsylvania State University

The Graduate School

College of the Liberal Arts

VALIDATION OF THE AFFECTIVE WORD LIST AS A MEASURE OF VERBAL LEARNING AND MEMORY

A Thesis in

Psychology

by

Jessica E. Meyer

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Master of Science

December 2013

The thesis of Jessica E. Meyer was reviewed and approved* by the following:

Peter A. Arnett
Professor of Clinical Psychology
Thesis Adviser

Brady Wyble
Assistant Professor of Cognitive Psychology

Brian Sharpless
Assistant Professor of Clinical Psychology

Melvin Mark
Professor of Psychology
Department Head

*Signatures are on file in the Graduate School.

Abstract

A major issue in concussion research is the lack of a consistent post-concussion cognitive profile. Great variability is seen in the mechanisms of injury, the presentation of symptoms, and the recovery from concussion. While no consistent post-concussive cognitive profile has emerged from the literature, the areas of verbal learning and memory are frequently found to be impacted by concussion. This study evaluated the reliability and validity of the Affective Word List (AWL), a new measure of verbal learning and memory. The AWL was developed to assess affective bias in order to overcome the tendency of some examinees to minimize self-report of depression symptoms. Because it is designed as a traditional list learning task, the AWL has the potential to additionally be used as a measure of verbal learning and memory. It was hypothesized that the AWL would be a valid and reliable measure of verbal learning and memory and would be more sensitive to the effects of concussion than currently used measures of verbal learning and memory (HVLT-R and ImPACT). Results of this study supported these hypotheses, showing that the AWL demonstrated moderate test-retest reliability, moderate convergent validity with other measures of verbal learning and memory and strong discriminant validity with measures of processing speed and reaction time. The AWL was found to be comparably sensitive to the cognitive effects of concussion as the ImPACT Verbal Memory Composite and the HVLT-R.

TABLE OF CONTENTS

List of Tables

List of Figures

Introduction

The management of sports-related concussions has become an increasingly popular area of research in recent years due to highly publicized cases of concussion in collegiate and professional athletes. Despite the increased focus of research on this issue, little consensus has been reached as how to best manage concussions and return to play decisions. A primary issue facing researchers is the lack of consistency in results of studies of the neurocognitive effects of concussion. Results vary not only between studies but also between participants, as response to concussion has been found to be very idiosyncratic (Iverson, Brooks, Collins, & Lovell, 2006; McCrea, Guskiewicz, Marshall, & Barr, 2003). While no consistent post-concussive cognitive profile has emerged from the literature, commonly observed deficits are in the areas of learning and memory (Guilmette & Rasile, 1995; McClincy, Lovell, Pardini, Collins, & Spore, 2006). The following study was designed to validate a new measure of verbal learning and memory and evaluate its sensitivity to post-concussive cognitive deficits.

Epidemiology of Concussion

The increased focus on sports-related concussions in the media and research in recent years has elucidated the high incidence of sports-related concussions and the potential for long-term consequences that make concussions a major public health issue. In the United states the annual incidence of concussion or mild traumatic brain injury (mTBI) is estimated to be 160 to 375 cases / 100,000 persons, while the number of sports concussions, sustained each year across all sports, has reached at least 300,000 (Hootman, Dick, & Agel, 2007; Pellman et al., 2004). Research shows that rates of concussion have been increasing by as much as 7% annually, though this may be due primarily to better identification of concussion rather than an actual increase in injury (Hootman et al., 2007).

Studies show that concussions account for 5.9 – 13.2 percent of injuries in high school and collegiate sports and it has been estimated that 3.9 – 7.7 % of high school and college athletes sustain mTBIs each year (Gessel, Fields, Collins, Dick, & Comstock, 2007; Marar, McIlvain, Fields, & Comstock, 2012; Powell & Barber-Foss, 1999). Concussion rates have been found to differ between high school and collegiate athletes (Gessel et al., 2007; Guskiewicz, Weaver, Padua, & Garrett, 2000). In a 2007 study, concussions accounted for 8.9% of all injuries in a sample of 100 high schools and 5.8% of all injuries in a sample of 180 colleges. Collegiate athletes, however, had a higher overall rate of concussions than did high school athletes. The increase in concussions amongst collegiate athletes may be due to the increased intensity of play resulting in a greater number of impacts, or the increased size and strength of the athletes resulting in more powerful impacts (Gessel et al., 2007).

Between puberty and middle age, sex differences in the incidence of traumatic brain injury are striking with men sustaining two to three times as many TBIs as women (Farace & Alves, 2000). No sex differences are seen in children or in adults over the age of 65 (Bruns & Hauser, 2003). Interestingly, a meta-analysis conducted by Farace and Alves found that, although women sustain fewer TBIs, they have worse outcome from TBI. This meta-analysis included studies evaluating all levels of TBI and a variety of outcome variables. Women fared worse than men on variables such as length of hospitalization, death, post-concussive symptoms (e.g. dizziness, impaired memory, headache), depression, and anxiety.

In sports concussions, sex differences have been observed in both incidence of concussion and recovery from concussion (Broshek et al., 2005; Covassin et al., 2006; Gessel et al., 2007). It has been found that females are less likely to sustain concussions in practice, but more likely to sustain concussions in games (Covassin, Swanik, & Sachs, 2003). Gessel et al. found that, in sports played by both sexes, females had a higher rate of concussion in both high school and college samples (2007). There is some evidence that the smaller head to ball ratio and weaker necks of females is driving this discrepancy in concussion rates (Barnes et al., 1998; Mansell, Tierney, Sitler, Swanik, & Stearne, 2005). Alternatively, females may simply be more honest when reporting injuries, causing concussions sustained by males to more often go undiagnosed (Marar et al., 2012).

Most sports concussions result from player to player contact (Dick, 2009). Sex differences are also seen, however, in mechanism of injury, with males showing an absolute higher percentage of concussions resulting from player to player contact and females showing absolute higher percentages of concussions resulting from contact with the ball or playing surface (Dick, 2009). The mechanism of injury also varies between sports, with football, soccer, basketball, baseball, hockey, and men's lacrosse showing higher rates of concussions from player to player contact; volleyball and wrestling showing higher rates of concussions from contact with the surface; and softball, field hockey, and women's lacrosse showing higher rates of concussions from contact with equipment (Marar et al., 2012). Research has shown that head contact with artificial turf more often results in concussion, and results in more severe concussions, than does head contact with natural grass (Guskiewicz et al., 2000).

Pathophysiology of Concussion

Concussion is often considered to be an "invisible injury," leaving neither a visible wound nor neuronal lesions. Concussion has been defined as a "transient neurologic dysfunction resulting from a biomechanical force (Giza & Hovda, 2001). The signs and symptoms of concussion are believed to result from "sequential neuronal dysfunction" rather than lesions of the brain, reinforcing the status of concussions as an "invisible injury." While concussions often do not result in "destruction" of neurons, neuronal dysfunction occurs as a result of a variety of pathophysiological changes in the brain, termed the "neurometabolic cascade."

The initial step in the "neurometabolic cascade" is a widespread nonspecific depolarization and initiation of action potentials, resulting in a release of excitatory neurotransmitters such as glutamate, and a large efflux of potassium and influx of calcium. This depolarization results in widespread suppression of neuronal activity. This ionic shift also leads the sodium potassium pump to require more energy, in the form of adenosine triphosphate (ATP), in order to restore homeostasis. This increased need for ATP leads to a need for a dramatic increase of glucose metabolism, which is achieved through hyperglycolysis. The "energy crisis" resulting from the lack of ATP necessary to regain homeostasis is thought to make the brain less able to recover from a second injury (Giza & Hovda, 2001). During this period of hyperglycolysis, lactate production increases in conjunction with a decrease in lactate metabolism leading to an accumulation of lactate that can lead to neuronal dysfunction.

This initial phase of hyperglycolysis is followed by a decrease in cerebral glucose. In severe TBI, cerebral glucose metabolism has been found to be diminished for 2-4 weeks (Bergsneider et al., 2000).

Neuroimaging of Concussion

Neuropathophysiology of concussion, as described above, has been investigated primarily in animal models and verified by studies of neuropathological findings in humans who sustained concussions but died of other causes (McCrea & Powell, 2012). Neuroimaging has served the primary method of attempting to understand the neuropathophysiology of concussion in humans. Initially, as concussions rarely lead to discrete lesions in the brain, it was thought that structural neuroimaging techniques would be unable to capture the effects of concussion on the brain.

Although the nature of the microstructural damage caused by concussion may cause brain abnormalities to go undetected with traditional structural imaging techniques, the use of diffusion tensor imaging (DTI) has shed light on some structural damage resulting from concussion. More specifically, DTI allows for detection of diffuse axonal injury by measuring the integrity of the white matter in the brain through fractional anisotropy (FA). DTI studies have provided promising results in studies of individuals in both the acute and chronic phases of concussion. Niogi et al. (2008) found reduced FA in the anterior corona radiata and the uncinate fasciculus in individuals with a history of concussion, and interestingly also found that the FA values correlated with cognitive performance in both the healthy and concussed individuals (Niogi et al., 2008).

Studies that have measured FA in acutely concussed individuals have found DTI abnormalities in a variety of regions but most often reduced FA has been observed in the corpus callosum and internal capsules (Arfanakis et al., 2002; Inglese et al., 2005), suggesting that these areas may be most susceptible to axonal injury. DTI results confirm that concussions may lead to compromised white matter integrity in the brain, which likely decreases the processing efficiency of the neurons. The functional consequences of this decreased neuronal efficiency are observed in many functional magnetic resonance imaging (fMRI) studies.

Functional imaging provides insight into the pathophysiological and functional abnormalities that occur post-concussion. The use of fMRI has resulted in varied findings across studies, perhaps due to the variability in inclusion criteria including severity of mTBI and time since injury. One consistent finding has been increased bifrontal and biparietal activation in response to increasing working memory loads (McAllister et al., 1999; McAllister et al., 2001; Smits et al., 2008). Some studies have additional reported activation outside of the typical working memory circuitry during these tasks, suggesting possible compensatory recruitment of other networks (Smits et al., 2008). Cognitive performance of mTBI individuals in these imaging studies has most often been reported as being equal to that of healthy individuals, though individuals with mTBI tend to report more cognitive symptoms (Smits et al., 2008). It is hypothesized that the perception of increased cognitive difficulties is a result of the increased cognitive energy and neuronal activation necessary in individuals with mTBI to achieve the same level of cognitive performance.

Biomechanics of Concussion

The variability in concussion assessment is further complicated by the inconsistent nature of injury that results in concussion. Researchers have attempted to establish a minimal biomechanical threshold for concussion using animal models (Ommaya & Gennarelli, 1974), video analyses and dummy reenactments (Newman et al., 2000; Newman et al., 1999; Pellman, Viano, Tucker, Casson, & Waeckerle, 2003), as well as in-helmet systems measuring linear head acceleration (Brolinson et al., 2006; Guskiewicz et al., 2005; McCaffrey, Mihalik, Crowell, Shields, & Guskiewicz, 2007). Ommaya and Gennarelli's early work on the biomechanics of concussion in animal models set the foundation for further use of animal models in studying mathematical and physical models of concussion. Their focus on linear versus rotational forces has contributed to the current attempts researchers are making to elucidate the injury threshold. More recent animal model studies have used information from the in-helmet system studies to create animal models of concussion more closely resembling the impact seen in sports concussion. These models have been used to evaluate the immediate effects of concussion in an animal model (Viano, Hamberger, Bolouri, & Säljö, 2009).

While animal models have provided a foundation for studying the biomechanics of concussion, more recent studies have utilized video analysis and dummy reenactment of collisions in sporting events. Pellman et al. (2003) found that in football players, impacts to the facemasks resulted in concussion at lower accelerations than impacts on other quadrants of the head. Zhang et al. also used video analysis and dummy reenactments to study 24 head-to-head collisions in professional football games. Interestingly, they found that translational (linear) acceleration had a greater impact on concussion development than did rotational accelerations (Zhang, Yang, & King, 2004), a finding inconsistent with previous research (McCrea & Powell, 2012). Zhang et al. developed a system of deducing probability of concussive injury based on the combination of translational and rotational acceleration (Zhang et al., 2004).

Current research on the biomechanics of concussion primarily utilizes in-helmet accelerometers, such as the Head Impact Telemetry (HIT) system to measure the frequency and intensity of concussive and subconcussive hits. In a study analyzing the hits that resulted in concussion in collegiate football players, Guskiewicz et al., found that the average linear acceleration recorded in concussion was between 60 and 120g (Guskiewicz et al., 2005). They also reported, however, that the majority of hits over 80g did not result in concussion, suggesting that this amount of linear acceleration is necessary, but not sufficient, to cause concussion

(Guskiewicz et al., 2005). The rotational acceleration, unrecorded by the HIT system, likely lowers the threshold of linear acceleration necessary for concussion and heavily influences which hits result in concussion.

Neuropsychology of Concussion

The variability in injuries that complicates the assessment of the biomechanics of concussion also complicates the clinical outcomes. Because concussions occur with impact to different areas of the head, at a variety of linear and rotational accelerations, and individuals have different injury thresholds, influenced by factors such as neck strength (Mansell et al., 2005), the clinical outcome of concussive injuries can be very idiosyncratic.

Neuropsychological testing has played an important role in concussion assessment since Barth et al. (1989) began using baseline testing as part of their Sports as a Laboratory Assessment Model (SLAM). Barth et al. employed baseline testing in a 4-year prospective study of concussions in 10 university football programs (Barth et al., 1989). Baseline testing was used to establish pre-season cognitive functioning, allowing athletes to serve as their own controls and allowing for detection of idiosyncratic and subtle changes in individual athletes' performance post-concussion (Barth et al., 1989). Since this seminal study, in which most athletes were found to recover from the cognitive consequences of concussion within 10 days, baseline testing has become the standard for concussion assessment.

While pre-post-injury test comparisons have resulted in significant research findings, described below, as well as great clinical utility, these comparisons may be complicated by factors such as depression and intraindividual variability. Intraindividual variability seems to be indicative of more than just random error and has been linked with clinical phenomena such as

TBI, depression, and anxiety (MacDonald, Li, & Bäckman, 2009), and has been found to be sensitive to cognitive decline in normal aging (Lövdén, Li, Shing, & Lindenberger, 2007). A recent study found that high variability in cognitive performance at baseline predicted an increase in cognitive performance variability and a decrease in overall cognitive performance post-concussion in a sample of collegiate athletes (Rabinowitz & Arnett, in press). These findings suggest that intraindividual variability at baseline may have clinical utility as a predictor of who will exhibit neurocognitive decline post-concussion.

Without factoring in intraindividual variability, Pre-post-injury test comparisons have led to findings of cognitive impairment in a variety of domains post-concussion. A 2005 meta-analysis explored the effects of concussion across nine cognitive domains: orientation, global cognitive ability, attention, executive functioning, memory acquisition, delayed memory, language, visuospatial ability, and motor abilities. The largest effect sizes were seen in the domains of memory acquisition and global cognitive ability and the smallest effect sizes were seen in attention (Belanger & Vanderploeg, 2005).

Consistent with the findings of Belanger and Vanderploeg (2005), verbal learning and memory have often been found to be impacted by concussion (Collins et al., 1999; Field, Collins, Lovell, & Maroon, 2003; McCrea et al., 2003). Verbal learning and memory are often evaluated in concussion work through the administration of the Hopkins Verbal Learning Test – Revised (HVLT-R). The HVLT-R is a well-validated and reliable measure that is used in concussion management and research due to the availability of six alternate forms, helping to minimize practice effects in repeat assessments (Benedict, Schretlen, Groninger, & Brandt, 1998). Although the HVLT-R has been found to be sensitive to concussion in many studies (Collins et al., 1999; Field et al., 2003), the findings are not entirely consistent (Guskiewicz, Ross, &

Marshall, 2001). A potential cause of the variability in HVLT-R findings is that a ceiling effect on the test reduces the variability in scores necessary to detect changes due to concussion. In a 1998 study by Lacritz and Cullum, the majority of older adults (mean age= 70.7) tested approached ceiling level on the HVLT-R by trial 3 (Lacritz & Cullum, 1998). Lacritz and Cullum concluded that the HVLT-R may not be long enough or complex enough to elicit sufficient variability in scores and error types.

It is hypothesized that this ceiling effect on the HVLT-R detected in older adults would be present to an even greater extent in young athletes, and this could have considerable impact on the results of concussion studies. Thus, it is likely that a longer and more complex measure of verbal learning and memory will be more sensitive to cognitive changes following concussion.

The Affective Word List (AWL) was designed as a performance based test of affect to circumvent the tendency of some examinees to minimize self-report of depression symptoms (Ramanathan, Rabinowitz, Barwick, & Arnett, 2012). This test has significant clinical utility in a sample of athletes due to several factors affecting athletes' willingness to report affective symptoms post-concussion, including stigma associated with depression and motivation to return to play. The AWL has been found to be a valid measure of affective bias in a sample of collegiate athletes at a baseline assessment (Ramanathan et al., 2012). In addition to measuring affective bias, the AWL has the potential to be explored as a measure of verbal learning and memory, as it is designed in the same manner as a traditional list-learning task.

The current study was designed to test the validity and reliability of the AWL as a measure of verbal learning and memory in a sample of collegiate athletes. Due to the increased length of the word list and the more difficult nature of the task without the assistance of semantic clustering, it was hypothesized that the participants would display a wider range of scores, something that is likely to increase the sensitivity of the task to cognitive changes due to concussion. In 2005, Randolph, McCrea, and Barr, outlined the necessary steps to validate a sport concussion test by establishing its reliability, sensitivity, validity, change scores/classification rates, and clinical utility. (Randolph, McCrea, & Barr, 2005). Randolph et al. (2005) limited clinical utility to the test's ability to detect neurocognitive impairments after self-reported post-concussive symptoms had resolved, though others have suggested different approaches to clinical utility. Because recovery patterns from concussion are variable (Iverson et al., 2006) and athletes' symptom reporting is often unreliable (McCrea, Hammeke, Olsen, Leo, & Guskiewicz, 2004), the proposed study did not evaluate clinical utility in this way. Because clinical utility is heavily influenced by the time requirement and difficulty of administration of a test (Randolph et al., 2005), the AWL's clinical utility was evaluated by exploring its validity as a verbal learning and memory measure relative to the HVLT-R. Characteristics of athletes who decline on the AWL were explored to further evaluate the AWL's sensitivity to concussion. Specifically, variability in baseline performance was evaluated as a predictor of decline on the AWL post-concussion.

Given that the AWL has already been shown to be sensitive to affective bias in detecting depression, it would be useful to know whether it was also sensitive to verbal learning and memory problems post-concussion. If this is the case, then the instrument could serve a dual purpose: Detection of mood problems and cognitive difficulties. Using these guidelines, the current study evaluated the following hypotheses:

Hypothesis 1)

The AWL will have a normal distribution of scores in the sample and the four forms will be equivalent. The order of test administration (HVLT-R – AWL vs. AWL – HVLT-R) will have no influence on test performance.

Hypothesis 2)

The AWL will demonstrate moderate stability over time as evaluated through test-retest reliability. This stability will be influenced by changes in state-dependent traits at each session, specifically, depression.

Hypothesis 3)

The AWL will demonstrate convergent validity with other measures of verbal learning and memory (HVLT-R, ImPACT Verbal Learning Composite, Rivermead Behavioral Memory Test – Story Memory subtest) and discriminant validity with measures of processing speed and attention (PSU Cancellation Task, VIGIL, and CARB).

Hypothesis 4)

a) The AWL will be able to detect decline relative to baseline following concussion. Mean change from baseline to post-concussion will be examined, in addition to reliable change scores.

b) The AWL will demonstrate greater sensitivity to concussion than the HVLT-R.

c) Athletes with more variable performance at baseline will be more likely to decline on the AWL post-concussion than athletes with stable baseline performance.

Method

Participants

Three samples of participants were used for different aspects of this study. Most analyses were conducted using data from the Penn State Concussion program, in which college athletes at risk for concussion are assessed prior to and following concussion. Athletes were assessed before starting their sports participation at Penn State to determine their baseline level of functioning. At these assessment sessions the athletes completed a thorough neuropsychological battery in addition to questionnaires to obtain data on previous head injuries and background demographic information. Only those athletes who have undergone a baseline assessment in which both the Affective Word List and the Hopkins Verbal Learning Test – Revised were administered will be included in this study. Additionally, those athletes from the baseline sample who sustained a concussion during their time at Penn State were included in the analyses related to sensitivity to concussion. 379 athletes were included as participants in the baseline sample and 45 athletes were included in the post-concussion sample. Participant characteristics are outlined in Tables 1 and 2.

Table 1. Sample characteristics: participants tested at baseline

	N	Mean	SD
Age	379	18.3	0.79

Previous Head Injuries	379	0.55	0.82
BDI-FS	379	0.96	1.66
	N		**N**
Gender		**Sport**	
Male	284	Football	146
Female	95	Men's Soccer	43
Ethnicity		Women's Soccer	40
Caucasian American	276	Wrestling	10
African American	72	Men's Lacrosse	39
Asian American	3	Women's Lacrosse	54
Hispanic American	7	Men's Ice Hockey	6
Biracial or Multiracial	12	Men's Basketball	13
Latin American	1	Women's Basketball	26
Other	6	Softball	2
Unknown	2		

Note: BDI-FS = Beck's Depression Inventory- Fast Screen

Table 2. Sample characteristics: participants tested post-concussion

	N	**Mean**	**SD**
Age Post-Concussion	45	19.78	1.48
Previous Head Injuries	45	1.04	2.28
BDI-FS (BL)	45	1.22	2.20
BDI-FS (Post-Concussion)	45	2.16	2.36
	N		**N**
Gender		**Sport**	

Male	36	Football	21
Female	9	Men's Soccer	1
Ethnicity		Women's Soccer	2
Caucasian American	26	Wrestling	3
African American	15	Men's Lacrosse	6
Biracial or Multiracial	3	Women's Lacrosse	3
Unknown	1	Men's Ice Hockey	1
		Men's Basketball	5
		Women's Basketball	3

Note: BDI-FS = Beck's Depression Inventory- Fast Screen;

In order to study the effects of test order (AWL-HVLT-R vs. HVLT-R-AWL) on performance on the AWL, a sample of 31 undergraduate students, active in intramural or club athletics, were administered the standard baseline battery of tests, with half of them receiving the AWL first and the other half receiving the HVLT-R first.

An additional sample of 42 undergraduate students, active in intramural or club athletics, served as participants for a test-retest reliability study. These participants were recruited from the psychology department subject pool and were given credit for their psychology courses for participating in this study.

Measurements

Affective Word List

The Affective Word List (AWL) was developed to measure affective bias (Ramanathan et al., 2012), specifically to assess the proportion of positive to negative information remembered by the examinee. A negative affective bias, in which one remembers more negatively valenced stimuli than positively valenced stimuli, has been found to be associated with depression in many studies (Maalouf et al., 2012; Scher, Ingram, & Segal, 2005). The AWL is designed in the same format as a traditional list-learning task. The examinee is read a list of 16 words, 8 positively valenced and 8 negatively valenced, and then asked to recall as many of the words as possible. This is repeated two additional times for a total of three learning trials. After a 20 to 25 minute delay, examinees are again asked to recall as many of the words as they can remember. There are four forms of the AWL that have been matched on the frequency with which the words are used in the English language and number of verbs, nouns, and adjectives. To minimize any possible differential influence of the word length effect on recall (Baddeley, Thomson, & Buchanan, 1975), all words on all forms of the AWL are one syllable.

Hopkins Verbal Learning Test – Revised (HVLT-R)

The HVLT-R was developed in 2001 by Brandt and Benedict to measure verbal learning and memory in moderately demented patients (Strauss, Sherman, & Spreen, 2006). In this task examinees are read a list of 12 words, with four words from each of three semantic categories, and asked to recall as many words as they can remember. This is repeated over three learning trials and then a delayed recall trial is administered 20-25 minutes later, followed immediately by a recognition trial. A 1998 study by Benedict at al. reported adequate test-retest reliability for total recall (r=0.74), slightly lower reliability for delayed recall (r=0.66), and low reliability for percent retention (r=0.39) and recognition discrimination index (r=0.40) (Benedict et al., 1998).

This study was limited, however, to a small sample (n=40) of elderly participants. Additionally, there was a wide range of test-retest intervals ranging from 14 to 134 days (mean =46.6, SD = 30.1). In a 2003 study using a sample of high school athletes, Barr reported generally lower (<.60) test-retest reliability for all indices of the HVLT-R (Barr, 2003). It has been suggested that the low test-retest reliability may be due to the restricted and non-normal distribution of the test variables (Benedict et al., 1998). When convergent validity of the HVLT-R with the CVLT was explored in a sample of patients with Alzheimer's Dementia, Lacritz et al. (2001) found that the two measures correlated moderately on total learning (r=0.36), delayed recall (r=0.62), intrusion errors (0.34), and recognition hits (0.48) (Lacritz, Cullum, Weiner, & Rosenberg, 2001). Additionally, in a 1999 study, the HVLT-R was found to correlate more strongly with verbal memory (WMS-R logical memory; r=0.75) than with visual memory (WMS-R visual reproduction; r=0.43) (Shapiro, Benedict, Schretlen, & Brandt, 1999).

Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) Verbal Memory Composite Score

The ImPACT Verbal memory composite score summarizes the average scores on a word recognition task, a letter memory task with an accompanying interference task, and a symbol number match task. When a sample of young adults was tested on two occasions, an average of 5.8 days apart (range = 1-13 days), the test-retest correlation coefficient was found to be 0.70 (Iverson, Lovell, & Collins, 2003). In a 2007 study, the test–retest reliability of several computerized concussion assessment programs was examined using test-retest intervals more closely resembling those more typically seen in concussion evaluations. After an interval of 45 days, the Verbal Memory Composite of the ImPACT was found to have an intraclass correlation coefficient (ICC) of 0.23, suggesting that the test-retest reliability of this measure may decrease significantly over longer time intervals (Broglio, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007).

Rivermead Behavioral Memory Test (RBMT) – Story Memory

In this RBMT Story Memory task, the participant listens to a short story and then is asked to recall it immediately after the reading and again 20-25 minutes later. The story memory task is part of a larger battery of tasks comprising the RBMT. Very high test-retest reliability (r=0.96) has been reported at short retest intervals (1 week) with the RBMT English version composite score (Makatura, Lam, Leahy, Castillo, & Kalpakjian, 1999). A 2001 study of the RBMT-Chinese Version also found high test-retest reliability of the composite score (r=0.89) after a two-week interval with alternative forms (Man & Li, 2002). The above test-retest reliabilities may be inflated by the stability of a large composite score in comparison to an individual test score. The RBMT has been found to have moderate correlations with the Wechsler Memory Scale and the Rey Auditory Verbal Learning Test (Strauss et al., 2006).

PSU Cancellation Task

In the PSU Cancellation task, a measure of visual attention, participants are asked to cross out every symbol on a page that exactly matches a target symbol. They are given 90 seconds to cross out as many matching symbols as possible, working left to right across the rows. The PSU Cancellation Task has four equivalent forms and has been found to demonstrate adequate test-retest reliability (Echemendia, Putukian, Mackin, Julian, & Shoss, 2001).

Vigil/W Continuous Performance Task

The Vigil/W Continuous Performance Task serves as a measure of reaction time and sustained attention (Cegalis & Cegalis, 1994). In this computerized task a series of letters are presented in the middle of the screen and the participant is required to hit the space bar as quickly as possible each time he or she sees the letter “K.” The test used in the concussion program lasts for about five minutes and results provide the average response speed to the targets, the number of omissions (missed target letters), and number of commissions (number of non-target letters responded to).

Computerized Assessment of Response Bias (CARB)

The CARB is a computer-based task designed to assess inadequate effort during neuropsychological testing. On this task participants view a five digit number for several seconds. After a short delay they must choose the most recently displayed number from two options that appear on the screen. Participants with mild to severe traumatic brain injuries have been found to perform near perfect levels on this task, suggesting that it is not sensitive to effects of actual TBI (Green & Iverson, 2001). Deficits (measured by errors) on the CARB in this sample of athletes at baseline can be viewed as a sign of insufficient effort rather than true cognitive deficits. However, we also used the CARB mean reaction time as a measure of sustained attention and processing speed in the current study.

WAIS-III Digit Span

The Digit Span test is a subtest of the WAIS-III designed to measure span of immediate verbal recall (Wechsler, 1997). The Digit Span test consists of two tasks - Digit Span Forward

(DSF) and Backward (DSB). These tasks consist of the test administrator reading aloud strings of digits (ranging from 2 to 9 digits) at a rate of one word per second. In the DSF task, the participant is asked to repeat the numbers as they were presented, while in the DSB task, the participant is asked to repeat the digits in the reverse order. Participants are asked to successfully repeat digits of increasing span length, with two strings of digits at each span length. In the PSU Concussion program, and in this study, a slightly modified version of the Digit Span test is used, in which the participant must only accurately repeat one string of digits at each span length before moving to the next span length. The Digit Span test is a frequently used and well-validated test. It has been found to be sensitive to a number of cognitive and neuropsychological conditions (Lezak, Howieson, & Loring, 2004), including concussion (Vanderploeg, Curtiss, & Belanger, 2005). This task has been found to have excellent test-retest reliability in healthy adult males (Levine, Miller, Becker, Selnes, & Cohen, 2004) and in high school athletes (Barr, 2003).

Symbol Modalities Test (SDMT)

The SDMT is a test designed to assess speeded information processing and visual tracking (Lezak, Howieson, & Loring, 2005). The SDMT, similar to the Digit Symbol subtest of the WAIS-III, requires the participant to pair a number to a symbol based on a key at the top of the page consisting of 9 unique symbols paired with each of the 9 numerals. There are a total of 110 symbols on the page, and the participant has 90 seconds to write the correct numbers under their corresponding symbols. The SDMT has been found to be sensitive to severe TBI and other neuropsychological conditions (Pfeffer et al., 1981; Ponsford & Kinsella, 1992). Research has demonstrated that the written SDMT has excellent test-retest reliability over a one-month interval (Levine et al., 2004).

Comprehensive Trail-Making Test (CTMT).

The CTMT is designed to measure attention, cognitive flexibility, and visual-motor speed (Reynolds, 2002). It consists of five unique trails, all of which require the participant to connect numbers and/or letters in a particular order as quickly as possible. Trail-making tasks have been found to be sensitive to a variety of neuropsychological disorders (Lezak et al., 2004), including concussion (Belanger & Vanderploeg, 2005; Leininger, Gramling, Farrell, Kreutzer, & Peck, 1990). These tasks have demonstrated excellent test-retest reliability in healthy adult males (Reynolds, 2002) and in high school athletes (Barr, 2003).

Controlled Oral Word Association Test (COWAT)

The COWAT is a test designed to measure phonemic fluency. The test administrator states a specific letter and asks the participant to produce as many words as they can in 60 seconds. This is repeated with two additional letters, for a total of 3 trials. The COWAT has been found to be sensitive to a variety of neuropsychological conditions including dementia, mood disorders, and TBI (Strauss et al., 2006). The COWAT has been found to have high alternate form reliability for different sets of letters (Cohen & Stanczak, 2000; Lacy et al., 1996) and high test retest reliability in a sample of high school athletes (Barr, 2003)

Beck Depression Inventory- Fast Screen (BDI-FS)

The BDI-FS serves as a brief self-report measure of depression. It contains 7 statements corresponding to symptoms of depression. Each statement is rated on a 4-point scale (0-3) with higher numbers corresponding to more depressive symptoms. The overall score is used as a measure of depression, with higher scores indicating greater depression. The BDI-FS is modeled after the BDI, and serves as an abbreviated measure of depression. The BDI-FS has been found to be a valid and reliable measure of depression in both healthy and clinical populations (Beck, Steer, & Brown, 2000; Benedict, Fishman, McClellan, Bakshi, & Weinstock-Guttman, 2003).

Post-Concussive Symptoms Scale (PCSS)

The PCSS is a 22-item self-report measure of post-concussive symptoms. The participant rates the extent to which they are experiencing each symptom on a 7-point Likert scale (0-none to 6-severe). The PCSS is modeled after the Post-Concussion Scale (PCS), which was designed to measure post-concussive symptom severity in athletes (Lovell et al., 2006). This scale has been found to be a reliable measure of post-concussive symptoms and variations of this scale have been used throughout concussion research (Chen, Johnston, Collie, McCrory, & Ptito, 2007; Kontos et al., 2012; Lovell et al., 2006).

Procedure

The above measurements of interest were included in a larger neuropsychological battery that all participants completed at baseline. The battery was administered by an undergraduate research assistant or graduate student under the supervision of a clinical neuropsychologist and licensed psychologist.

An additional sample of 31 athletic undergraduates with no history of concussion completed the same battery. 17 completed the standard battery and 14 completed the battery with the HVLT-R and AWL in reverse order.

To explore test- retest reliability, a sample of 42 athletic undergraduate students was recruited. Test-retest reliability studies vary in test battery content, participant selection, and test-retest time intervals (Barr, 2003; Benedict et al., 1998; Woods, Delis, Scott, Kramer, & Holdnack, 2006). This study used procedures consistent with the procedures used by Barr et al. (2003) who established the test-retest reliability of a short battery of tests in a sample of high school athletes. Participants were tested during two sessions approximately 7 weeks apart, consistent with studies evaluating test-retest reliabilities at time intervals more closely resembling those seen in concussion management (Barr, 2003; Broglio et al., 2007). These participants were given an abbreviated and modified version of the battery of tests that was administered to the athletes at baseline and post-concussion, and received alternate forms of the AWL at the second testing session. The test battery consisted of the AWL, Digit Span, SDMT, CARB, PSU Cancellation Task, and the Comprehensive Trail-Making Test.

Low test-retest reliabilities have often been observed in verbal learning measures (Barr, 2003; Broglio et al., 2007). Because of this finding, and the potential additional influence of affect on AWL test performance, additional questionnaires, described in the measures section of this proposal, were administered at the end of each testing session to assess factors potentially affecting test-retest reliability. Depression was evaluated as a potential moderator of test-retest reliability.

Data Analysis

Hypothesis 1:

The AWL will have a normal distribution of scores in the sample and the four forms will be found to be equivalent. The order of test administration (HVLT-R – AWL vs. AWL – HVLT-R) will have no influence on test performance.

Descriptive statistics were run on all indices of the AWL to evaluate its basic properties including its distribution and normality. A one-way ANOVA, accompanied by appropriate post-hoc comparisons, was used to compare test forms on performance on the AWL. A MANOVA was used to evaluate any effects of the order of test administration on AWL and HVLT-R performance in a sample of subject pool participants.

Hypothesis 2:

The AWL will demonstrate moderate stability over time as evaluated through test-retest reliability. This stability will be influenced by changes in state-dependent traits at each session, specifically, depression.

Test-retest reliability was assessed in the subject pool sample by correlating performance on all indices of the AWL at the time of the first assessment with performance on all indices of the AWL at the time of the second assessment. This provided Pearson's r values representing the stability of each of the indices over time, which were used in RCI analyses. Intra-class correlations (ICCs) were also calculated to account for individual variation in scores from Time 1 to Time 2. While Pearson's r values establish whether paired values are ranked in a similar order from test to retest, ICCs additionally establish the absolute level of agreement between individual values at test and retest (P. Schatz, Kontos, & Elbin, 2012).

To assess the influence of depression on test-retest reliability, Reliable Change Indices were used to identify participants with a reliable change in depression between testing sessions. These participants were then removed from the sample and the intraclass correlation coefficient was recalculated to determine whether changes in mood had a negative impact in test-retest reliability.

Hypothesis 3:

The AWL will demonstrate convergent validity with other measures of verbal learning and memory (HVLT-R, ImPACT Verbal Learning Composite, Rivermead Behavioral Memory Test) and discriminant validity with measures of processing speed and attention (PSU Cancellation Task, VIGIL, and CARB).

To establish convergent validity, the total recall and delayed recall scores from the AWL and HVLT-R were correlated. Additionally, the immediate and delayed recall of the AWL were correlated with ImPACT Verbal Memory Composite and the immediate and delayed recall of the RBMT Story Memory Test. To establish discriminant validity, correlations were run between the total and delayed recall scores from the AWL and the number of correct target shapes on the PSU Cancellation Task, the average response time on the VIGIL, and the mean response time on the CARB. The convergent and discriminant validity correlation values were compared (via Fisher's r to z) to assess whether they were significantly different.

Convergent and discriminant validity were further evaluated using an exploratory factor analysis. All of the variables described above were included in the analysis. Factors were extracted using Principal Components Analysis with Varimax rotation and factor values greater than 0.4 were retained.

Hypothesis 4:

a) The AWL will be able to detect decline relative to baseline following concussion. Mean change from baseline to post-concussion will be examined, in addition to reliable change scores.

b) The AWL will demonstrate greater sensitivity to concussion than the HVLT-R.

c) Athletes with more variable performance at baseline will be more likely to decline on the AWL post-concussion than athletes with stable baseline performance.

Hypothesis 4 was evaluated in three ways. For 4a, paired sample t-tests comparing baseline and post-concussion performance on the total recall and delayed recall indices of the AWL were calculated. For 4b, a Chi-Square analysis was conducted examining the proportion of athletes who show a significant decline on the AWL Total Recall versus those who show a significant decline on the HVLT-R Total Recall. Evidence for significant decline was determined by using reliable change scores. Reliable change scores were calculated according to the method outlined by Jacobson and Truax (Jacobson & Truax, 1991), using the test-retest data established in hypothesis 2 to account for practice effects and the reliability of each test-index. Reliable change scores of +/- 1.64 (90% confidence interval) were considered to be significant.

To evaluate 4c, cluster analysis was first used to divide the sample into high variability and low variability groups. First, all test indices from the primary tests given in the cognitive battery for concussion assessment were converted into standard scores (SSs) using the baseline sample means and standard deviations. Performance variability was calculated by taking the standard deviation across all SS-converted test-indices for each individual at baseline. Baseline performance variability indices were entered into a k-means cluster analysis where k was set to equal 2. Chi-Square analysis was conducted examining the proportion of athletes with variable baseline performance who decline on the AWL versus athletes with less variable baseline performance.

Results

Hypothesis 1:

The AWL will have a normal distribution of scores in the sample and the four forms will be found to be equivalent. The order of test administration (HVLT-R – AWL vs. AWL – HVLT-R) will have no influence on test performance.

Descriptive and normality statistics for the AWL, as well as the HVLT-R and ImPACT Word Memory and Verbal Memory Composite for comparison are documented in Table 3. Notably, the HVLT-R total recall and delayed recall and all ImPACT indices show negatively skewed non-normal distributions. The AWL Delayed recall index showed a slight positive skew (but still within normal limits) and the AWL total recall index was normally distributed.

Table 3. Descriptive and normality statistics for the AWL, HVLT-R, and ImPACT

	Mean	SD	Skewness	Kurtosis	W-stat	p-value
AWL TR	26.38	5.01	-.065	-.220	.993	.09
AWL DR	7.26	2.94	.176	-.256	.984	<.05

HVLT-R TR	26.06	3.82	-.360	-.244	.982	<.05
HVLT-R DR	9.30	1.94	-.498	-.320	.941	<.05
ImPACT-WM-TR	11.36	0.91	-1.55	2.42	.718	<.05
ImPACT-WM-DR	10.78	1.30	-1.30	2.51	.828	<.05
ImPACT – VM Composite	85.08	11.43	-1.09	2.14	.925	<.05

Note: AWL-TR = Affective Word List – Total Recall; AWL-DR = Affective Word List- Delayed Recall; HVLT-R TR = Hopkins Verbal Learning Test – Revised Total Recall; HVLT-R DR = Hopkins Verbal Learning Test- Revised Delayed Recall; ImPACT WM-TR = Immediate Post-Concussion Assessment and Cognitive Testing Word Memory Total Recall; ImPACT WM-DR = Immediate Post-Concussion Assessment and Cognitive Testing Word Memory delay Recall; ImPACT VM Composite = Immediate Post-Concussion Assessment and Cognitive Testing Verbal Memory Composite

A statistically significant difference in performance on the total recall index of the AWL was found between test forms of the AWL (F(3)=3.20, p=.023). Post-hoc comparisons, however, revealed no statistically significant differences between test forms. A marginally statistically significant difference between test forms was also found in performance on the delayed recall index of the AWL (F(3)=2.53, p=.057).

There was no statistically significant difference in performance on the AWL or HVLT-R based on the order in which the tests were administered, F (4,30) = .713, p =0.589; Wilk's Λ = 0.913.

Hypothesis 2:

The AWL will demonstrate moderate stability over time as evaluated through test-retest reliability. This stability will be influenced by changes in state-dependent traits at each session, specifically, depression.

Participants were tested twice, using alternate forms of the AWL, an average of 49 days apart. AWL test scores at time 1 and time 2, as well as relevant test-retest indices are reported in table 4. The difference scores suggest a mild practice effect on the delayed recall index and a paired-sample t-test confirmed that performance on this index differs significantly between time 1 and time 2 (t(41) = -2.28, p = .028).

Table 4. AWL performance at Time 1 and Time 2, and test-retest reliability indices

	Time 1	Time 2	Time 2 – Time 1	Pearson's r	ICC
AWL-TR	26.26 (4.77)	26.83 (5.18)	0.57 (4.5)	0.590	0.588
AWL-DR	7.90 (2.86)	8.69 (2.87)	0.79 (2.24)	0.695	0.695

Note: AWL-TR = Affective Word List – Total Recall; AWL-DR = Affective Word List- Delayed Recall; ICC = Intraclass Correlation Coefficient

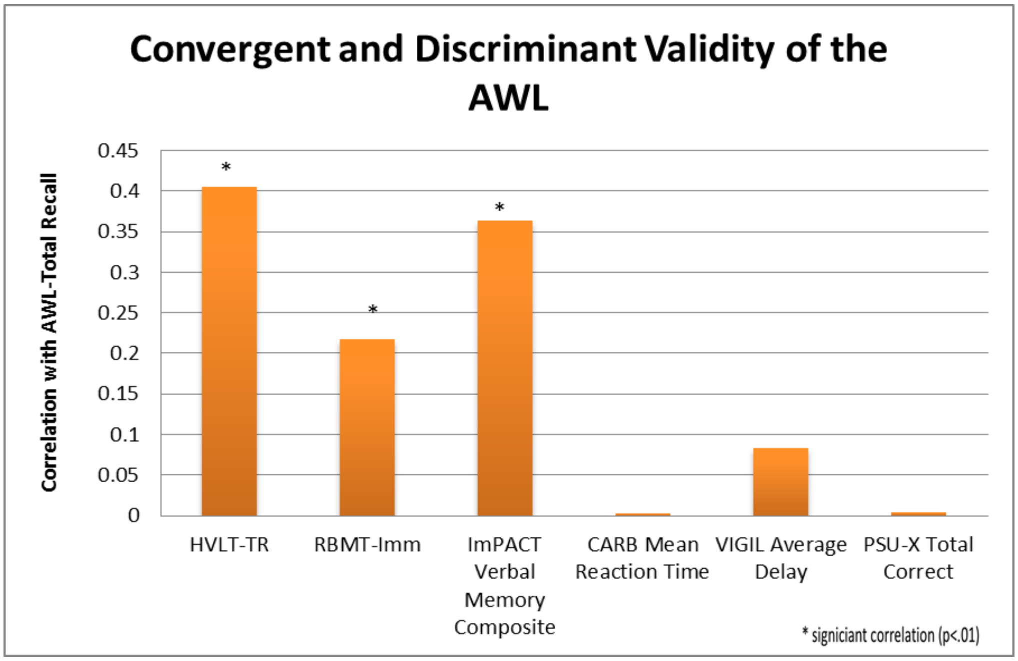
To assess the influence of depression on test-retest reliability, Reliable Change Indices were used to identify participants with a reliable change in depression between testing sessions. These RCI calculations resulted in 5 participants showing a reliable increase or decrease in BDI-FS scores between sessions. When these 5 participants were removed from the sample, there was a slight increase in the intraclass correlation coefficients representing test-retest reliability (full sample ICC = 0.596, stable BDI-FS sample ICC = 0.695). Although the AWL is designed to assess mood through affective bias, changes in mood only mildly influence the total number of words recalled by participants.

Hypothesis 3)

The AWL will demonstrate convergent validity with other measures of verbal learning and memory (HVLT-R, ImPACT Verbal Learning Composite, Rivermead Behavioral Memory Test – Story Memory subtest) and discriminant validity with measures of processing speed and attention (PSU Cancellation Task, VIGIL, and CARB).

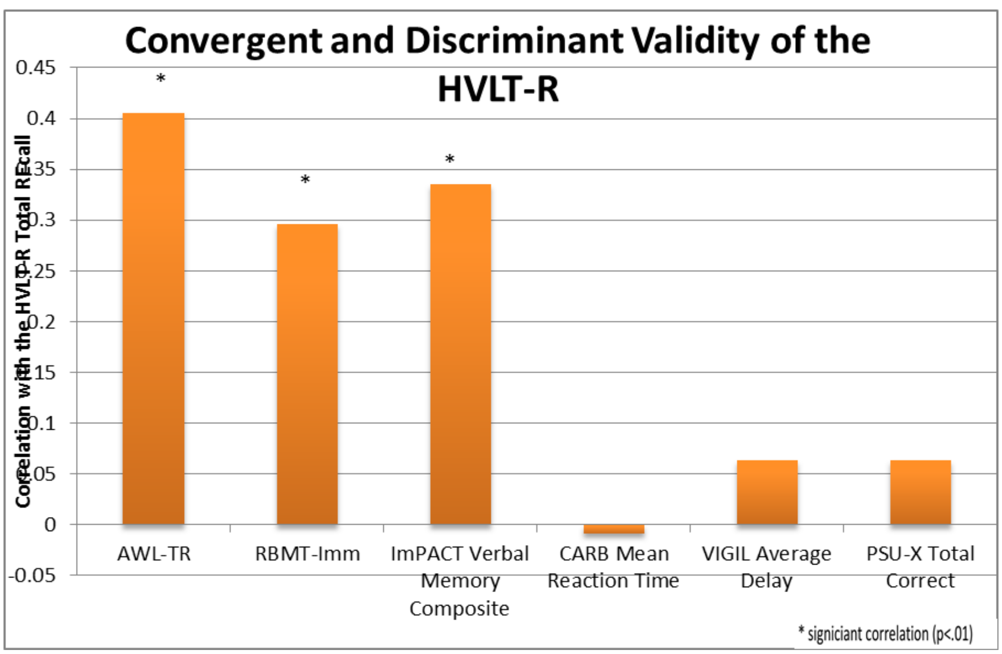
The AWL showed moderate convergent validity with other measures of verbal memory and learning and strong discriminant validity with measures of processing speed and reaction time (table 5). The convergent validity of the AWL with the RBMT story memory and ImPACT Verbal Memory Composite Score was similar to that of the HVLT-R with these measures (Figures 1& 2). The AVLT demonstrated strong discriminant validity with measures of processing speed and reaction time (figure 1), and again these correlations were similar to those seen in the HVLT-R (Figure 2). Fisher's r-to-z analyses comparing convergent and discriminant validity correlation values showed that each convergent validity correlation was significantly (p at least $< .05$) different from each discriminant validity correlation.

Figure 1: Convergent and discriminant validity of the AWL



Note: AWL=Affective Word List; HVLT-TR = Hopkins Verbal Learning Test – Revised Total Recall; RBMT-Imm = Rivermead Behavioral Memory Test, Story Memory, Immediate Recall; PSU-X = PSU Cancellation Test

Figure 2: Convergent and discriminant validity of the HVLT-R



Note: AWL-TR = Affective Word List Total Recall; RBMT-Imm = Rivermead Behavioral Memory Test, Story Memory, Immediate Recall; HVLT-TR = Hopkins Verbal Learning Test – Revised Total Recall; PSU-X = PSU Cancellation Test

Table 5. Correlation matrix for convergent and discriminant validity of the AWL

	AWL-TR	HVLT-TR	RBMT-IR	ImPACT VM	CB MRT	Vigil AD	PSU CR
AWL-TR	1.0						
HVLT-TR	.405*	1.0					
RBMT-IR	.218*	.296*	1.0				
ImPACT VM	.363*	.335*	.202*	1.0			
CB MRT	.003	-.009	-.048	-.032	1.0		
VIGIL AD	.084	.064	.056	.207*	.068	1.0	
PSU CR	.005	.064	-.199*	.079	.218*	.103*	1.0

* significant at p<.05

Note: AWL-TR = Affective Word List – Total Recall; HVLT-TR = Hopkins Verbal Learning Test – Revised Total Recall; RBMT-IR = Rivermead Behavioral Memory Test, Story Memory, Immediate Recall; ImPACT VM- Immediate Post-Concussion Assessment and Cognitive Testing Verbal Memory Composite; CB MR T = CARB Mean Reaction Time; VIGIL-AD = VIGIL Average Delay; PSU CR = PSU Cancellation Test Total Correct

The above tests were entered into Principal Components Factor Analysis with Varimax rotation, revealing a two-factor solution interpreted as verbal memory and processing speed/reaction time (table 6). Correlation matrix eigenvalues were 1.95 and 1.35 before rotations (47.2% of total variance explained). The first factor, interpreted as "Verbal Memory" accounted for 27.9% of the variance and the second factor, interpreted as "processing speed/ reaction time," accounted for 19.3% of the variance. A varimax rotation with Kaiser normalization was applied to the components. The varimax factor loadings are shown in table 6. One variable, Vigil Average

Delay, did not load onto either factor. The RBMT Immediate Recall loaded positively on verbal memory; it also loaded negatively, but to a lesser degree, onto the Processing Speed / Reaction Time factor.

Table 6. Convergent and discriminant validity factor loadings

	Verbal Memory	Processing Speed/Reaction Time
ImPACT VM Composite	**0.713**	.109
HVLT-R TR	**0.735**	-.001
AWL-TR	**0.723**	.011
RBMT-IR	**0.533**	**-.420**
VIGIL-Average Delay	.307	0.385
CARB- Mean Response Time	-.031	**0.626**
PSU- Cancellation Correct	.050	**0.788**

Note: ImPACT VM Composite = Immediate Post-Concussion Assessment and Cognitive Testing Verbal Memory Composite; HVLT-R TR = Hopkins Verbal Learning Test – Revised Total Recall; AWL-TR = Affective Word List – Total Recall; *RBMT-IR = Rivermead Behavioral Memory Test, Story Memory, Immediate Recall;*

Hypothesis 4)

a) The AWL will be able to detect decline relative to baseline following concussion. Mean change from baseline to post-concussion will be examined, in addition to reliable change scores.

b) The AWL will demonstrate greater sensitivity to concussion than the HVLT-R.

c) Athletes with more variable performance at baseline will be more likely to decline on the AWL post-concussion than athletes with stable baseline performance.

Paired sample t-tests showed no statistically significant mean change from baseline to post-concussion for the AWL Total Recall, AWL Delay Recall, HVLT-R Total Recall, HVLT-Delay Recall, or the ImPACT Verbal Memory Composite Score (Table 7)

Table 7. Mean change of verbal memory measures from baseline to post concussion

	t(44)	p
AWL TR	.774	0.443
AWL DR	1.345	0.185
HVLT-R TR	0.623	0.536
HVLT-R DR	1.415	0.164
ImPACT VM Composite	1.143	0.259

Note: AWL TR = Affective Word List – Total Recall; AWL DR = Affective Word List- Delayed Recall; HVLT-R TR = Hopkins Verbal Learning Test – Revised Total Recall; HVLT-R DR = Hopkins Verbal Learning Test- Revised Delayed Recall; ImPACT VM Composite = Immediate Post-Concussion Assessment and Cognitive Testing Verbal Memory Composite

Reliable change indices showed that the AWL's sensitivity to the effects of concussion was comparable to that of the HVLT-R and the ImPACT Verbal Memory Composite Score (Table 8). Reliable change on delay recall measures was comparable for the AWL and the HVLT-R. The ImPACT Verbal Memory Composite score does not have a delay index. Evidence for regression to the mean was seen in all of the above measures and baseline scores were therefore adjusted to account for this using the Edwards-Nunnally Correction (Speer, 1992) Additionally, the AWL Total Recall, AWL Delay Recall, HVLT-R Total Recall, and HVLT-R Delay Recall showed evidence of practice effects; therefore, the RCI scores were adjusted to account for this as well. AWL scores were standardized to account for differences between forms. A Chi-square analysis showed no effect of test on reliable decline (χ^2 (4) = 3.302, p =0.509).

Table 8. Percentage of participants showing reliable decline on verbal memory measures post-concussion

	AWL TR	HVLT-R TR	ImPACT VM Composite	AWL-DR	HVLT-R DR
% Decline	35.6%	31.1%	20%	26.7%	33.3%
% no Decline	64.4 %	68.9%	80%	73.3%	66.7%

Note: AWL-TR = Affective Word List – Total Recall; AWL-DR = Affective Word List- Delayed Recall; HVLT-R TR = Hopkins Verbal Learning Test – Revised Total Recall; HVLT-R DR = Hopkins Verbal Learning Test- Revised Delayed Recall; ImPACT VM Composite = Immediate Post-Concussion Assessment and Cognitive Testing Verbal Memory Composite

To examine the effects of baseline performance variability on post-concussion performance on the AWL, cluster analysis was first used to divide the sample into high variability and low variability groups. First, test indices from the primary tests given in the cognitive battery for concussion assessment were converted into standard scores (SSs) using the baseline sample means and standard deviations. The test indices used in these calculations were determined through previous work in which Principal Components Factor Analysis was used to determine which indices load together (Rabinowitz & Arnett, 2013). Performance variability was calculated by taking the standard deviation across all SS-converted test-indices for each individual at baseline. Baseline performance variability indices were entered into a k-means cluster analysis where k was set to equal 2. The cluster center for the low variability cluster (N=28) was 9.63 SS points and the cluster center for the high variability cluster (N=17) was 18.62. Chi-square analyses for each test showed no effect of baseline variability on reliable decline in performance Tables 9 and 10.

Table 9. Percentage of participants with low and high baseline variability showing reliable decline on the AWL Total Recall

	% Decline	**% No decline**
Low Variability	39.3%	60.7%
High Variability	29.4%	70.6% %

Table 10. Percentage of participants with low and high baseline variability showing reliable decline on the AWL Delay Recall

	% Decline	**% No decline**
Low Variability	28.6%	71.4%
High Variability	23.5%	76.5%

Discussion

Summary of Findings

The results of this study support the hypothesis that the AWL is a valid and reliable measure of verbal learning and memory in collegiate athletes and its sensitivity to concussion is comparable to other established measures of verbal learning and memory.

The first hypothesis tested in this study was designed to establish the normative properties of this measure. The total recall index of the AWL was found to have a normal distribution at baseline and showed no indication of a ceiling effect. The delayed index of the AWL, however, had a non-normal distribution with a slightly positive skew (though still within normal limits). All other measures of verbal learning and memory evaluated showed a negative skew, suggesting that many participants were performing very well on these tasks, a pattern indicative of a ceiling effect.

The AWL may be capable of standing alone as both a measure of verbal learning and memory and affective bias, but prior to the reliability and validity analyses for this study it had only been used in the context of a battery with other measures of verbal learning and memory. Because the order in which the two list learning tasks (AWL and HVLT-R) were administered had no effect on performance on either measure, it can be concluded that normative performance on the AWL, as evaluated in this current study, would not differ if the test was administered independent of other list learning tasks.

The second hypothesis of this study was evaluated to establish the stability of performance on the AWL over time. Test-retest reliability is particularly important in concussion assessment as the baseline testing paradigm necessitates multiple testing sessions over varying time frames. Measures of verbal learning and memory have been found to have low to moderate test-retest reliability over different time frames (Barr, 2003; Broglio et al., 2007), and it was hypothesized that the AWL would show moderate stability. Results of the test-retest reliability study showed that the AWL demonstrated moderate stability over time. Due to the affective nature of the task it was hypothesized that changes in mood may affect performance on the AWL. Because only a small number of participants displayed a reliable change in mood over time ($n = 5$), it was not possible to test differences in reliability between change and no-change groups. However, when the participants with a reliable change in mood were removed from the sample, there was a slight increase in the test-retest reliability of the measure of about .10. This suggests that changes in mood may affect the test-retest reliability of the AWL, but further study of this issue is necessary given the power limitations in the current study.

The AWL was found to have moderate convergent validity with other measures of verbal learning and memory (HVLT-R, ImPACT Verbal Learning Composite Rivermead Behavioral Memory Test- Story Memory subtest) and strong discriminant validity with measures of processing speed and attention (PSU Cancellation Task, VIGIL, and CARB). The convergent and discriminant validity demonstrated by the AWL was comparable to that demonstrated by the HVLT-R, suggesting that the AWL is targeting similar constructs as currently used measures of verbal learning and memory.

Mean change analyses showed no significant mean change on the AWL or the HVLT-R from baseline to post-concussion. When reliable change indices were used, however, it was found that the sensitivity of the AWL to concussion was comparable to that of the HVLT-R. While the AWL did not show greater sensitivity to concussion, it has the added benefit of being sensitive to affective changes post-concussion. The dual use of this measure has many potential benefits to be discussed below.

In order to begin to determine what factors influence whether or not an athlete declines on the AWL post-concussion, the effects of baseline variability were evaluated. Despite an emerging literature suggesting that baseline variability may be indicative of sensitivity of cognitive functioning to injury (Rabinowitz & Arnett, 2013), baseline variability had no significant effect on performance on the AWL post-concussion.

In summary the AWL was found to be a reliable and valid measure of verbal learning and memory. Its test-retest reliability, convergent and discriminant validity, and normative properties were comparable to those of the HVLT-R. Similar numbers of athletes showed a post-concussion decline on the verbal learning and memory indices of the AWL as on the HVLT-R and the

ImPACT Verbal Learning Composite Score. These results suggest that the AWL is able to stand alone as a valid and reliable measure of verbal learning and memory.

Interpretation of Results

Although the AWL was not found to be more sensitive to concussion than current validated measures of verbal learning and memory, its sensitivity was comparable and it provides the additional strength of assessing mood and cognitive symptoms simultaneously. The AWL was designed as a performance based test of affect that is resistant to the possibility of some athletes' tendency to underreport symptoms such as depression post-concussion. Athletes may be unwilling to report affective symptoms post-concussion due to several factors including stigma associated with depression, as well as motivation to return to play. In a 2012 study using the AWL, athletes were divided into positive and negative bias groups using a cutoff of 1.5 SD from the mean affective bias for the group. It was found that at baseline, a significantly greater proportion of athletes with a BDI-FS score of 2 (mild) or above was in the negative bias group than the positive bias group (80% vs. 20%), suggesting that the AWL may be a useful indicator of self-reported depression scores in collegiate athletes (Ramanathan et al., 2012). Athletes may be more likely to endorse symptoms of depression at baseline, rather than post-concussion, because there is no influence of motivation to return to play. It is therefore hypothesized that the AWL and BDI-FS scores would be consistent with each other at baseline but may differ post-concussion. Future studies will need to focus on the effects of concussion on affective bias.

Because the AWL can provide an objective measure of mood and verbal learning and memory simultaneously, its implementation has the possibility to shorten the test battery. A shorter test battery could lead to reduced fatigue and improved motivation in athletes, both of which would improve the validity of all assessments in the battery.

Motivation for testing has long been a concern for neuropsychological assessments. Traditionally the concern has been for head injury patients underperforming for secondary gain, that is, exaggerating their symptoms for financial compensation or other benefits (Binder, 1993; Binder & Rohling, 1996). In sports concussion clinical work and research, however, the problem is often athletes exhibiting very high motivation post-concussion with a desire to minimize symptoms and return to play. High motivation post-concussion is only a problem due to frequent sub-optimal motivation at baseline due to factors such as apathy, boredom, and at times even intentionally underperforming to prevent future detection of post-concussive symptoms (Bailey, Echemendia, & Arnett, 2006). Because of the comparative nature of concussion assessments, differential motivation at baseline and post-concussion can lead to undetected post-concussive cognitive symptoms.

In a 2006 study evaluating the impact of motivation on neuropsychological concussion assessments, it was found that when athletes were divided into “suspect motivation” and “high motivation” groups based on their baseline performance relative to mean performance, the “suspect motivation” group demonstrated greater improvement on several tests post-concussion than the high motivation group (Bailey et al., 2006). A higher percentage of suspect motivation participants additionally improved reliably (as established by reliable change indices) on several tests compared with the high motivation participants. These data suggest that low motivation at baseline can lead to an inaccurate interpretation of post-concussion assessment data. An additional finding of this study was that individuals’ motivation level categorization changed depending on the measure being evaluated and no individual showed suspect motivation for every measure in the battery. The authors suggested that this may be indicative of some measures being more resistant to motivation changes, but also may indicate fluctuations in motivation across the test battery. Shortening the test battery would provide fewer opportunities for motivation fluctuations and allow for maximum motivation across tasks, and thereby improving the validity of the concussion assessment.

The effects of cognitive exertion on neuropsychological test performance have not been well studied to date. It has been suggested, however, that like physical exertion cognitive exertion may worsen concussion symptoms (Majerske et al., 2008). A study in which cognitive and physical exertion were studied as one entity, found that athletes with the highest levels of exertion post-concussion showed the largest decrease in neurocognitive scores and the highest post-concussion symptom scores. The authors of this study noted that anecdotally cognitive activity related to school was a frequently reported problem for many athletes (Majerske et al., 2008). Previous research has suggested that like physical exertion, cognitive exertion can change metabolic activity in the brain, meaning that cognitive exertion could exacerbate the “neurometabolic cascade” that occurs after concussion. More research is needed to establish the effects of cognitive exertion on post-concussive symptoms and cognitive performance. It is possible, however, that engaging in a long neuropsychological testing session may have negative effects on concussion symptoms and recovery. Reducing the length of the testing battery and removing redundant tests would likely help to minimize these potential negative consequences of cognitive exertion.

While shortening the test battery has clear benefits, a balance must be struck between maximizing athlete’s motivation and the accuracy of test performance and retaining enough indices to detect the wide variety of post-concussive changes that can be observed. As discussed above, there is no consistent post-concussive cognitive profile that has been observed. It is necessary, therefore, to retain enough testing indices to pick up on the wide variety of possible post-concussive cognitive difficulties. Current assessment batteries may be improved, however, by efforts to minimize the redundancy seen in many testing batteries. Evaluating multiple tests given within the same neurocognitive domain provides the opportunity to assess for redundancy in specific domains. In the above study, the AWL was found to have moderate convergent validity with the HVLT-R and ImPACT Verbal Learning Memory Composite, and showed similar rates of detecting post-concussion cognitive decline. This suggests that there may be no added benefit to administering all of these measures and this is an area in which the battery could be shortened. Because the AWL has the added property of providing a performance-based measure of mood, and additionally shows no ceiling effect as the other measures do, it is likely a stronger measure of post-concussion changes than the currently widely used HVLT-R.

Limitations

The conducted study has several limitations. For all analyses other than the test-retest reliability analyses, the AWL was evaluated in the context of a battery that contained another list learning task (the HVLT-R). This potential confound was addressed through a normal control study to evaluate the effect of test order on test performance and it was found that there was no statistically significant difference in performance on either test based on the order in which they were administered.

The test-retest reliability study conducted as part of this project had a mean inter-test time of 7 weeks. Concussion assessments may occur at any time after baseline, meaning that this test-retest interval may in some cases be significantly longer than that seen in concussion assessment, or in many cases much shorter. Many test-retest reliability studies use even shorter time frames, often testing participants between one week and one month apart (Broglio et al., 2007; Philip Schatz & Ferris, 2013; Woods et al., 2006). The current study is thought to more closely resemble a typical time frame seen in concussion assessment than many existing studies, however, it still is likely an underrepresentation of the time between most baseline and concussion assessments.

Summary and Recommendations

Future research should focus on the ability of the AWL to detect cognitive and affective changes post-concussion simultaneously. Previous work has found that the affective bias score on the AWL at baseline is consistent with the athlete's self-reported level of depression on the BDI-FS (Ramanathan et al., 2012). Change in affective bias post-concussion has not, however, been evaluated. In order to determine the sensitivity of the AWL to affective changes post-concussion, it will be necessary to evaluate the proportion of athletes who show a reliable decrease in affective bias (representing an increase in depression) in comparison with the proportion of athletes who decline on other measures of mood.

Additional research should also be performed on the influence of mood on performance on the cognitive indices of the AWL. While a preliminary analysis of the effect of mood on stability of test performance over time indicated that changes in mood mildly affected performance on the cognitive indices of the AWL, the study had a small sample size and statistical analysis of this question was not possible. It is possible that because mood affects which words are remembered (seen in changes in affective bias), it may also influence the number of words recalled. Future studies could compare performance on the AWL in low and high depression groups, based both on affective bias and athlete self-report.

Lastly, future research should focus on the development of a recognition task to accompany this measure. While verbal memory recognition is not often affected by concussion (Collins et al., 1999; Field et al., 2003), if the AWL were to be used in other neuropsychological populations this additional index may be sensitive to different cognitive profiles. A forced-choice index would additionally provide an embedded measure of effort to further help to detect poor motivation for testing.

References

Arfanakis, K., Haughton, V.M., Carew, J.D., Rogers, B.P., Dempsey, R.J., & Meyerand, M.E. (2002). Diffusion tensor MR imaging in diffuse axonal injury. *American Journal of Neuroradiology, 23*(5), 794-802.

Baddeley, A.D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior, 14*(6), 575-589.

Bailey, C.M., Echemendia, R.J., & Arnett, P.A. . (2006). The Impact of Motivation on Neuropsychological Performance in Sports-Related Mild Traumatic Brain Injury. *Journal of the International Neuropsychological Society, 12*, 475-484. .

Barnes, B.C., Cooper, L., Kirkendall, D.T., McDermott, T.P., Jordan, B.D., & Garrett, W.E. (1998). Concussion history in elite male and female soccer players. *The American Journal of Sports Medicine, 26*(3), 433-438.

Barr, W.B. (2003). Neuropsychological testing of high school athletes: Preliminary norms and test–retest indices. *Archives of Clinical Neuropsychology, 18*(1), 91-101.

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

Beck, A.T., Steer, R.A., & Brown, G.K. (2000). *BDI-FastScreen for Medical Patients Manual*. San Antonio:: The Psychological Corporation.

Belanger, H.G., & Vanderploeg, R.D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*, 345-357.

Benedict, R.H.B., Fishman, I., McClellan, M. M., Bakshi, R., & Weinstock-Guttman, B. . (2003). Validity of the Beck Depression Inventory - Fast Screen in multiple sclerosis. *Multiple Sclerosis, 9*, 393–396.

Benedict, R.H.B., 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*, 43-55.

Bergsneider, M. , Hovda, D.A. , Lee, S.M. , Kelly, D.F., McArthur, D.L., Vespa, P.M., . . . Phelps, M.E. (2000). Dissociation of cerebral glucose metabolism and level of consciousness during the period of metabolic depression following human traumatic brain injury. *Journal of neurotrauma, 17*(5), 389-401.

Binder, Laurence M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 15*(2), 170-182.

Binder, Laurence M, & Rohling, Martin L. (1996). Money matters: a meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *American Journal of Psychiatry, 153*(1), 7-10.

Broglio, S.P., Ferrara, M.S., Macciocchi, S.N., Baumgartner, T.A., & Elliott, R. (2007). Test–retest reliability of computerized concussion assessment programs. *Journal of athletic training, 42*, 509–514.

Brolinson, P.G., Manoogian, S., McNeely, D., Goforth, M., Greenwald, R., & Duma, S. (2006). Analysis of linear head accelerations from collegiate football impacts. *Current sports medicine reports, 5*(1), 23-28.

Broshek, D.K., Kaushik, T., Freeman, J.R., Erlanger, D., Webbe, F., & Barth, J.T. (2005). Sex differences in outcome following sports-related concussion. *Journal of neurosurgery, 102*(5), 856-863.

Bruns, J., & Hauser, W.A. (2003). The epidemiology of traumatic brain injury: a review. *Epilepsia, 44*(s10), 2-10.

Cegalis, J.A., & Cegalis, S. . (1994). *The Vigil/W Continuous Performance Test (manual)*. New York: ForThought.

Chen, J.K., Johnston, K.M., Collie, A., McCrory, P., & Ptito, A. (2007). A validation of the post concussion symptom scale in the assessment of complex concussion using cognitive testing and functional MRI. *Journal of Neurology, Neurosurgery & Psychiatry, 78*(11), 1231-1238.

Cohen, M.J., & Stanczak, D.E. (2000). On the reliability, validity, and cognitive structure of the Thurstone Word Fluency Test. *Archives of Clinical Neuropsychology, 15*(3), 267-279.

Collins, M.W., Grindel, S.H., Lovell, M.R., Dede, D.E., Moser, D.J., Phalin, B.R., . . . Daugherty, M.K. (1999). Relationship between concussion and neuropsychological performance in college football players. *JAMA: the journal of the American Medical Association, 282*(10), 964-970.

Covassin, T., Swanik, C.B., Sachs, M., Kendrick, Z., Schatz, P., Zillmer, E., & Kaminaris, C. (2006). Sex differences in baseline neuropsychological function and concussion symptoms of collegiate athletes. *British journal of sports medicine, 40*(11), 923-927.

Covassin, T., Swanik, C.B., & Sachs, M.L. (2003). Sex differences and the incidence of concussions among collegiate athletes. *Journal of athletic training, 38*(3), 238.

Dick, RW. (2009). Is there a gender difference in concussion incidence and outcomes? *British journal of sports medicine, 43*(Suppl 1), i46-i50.

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

Farace, E., & Alves, W.M. (2000). Do women fare worse: a metaanalysis of gender differences in traumatic brain injury outcome. *Journal of neurosurgery, 93*(4), 539-545.

Field, M., Collins, M.W., Lovell, M.R., & Maroon, J. (2003). Does age play a role in recovery from sports-related concussion? A comparison of high school and collegiate athletes. *Journal of Pediatrics, 142*, 546-553.

Gessel, L.M., Fields, S.K., Collins, C.L., Dick, R.W., & Comstock, R.D. (2007). Concussions among United States high school and collegiate athletes. *Journal of athletic training, 42*(4), 495.

Giza, C.C., & Hovda, D.A. (2001). The neurometabolic cascade of concussion. *Journal of Athletic Training, 36*(3), 228.

Green, P., & Iverson, G.L. (2001). Validation of the computerized assessment of response bias in litigating patients with head injuries. *The Clinical Neuropsychologist, 15*(4), 492-497.

Guilmette, T.J., & Rasile, D. (1995). Sensitivity, specificity, and diagnostic accuracy of three verbal memory measures in the assessment of mild brain injury. *Neuropsychology, 9*(3), 338.

Guskiewicz, K.M., Mihalik, JP, Notebaert, AJ, Hooker, DN, Oliaro, SM, Crowell, DH, & Ciocca, MF. (2005). Recurrent concussion in a collegiate football player equipped with a Head Impact Telemetry System. *J Athl Train, 40*.

Guskiewicz, K.M., Ross, S.E., & Marshall, S.W. (2001). Postural stability and neuropsychological deficits after concussion in collegiate athletes. *Journal of athletic training, 36*(3), 263.

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

Hootman, J.M., Dick, R., & Agel, J. (2007). Epidemiology of collegiate injuries for 15 sports: summary and recommendations for injury prevention initiatives. *Journal of athletic training, 42*(2), 311.

Inglese, M., Makani, S., Johnson, G., Cohen, B.A., Silver, J.A., Gonen, O., & Grossman, R.I. (2005). Diffuse axonal injury in mild traumatic brain injury: a diffusion tensor imaging study. *Journal of neurosurgery, 103*(2), 298-303.

Iverson, G.L., Brooks, B.L., Collins, M.W., & Lovell, M.R. (2006). Tracking neuropsychological recovery following concussion in sport. *Brain Injury, 20*(3), 245-252.

Iverson, G.L., Lovell, M.R., & Collins, M.W. (2003). Interpreting change on ImPACT following sport concussion. *The Clinical Neuropsychologist, 17*(4), 460-467.

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.

Kontos, A.P., Elbin, RJ, Schatz, P., Covassin, T., Henry, L., Pardini, J., & Collins, M.W. (2012). A Revised Factor Structure for the Post-Concussion Symptom Scale Baseline and Postconcussion Factors. *The American Journal of Sports Medicine*.

Lacritz, L.H., & Cullum, C.M. (1998). The hopkins verbal learning test and CVLT: a preliminary comparison. *Archives of Clinical Neuropsychology, 13*(7), 623-628.

Lacritz, L.H., Cullum, C.M., Weiner, M.F., & Rosenberg, R.N. (2001). Comparison of the hopkins verbal learning test-revised to the California verbal learning test in Alzheimer's disease. *Applied neuropsychology, 8*(3), 180-184.

Lacy, M.A., Gore Jr, P.A., Pliskin, N.H., Henry, G.K., Heilbronner, R.L., & Hamer, D.P. (1996). Verbal fluency task equivalence. *The Clinical Neuropsychologist, 10*(3), 305-308.

Leininger, B.E., Gramling, S.E., Farrell, A.D., Kreutzer, J.S., & Peck, EA. (1990). Neuropsychological deficits in symptomatic minor head injury patients after concussion and mild concussion. *Journal of Neurology, Neurosurgery & Psychiatry, 53*(4), 293-296.

Levine, A.J., Miller, E.N., Becker, J.T., Selnes, O.A., & Cohen, B.A. (2004). Normative data for determining significance of test–retest differences on eight common neuropsychological instruments. *The Clinical Neuropsychologist, 18*(3), 373-384.

Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological Assessment, 4th Edition* (4th ed.). New York: Oxford University Press.

Lezak, M.D., Howieson, D.B., & Loring, D.W. (2005). *Neuropsychological Assessment, 4th Edition* (4th ed.). New York: Oxford University Press.

Lövdén, M., Li, S.C., Shing, Y.L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. *Neuropsychologia, 45*(12), 2827-2838.

Lovell, M.R., Iverson, G.L., Collins, M.W., Podell, K., Johnston, K.M., Pardini, D., . . . Maroon, J.C. (2006). Measurement of symptoms following sports-related concussion: reliability and normative data for the post-concussion scale. *Applied neuropsychology, 13*(3), 166-174.

Maalouf, F.T., Clark, L., Tavitian, L., Sahakian, B.J., Brent, D., & Phillips, M.L. (2012). Bias to negative emotions: A depression state-dependent marker in adolescent major depressive disorder. *Psychiatry Research*.

MacDonald, S.W.S., Li, S.C., & Bäckman, L. (2009). Neural underpinnings of within-person variability in cognitive functioning. *Psychology and aging, 24*(4), 792.

Majerske, Cynthia W, Mihalik, Jason P, Ren, Dianxu, Collins, Michael W, Reddy, Cara Camiolo, Lovell, Mark R, & Wagner, Amy K. (2008). Concussion in sports: postconcussive activity levels, symptoms, and neurocognitive performance. *Journal of athletic training, 43*(3), 265.

Makatura, T.J., Lam, C.S., Leahy, B.J., Castillo, M.T., & Kalpakjian, C.Z. (1999). Standardized memory tests and the appraisal of everyday memory. *Brain Injury, 13*(5), 355-367.

Man, D.W., & Li, R. (2002). Assessing Chinese Adults' Memory Abilities. *Clinical gerontologist, 24*(3-4), 27-36.

Mansell, J., Tierney, R.T., Sitler, M.R., Swanik, K.A., & Stearne, D. (2005). Resistance training and head-neck segment dynamic stabilization in male and female collegiate soccer players. *Journal of athletic training, 40*(4), 310.

Marar, M., McIlvain, N.M., Fields, S.K., & Comstock, R.D. (2012). Epidemiology of concussions among United States high school athletes in 20 sports. *The American Journal of Sports Medicine, 40*(4), 747-755.

McAllister, T.W., Saykin, AJ, Flashman, LA, Sparling, MB, Johnson, SC, Guerin, SJ, . . . Yanofsky, N. (1999). Brain activation during working memory 1 month after mild traumatic brain injury A functional MRI study. *Neurology, 53*(6), 1300-1300.

McAllister, T.W., Sparling, M.B., Flashman, L.A., Guerin, S.J., Mamourian, A.C., & Saykin, A.J. (2001). Differential working memory load effects after mild traumatic brain injury. *Neuroimage*, *14*(5), 1004-1012.

McCaffrey, M.A., Mihalik, J.P., Crowell, D.H., Shields, E.W., & Guskiewicz, K.M. (2007). Measurement of head impacts in collegiate football players: clinical measures of concussion after high-and low-magnitude impacts. *Neurosurgery*, *61*(6), 1236-1243.

McClincy, M.P., Lovell, M.R., Pardini, J., Collins, M.W., & Spore, M.K. (2006). Recovery from sports concussion in high school and collegiate athletes. *Brain Injury, 20*(1), 33-39.

McCrea, M., Guskiewicz, KM, Marshall, SW, & Barr, W. (2003). Acute effects and recovery time following concussion in collegiate football players. *JAMA, 290*, 2556-2563.

McCrea, M., Hammeke, T., Olsen, G., Leo, P., & Guskiewicz, K. (2004). Unreported concussion in high school football players: implications for prevention. *Clinical Journal of Sport Medicine*, *14*(1), 13-17.

McCrea, M., & Powell, M.R. (2012). Biomechanics and Pathophysiology of Concussion. *Pediatric and Adolescent Concussion*, 29-53.

Newman, J., Barr, C., Beusenberg, M., Fournier, E., Shewchenko, N., Welbourne, E., & Withnall, C. (2000). *A new biomechanical assessment of mild traumatic brain injury. Part 2: results and conclusions.* Paper presented at the Proceedings of the 2000 International IRCOBI Conference on the Biomechanics of Impact, September 20-22, 2000, Montpellier, France.

Newman, J., Beusenberg, M., Fournier, E., Shewchenko, N., Withnall, C., King, A., . . . Thibault, L. (1999). *A new biomechanical assessment of mild traumatic brain injury. Part I: Methodology.* Paper presented at the Proceedings of the 1999 International IRCOBI Conference on the Biomechanics of Impact, September 23-24, 1999, Sitges, Spain.

Niogi, SN, Mukherjee, P., Ghajar, J., Johnson, C., Kolster, RA, Sarkar, R., . . . Manley, GT. (2008). Extent of microstructural white matter injury in postconcussive syndrome correlates with impaired cognitive reaction time: a 3T diffusion tensor imaging study of mild traumatic brain injury. *American Journal of Neuroradiology, 29*(5), 967-973.

Ommaya, A.K., & Gennarelli, TA. (1974). Cerebral concussion and traumatic unconsciousness correlation of experimental and clinical observations on blunt head injuries. *Brain, 97*(1), 633-654.

Pellman, E.J., Powell, J.W., Viano, D.C., Casson, I.R., Tucker, A.M., Feuer, H., . . . Robertson, D.W. (2004). Concussion in professional football: Epidemiological features of game injuries and review of the literature-Part 3. *Neurosurgery, 54*(1), 81-96.

Pellman, E.J., Viano, D.C., Tucker, A.M., Casson, I.R., & Waeckerle, J.F. (2003). Concussion in professional football: reconstruction of game impacts and injuries. *Neurosurgery, 53*(4), 799-814.

Pfeffer, R. I., Kurosaki, T. T., Harrah Jr, C. H., Chance, J. M., Bates, D., Detels, R., & Butzke, C. (1981). A survey diagnostic tool for senile dementia. *American journal of epidemiology, 114*(4), 515.

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

Powell, J.W., & Barber-Foss, K.D. (1999). Traumatic brain injury in high school athletes. *JAMA: the journal of the American Medical Association, 282*(10), 958-963.

Rabinowitz, A.R., & Arnett, P. A. (2013). Intraindividual Cognitive Variability Before and After Sports-Related Concussion. *Neuropsychology, 27*(4), 481-490. doi: Doi 10.1037/A0033023

Rabinowitz, A.R., & Arnett, P.A. (in press). Cross-domain intraindividual cognitive variability pre- and post-sports related concussion. *Neuropsychology*.

Ramanathan, D.N., Rabinowitz, A.R., Barwick, F.H., & Arnett, P.A. . (2012). Validity of Affect Measurements in Evaluating Symptom Reporting in Athletes. *Journal of the International Neuropsychological Society, 18*, 101-107.

Randolph, C., McCrea, M., & Barr, W.B. (2005). Is neuropsychological testing useful in the management of sport-related concussion? *Journal of athletic training, 40*(3), 139.

Reynolds, C.R. . (2002). *Comprehensive Trail Making Test (CTMT)*. Austin, TX: Pro-Ed.

Schatz, P., Kontos, A., & Elbin, RJ. (2012). Response to Mayers and Redick:"Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: Psychometric issues". *Journal of Clinical and Experimental Neuropsychology, 34*(4), 428-434.

Schatz, Philip, & Ferris, Charles S. (2013). One-Month Test–Retest Reliability of the ImPACT Test Battery. *Archives of Clinical Neuropsychology*.

Scher, C.D., Ingram, R.E., & Segal, Z.V. (2005). Cognitive reactivity and vulnerability: Empirical evaluation of construct activation and cognitive diatheses in unipolar depression. *Clinical Psychology Review, 25*(4), 487-510.

Shapiro, A.M., Benedict, R.H.B., Schretlen, D., & Brandt, J. (1999). Construct and concurrent validity of the Hopkins Verbal Learning Test–revised. *The Clinical Neuropsychologist, 13*(3), 348-358.

Smits, M., Dippel, D.W.J., Houston, G.C., Wielopolski, P.A., Koudstaal, P.J., Hunink, MG, & van der Lugt, A. (2008). Postconcussion syndrome after minor head injury: brain activation of working memory and attention. *Human brain mapping, 30*(9), 2789-2803.

Speer, D.C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting & Clinical Psychology, 60,* 402-408.

Strauss, E., Sherman, E.M.S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Vanderploeg, R.D., Curtiss, G., & Belanger, H.G. (2005). Long-term neuropsychological outcomes following mild traumatic brain injury. *Journal of the International Neuropsychological Society, 11*(3), 228-236.

Viano, D.C., Hamberger, A., Bolouri, H., & Säljö, A. (2009). Concussion in professional football: animal model of brain injury-part 15. *Neurosurgery, 64*(6), 1162-1173.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale--Third Edition*. San Antonio: The Psychological Corporation.

Woods, S.P., Delis, D.C., Scott, J.C., Kramer, J.H., & Holdnack, J.A. (2006). The California Verbal Learning Test–second edition: test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. *Archives of Clinical Neuropsychology, 21*(5), 413-420.

Zhang, L., Yang, K.H., & King, A.I. (2004). A proposed injury threshold for mild traumatic brain injury. *Transactions-American Society of Mechanical Engineers Journal of Biomechanical Engineering*, *126*(2), 226-236.