AN ANALYSIS OF THE USE OF SUBJECTIVE ORGANIZATION ON THE SELECTIVE REMINDING TEST IN A SAMPLE OF ADULTS WITH TRAUMATIC BRAIN INJURY

A Dissertation

Presented to

The Faculty of the Department

Of Psychology

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

By

Fedora O. Biney

August, 2013

AN ANALYSIS OF THE USE OF SUBJECTIVE ORGANIZATION ON THE SELECTIVE REMINDING TEST IN A SAMPLE OF ADULTS WITH TRAUMATIC BRAIN INJURY

Fedora O. Biney

APPROVED:

H. Julia Hannay, Ph.D.
Committee Chair

Steven Pluth, Ph.D.

Paul Massman, Ph.D.

Siva Tian, Ph.D.

Margaret Struchen, Ph.D.

John W. Roberts, Ph.D.
Dean, College of Liberal Arts and Social Sciences
Department of English

USE OF SUBJECTIVE EXECUTIVE STRATEGIES IN VERBAL LEARNING IN A SAMPLE OF ADULTS WITH TBI

An Abstract of a Dissertation

Presented to

The Faculty of the Department

Of Psychology

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor in Philsophy

By

Fedora O. Biney

August, 2013

ABSTRACT

Objective: The current study evaluated the use of subjective organization (SO), an indicator of organized recall, on the Selective Reminding Test (SRT) in a TBI group at 3 and 6 months post injury as well as the relationship between SO and SRT performance. The study also examined changes in SO from 3 to 6 months post injury in a longitudinal subsample.

Participants and Methods: Patients with complicated mild, moderate and severe traumatic brain injury (TBI) were administered the Selective Reminding Test (SRT) at three (N = 121) and six months (N= 87) post injury. Hierarchical regression models were conducted to evaluate predictors of two SO measures at both time points. Repeated measures ANCOVA was used to evaluate changes in SO within a subsample of patients (N = 75) who completed the SRT at both time points to control for injury severity, age and gender.

Results: Best Day 1 GCS and age were consistent predictors of SO at three and six months post injury. No significant changes in SO were observed in a subsample of patients who completed the SRT at both three and six months post injury. SO was positively associated with SRT measures of long term storage and retrieval and negatively associated with short term storage and retrieval at both time points.

Conclusions: Injury severity and age are predictors of SO in the first six months after TBI. There are no significant increases in SO in the first six months after TBI. SO is positively related to SRT performance.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Julia Hannay, for going above and beyond in guiding me through the theoretical and practical aspects of this dissertation. This dissertation would not have been possible without her years of steadfast dedication to TBI research, which has produced a wealth of meticulous data that made this work possible. I thank Dr. Hannay for her mentorship throughout the years and teaching me to bring clinical compassion to my research and the spirit of scientific inquiry to my clinical work. I am a better clinician, researcher, scientist and writer because of her influence.

I would like to thank my committee members Dr. Paul Massman, Dr. Siva Tian, Dr. Margaret Struchen and Dr. Steve Pluth for their time, effort, patience and flexibility throughout the writing of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

DEDICATION

I would like to thank my family for their support – my brother and sisters for the joy they bring to me, my dad for his help with equations and math theories and my mom for her unfailing support in this long journey, her constant encouragement and for always believing in me.

An analysis of the use of subjective organization on the selective reminding test in a sample of adults with traumatic brain injury

Word list learning tasks have become useful tools for assessing memory status after acquired brain injury. Though these tasks tend to differ in a variety of ways such as list length and number of learning trials, word list learning tasks provide a method for determining several aspects of memory and learning including rate of learning, spontaneous recall of the words after varying delay lengths, the effect of interference on recall, recognition of words and subjective strategies an individual uses to learn and later recall list items. The constructs measured by list learning tasks include, for instance, encoding, attention, storage, spontaneous retrieval, and recognition. Furthermore, the free recall list learning paradigm allows one to evaluate how an individual may re-organize information to facilitate learning. The current study will evaluate the use of one subjective organization strategy (intertrial repetitions) by patients of varying traumatic brain injury (TBI) severities on the Selective Reminding Test (SRT).

To enhance the reader's appreciation of relevant issues, the paper will begin with a discussion of the two-store memory model on which the SRT is based and other widely accepted memory models. The remaining sections will review research on the function and measurement of organization in learning, memory deficit after TBI and the use of subjective organizational strategies on verbal learning tasks following TBI. Finally, the present problem will be readdressed in light of the research reviewed.

Short and Long Term Memory Models

Two Store Memory Models

Waugh and Norman (1965) proposed an early model focusing on short term memory (STM) that attempted to explain the short nature of digit span memory (less than 10 digits) and its rapid decay rate when there is no rehearsal of digits. In their model, two separate memory stores exist: Primary memory (PM) and secondary memory (SM) (terms borrowed from William James (1890)). PM was conceptualized as limited in nature and it is analogous to STM. All stimuli attended to were said to enter PM. SM, which is analogous to long term memory (LTM), was considered to be more permanent and information could be transferred from PM to SM with rehearsal (Waugh & Norman, 1965). Unrehearsed items in PM that were not transferred to SM could be displaced by newer items, and once displaced, these older items were lost permanently. Once rehearsal ended, and the recall of information began, items in PM were rapidly forgotten. The often found recency effect seen in free recall of items (e.g., words, digits, etc.) in which "just read items" tend to be recalled with ease was explained as the dumping of items from this temporary store before they are forgotten. As further evidence that the PM store contains the most recently presented items and that PM is subject to rapid forgetting, Lewis (1971) conducted a study in which participants listened to a list of 18 words and recalled the list twice immediately after the single presentation with a 30 second interval between recalls. Lewis (1971) found that word recall was influenced by the serial position of the words on the list. In the first recall, the highest recall was for words at the end and beginning of the list, while recall for words in the middle was the lowest. In the second recall, which was done only 30 seconds after the first, word recall was similar to that of the first recall, with the exception of words at the end of the list, which had a lower probability of being recalled. These findings were used as evidence in support of a two-store memory model. Items at the end of the list had a high initial recall presumably because these items were being dumped from STM. However, these items, because of the temporary storage nature of STM, were not available to be recalled even 30 seconds later during the second recall. Items earlier in the list were recalled at largely the same rate between the two recalls because these items had been transferred from the temporary STM to long term memory (LTM) where they were able to be recalled in both recall trials.

Atkinson and Shiffrin's Two Store Memory Model

Atkinson & Shiffrin (1968) proposed a model similar to Waugh and Norman in which memory was composed of two distinct dimensions: One dimension contained the permanent structural features and included three components: A sensory register (SR), a short term store (STS) and a long term store (LTS). The second dimension referred to control processes that could be selected at will by an individual and used to facilitate learning in different situations. The three structural components were considered fixed. Sensory information supposedly entered the SR, which holds information only temporarily before it decays and is permanently lost. The STS was considered "the subject's working memory" (p. 90) and received information from both the SR and the LTS (Atkinson & Shiffrin, 1968). The STS, like the SR, was a temporary store, but the delay in decay could be extended with rehearsal of information in this store. Finally, the LTS received information from the STS and was considered to be relatively permanent. Atkinson and Shiffrin (1968) emphasized that the transfer of information from one store to another (ex. from STS to LTS) was essentially a copy of information from one store to another, rather than a removal of information from one store to another store. Empirical evidence supporting this two store model come from several sources including (1) dissociations in memory impairments for patients with bitemporal hippocampal lesions and individuals with Korsakoff's syndrome and (2) the model's ability to account for phenomenon observed in empirical memory experiments such as the rapid forgetting observed in the Brown-Peterson paradigm, the primacy-recency effect and digit memory span.

Atkinson and Shiffrin (1968) reported the findings of Milner (1966) in which patients with bitemporal hippocampal lesions and individuals with Korsakoff's syndrome had relatively preserved SR and STS as evidenced by their ability to perform digit span and dichotic listening tasks. In addition, their recall for premorbid facts and events was also preserved. Their primary deficit was in the ability to add new memories to LTS and retain new information. In several related experimental studies, an experimental group composed of individuals with unilateral hippocampal damage was injected with sodium amytal to inactivate the opposite, undamaged hemisphere. Control group participants included patients with no hippocampal damage, as well as patients with unilateral hippocampal damage who received sodium amytal injections to inactivate their damaged hemisphere. All participants were presented with four pictures, distracted, and then asked to identify the original four pictures from nine pictures. Deficits in picture recall occurred in individuals with bilateral hippocampal damage and in individuals whose intact hemisphere was inactivated, leaving only the damaged hemisphere active (Atkinson & Shiffrin, 1968). Individuals whose damaged hemisphere was inactivated experienced no memory deficit on the task (Atkinson & Shiffrin, 1968). Memory deficits observed in individuals with Korsakoff's syndrome included deficits in their ability to retain information for longer than several minutes, while recall for information learned premorbidly was unimpaired, along with immediate memory span. Such findings offered support for the existence of a separate STS and LTS in Atkinson and Shiffrin's model.

Further evidence in support of a two-store model of memory comes from the Brown-Peterson paradigm in which participants are read a trigram of consonants such as *CBY,* and then are asked to complete a distracter task (e.g., such as a serial subtraction task) for a given number of seconds before being asked to recall the trigram. Experimenters found that as the length of the distracter task increased (from 3 to 18 seconds), the proportion of trigrams recalled decreased from 80% to 10% (Peterson & Peterson, 1959). In terms of Atkinson and Shiffrin's two-store memory model, the amount of information transferred to LTS and subsequently recalled depends on the extent to which the trigram was rehearsed while in STS. The rapid rate of forgetting of the trigram occurs because the distracter task limits the amount of rehearsal possible.

The primacy-recency effect describes what is commonly observed in a serial position curve in which the proportion of freely recalled items on a memory task is plotted against the absolute position of the item on the original presentation of the list. As described earlier, words at the beginning of the list are recalled in greater proportion (primacy effect) relative to words in the middle of the list. Words at the end of the list are recalled in an even greater proportion (recency effect) relative to items in the middle of the list. Atkinson and Shiffrin's explanation for this was similar to that of Waugh and Norman (1965) in that initial list items have greater rehearsal times in STS, allowing more complete transfer of these items into LTS where they are retrieved on subsequent recall trails. The recency effect is due to the fact that items at the end of the list have just been presented and can be immediately recited from STS with little effort. Alternatively, if recall is delayed, this model predicts that the primacy effect will hold, while the recency effect should disappear because these items are afforded little to no transfer to LTS for later recall. Experiments in long term free recall support this prediction as list items at the beginning of the list have a greater probability of being recalled after a longer delay compared to words at the end of the list (Craik, 1970; Glanzer & Cunitz, 1966). Finally, the two-store memory model is used to explain the relatively short serial digit span of approximately seven digits described by Miller (1956). The model proposes that STS is a limited store primarily because items can only be rehearsed one at a time. Therefore, as the length of items to be rehearsed increases, new items to be rehearsed must replace older items.

The second dimension of memory described by Atkinson and Shiffrin (1968) encompassed the existence of modifiable control processes that could be selected at will by an individual and used to facilitate learning in different situations. Examples of these control processes included rehearsal, specific search strategies and coding procedures. The nature of the three memory structures and associated control processes of each structure are described in more detail below.

Memory Structure: The Sensory Register

In Atkinson & Shiffrin's (1968) theoretical model of memory structure, sensory information entered the sensory register (SR), where it was stored for a very short time (several hundred milliseconds). The information then decayed and was permanently lost to this register. SR information could be sent to the STS where it remained for a longer period of time (in the order of 15-30 seconds). The mechanism of the SR was best understood in terms of short term visual images. For example, in partial report experimental techniques, participants were briefly shown a matrix of letters and numbers with a tachistoscope and then were immediately given a signal indicating which row of items they should report. Individuals were usually able to complete this task with success. However, this presentation was described as being sensitive to interference with subsequent visual stimuli that would either eliminate or change the original stimulus presented. In their model, Atkinson and Shiffrin assumed the report being made by participants was for information contained in the sensory register given the very brief stimulus presentation, immediate recall demands, very fast decay of the information and its sensitivity to interference. They noted that the processes of other sensory registers besides the visual sensory register were less well known.

Control Processes in the Sensory Register

Control processes in the SR functioned to help select which portions of incoming stimuli to attend to and transfer to the STS given that the SR received a large amount of information that decayed very quickly (Atkinson & Shiffrin, 1968). These control processes included selective attention to specific sensory registers (i.e., visual, auditory, etc.). As an example of the utilization of such a control process, they reported that individuals can report a specific sense modality when given simultaneous inputs from different senses easily if given specific instructions of which modality to report prior to receiving the inputs. However, their performance in reporting any single sense modality decreases when asked to report a specific modality after the simultaneous inputs are presented. Another control process described is the selective scanning of the SR, which allowed a select portion of large inputs to be transferred to the STS. The use of a matching strategy in which inputs to SR are matched to information from LTS in order to identify the input was considered another control process. For example, when tasked with finding either the letter *F* or *B* among a visual matrix of letters, two efficient strategies could be employed: Performing a match of features for each letter on the matrix against the features of the letters *F* and *B* (such information would be stored in LTS) to find a match. A second strategy involves performing a match with only one letter, *F* for example, and guessing the other letter, *B*, if the scan finishes and a match of the searched letter is not found.

Memory Structure: Short Term Store

The short term store (STS) was assumed to receive inputs from both the SR and the long term store (LTS), and inputs to the STS would decay rapidly (in about 30 seconds), after which, the input would be permanently lost to this store. Decay of STS inputs could supposedly be delayed through the use of control processes such as rehearsal, which kept the memory trace in STS refreshed. In their model, Atkinson and Shiffrin (1968) refer to a specific type of STS known as the auditory-verbal-linguistic short term store or a-v-l STS. The main sensory register they studied in their model was the visual sensory register. They noted that visual images in the SR could be transferred to an a-v-l STS rather than a visual STS either due to the response required of the participant (e.g., written response), or to constraints on the visual STS that prevent rehearsal capabilities. They hypothesized that the transfer process of a visual image from the SR to the a-v-l STS occurred through an intermediate step involving a search of information in long term storage. For example, if a person is presented with a visual image, for each visual item in the register, the person performs a search through the long term store for the verbal label associated with the visual image. Then this label is transferred to the STS, allowing for rehearsal in a new auditory-verbal-linguistic modality. They declared that distinctions between verbal, auditory and linguistic aspects of STS are difficult to make, and collapsed them into one store- the a-v-l STS. They applied this model of an a-v-l STS to explain findings of several experiments. In one experiment conducted by Milner (1968), patients were presented with stimuli that were not easily coded verbally such as light flashes and nonsense figures. Patients were presented with a specific stimulus, followed by a distraction, and then presented with another stimulus, and asked to determine if the two stimuli were the same or different. Atkinson & Shiffrin (1968) reported that patients with the most severe memory deficit performed at a level expected by chance after a 60 second delay, regardless of the presence of a distracting task. Atkinson and Shiffrin believed the poor performance indicated that either no verbal rehearsal took place (since the nature of the stimuli made this difficult), or that rehearsal strategies in other modalities were not effective. While Atkinson and Shiffrin did not dismiss the idea of other short term stores, such as a visual or kinesthetic STS, they did not find as convincing evidence in support of their structure and control processes as compared to the a-v-l STS, and essentially refer to the STS system as an auditory-verbal-linguistic store.

Control Processes in the Short Term Store

Rehearsal is a control process that serves to increase the length of time that information stays in the STS and therefore, increases the time that a trace or copy of the STS information is simultaneously created in LTS. It also provides time for coding and other processes to operate. Each rehearsal regenerates the STS trace, preventing decay. The amount of information that can be retained in STS through rehearsal was thought to be dependent on the rate of decay in STS and the trace being formed in STS through rehearsal. Atkinson and Shiffrin (1968) stated that ordered rehearsal, such as what typically occurs in a digit span task, resulted in the highest maintenance of STS items. On such tasks, rehearsal taking place repeatedly in a fixed order could adequately maintain a small number of items (5-8) indefinitely, as opposed to rehearsal in a random order, which was subject to item loss and decay (Atkinson & Shiffrin, 1968). Two experimental procedures which were assumed to minimize or eliminate rehearsal have been used to study decay in STS: Brief presentation of stimuli followed by an immediate recall test, and brief presentation of stimuli followed by an interference activity. However, Atkinson and Shiffrin called into question the ability of both of these methods to truly control rehearsal, suggesting that subjects could intentionally modify what stimuli they attend to in order to allow for rehearsal.

In free recall experimental paradigms, in which the number of items to be recalled exceeds the rehearsal capacity, a different rehearsal strategy called a rehearsal buffer was thought to be employed (Atkinson & Shiffrin, 1968). In this buffer, old rehearsed items in the STS were replaced by new items, to allow at least some rehearsal of each new input. Information in the buffer came from items in STS which had been transferred to STS from either the LTS or SR. As new items entered the buffer, old ones were displaced, and decayed at a fast rate. Any retention then for these displaced items was assumed to be due to whatever trace was created in LTS while the item was being rehearsed in the buffer. Other control processes associated with the buffer included choosing what items entered into and were removed from the buffer.

Coding was another control process identified by Atkinson and Shiffin that involved the addition of and/or modification of information in STS based on a search of LTS. As an example, Atkinson and Shiffrin (1968) point out that a word presented visually could not enter the a-v-l STS until the verbal presentation of the word had been found through a search and match process in the LTS. In the case of unfamiliar visual stimuli, they acknowledged that the visual image itself may enter a visual STS without any attempt at verbal coding. Atkinson and Shiffrin observed that in experimental procedures, participants tend to use a coding strategy in several situations, including when there are a large number of easily coded items, and when there is a long delay between the stimulus presentation and the recall test. Other related STS control processes they identified included grouping, or memorizing groups or sets based on a common element rather than individual items, as well as chunking and organizing.

Memory structure: Long Term Store

The LTS was the relatively permanent store among the three memory structures, and received information transferred from the STS. As an example, a series of items enters the sensory register, and is then transferred to the a-v-l STS, where it is rehearsed. During this rehearsal period, a trace or copy is created in LTS, and the more rehearsal that takes place, the stronger the LTS trace becomes. Control processes determine the nature and extent of the information transferred from STS to LTS. Transfer could be probabilistic, continuous, or some combination of the two and depended on the type of test and information being transferred. They acknowledged that each sensory modality must carry an analogous LTS and other types of LTS that are not modality specific such as temporal memory, were acknowledged.

The LTS was assumed to contain multiple traces or copies of information, which were conceived as being a combination of incomplete and complete traces. Searching through LTS may lead to an incomplete trace, and therefore no definite response in a free response format, which conceivably explained the tip-of-the-tongue phenomenon. When a task required only the recognition of a correct item among a set of alternatives, recognition of the correct response allowed the individual to access other traces through an associative process, and these traces had enough of the missing information to make subjects sure their responses were correct.

Control Processes in the Long Term Store

The above mentioned processes of coding and rehearsal that occur in STS contribute to the transfer of information from STS to LTS (Atkinson & Shiffrin, 1968). In their model, Atkinson and Shiffrin (1968) assert that information transferred through the process of rehearsal is weak and vulnerable to subsequent interference of other rehearsed items, and contrast this low quality of transfer to that induced by a coding process, which they assert strengthens the trace of stored information. In one example they used findings from an experiment conducted by Montague, Adams, & Kiess (1966) to support their proposal of coding strategy use in augmenting storage in LTS. In the experiment, nonsense syllables were presented in pairs to participants, who wrote any word, phrase or sentence they could think of associated with the pair. After 24 hours, the participants were asked to recall the paired syllables, as well the association they developed for each pair. There was greater retention of nonsense-syllable pairs on which the associations were retained, while the recall for syllable pairs on which the association was forgotten or changed was negligible. The ability of coding techniques to improve recall performance depends on the LTS structure and several possibilities were mentioned including the possibility that encoding makes use of pre-existing strong associations, and that coding techniques allow a smaller area of memory (just associates of the item rather than an entire memory store) to be searched during recall testing. A code could also conceivably increase the amount of information stored, and finally, coding could protect developing associations from interference of future items (Atkinson & Shiffrin, 1968).

Memory Retrieval in the Atkinson and Shiffrin Model

The search through LTS for stored information was thought to begin with finding a trace in LTS. The purpose of this search was two-fold: To locate a specific memory trace and to retrieve the trace. This process was assumed to be evident during the latency period while one attempted to recall well known information and in the tip-of-the tongue phenomenon, when an individual cannot recall a specific piece of information at the time. An example of the relationship between LTS traces and associative coding was presented by Atkinson and Shiffrin (1968). They observed that while quizzing a graduate student on the state capitals, he was unable to recall the capital of the state of Washington. Only when he recalled the capital of another state, Oregon, was he able to remember that Olympia was the capital of Washington, and explained that he had learned the two capitals together, and the remembrance of one capital conjured up the other. Atkinson and Shiffrin attributed the student's initial recall difficulty to a faulty initial search process that did not lead to a fruitful trace. An unsuccessful search process is likely to occur when recalling items with no natural order to them, such as in the verbal recall of a word list, which subsequently leads to gaps in recall. They expanded on the issue of how one determines when a search is unsuccessful and terminates it, and posited several rules someone may employ to decide to stop searching including a subjective time limit, an event counter that would stop searching when a number of events such as a number of searches, or number of incorrect traces were found, or an events counter that would determine whenever a number of searches recovered redundant unsuccessful traces (Atkinson & Shiffrin, 1968).

It is important to note that Atkinson and Shiffin (1968) differentiated the theoretical constructs of short term store (STS) and long term store (LTS) from short term memory (STM) and long term memory (LTM). STM and LTM were considered to be operational definitions used in experiments. STM was operationalized as the type of memory evaluated in single trial or short duration experiments, while LTM was operationalized as the memory evaluated in long duration experiments such as list learning experiments. Alternatively, STS and LTS referred to the theoretical constructs of stores that exist in human memory. The authors noted that STS and LTS are present in LTM and STM experiments. Others (Craik & Lockhart, 1972) have used the convention of referring to STS and LTS as storage systems, and STM and LTM to refer to experimental situations.

Criticisms of Atkinson and Shiffrin's Two Store Memory Model

Several criticisms of the Atkinson and Shiffrin (1968) memory model have been proposed. Criticisms center around the model's proposed unidirectional relationship between STS and LTS, the utility of rehearsal as a control process and lack of elaboration on the existence of multiple modalities of sensory registers in the model. One criticism of the Atkinson and Shiffrin model was its emphasis on the use of rehearsal as a control process, and particularly, in its proposed importance in the transfer of information from STS to LTS (Baddeley, 1976). Specifically, Baddeley (1976) noted that outside of a controlled laboratory setting information in everyday life is presented in a more meaningful form such as prose. Furthermore, in these everyday life situations, individuals tend to encode information semantically. Baddeley (1976) observed that even in the context of a laboratory setting, rote rehearsal often proved to be an inefficient method for transferring information to LTS. Of course, Atkinson and Shiffrin themselves acknowledged that rehearsal was an important control process in laboratory tasks due to the “concentrated, often meaningless, memory tasks” used in such settings (Atkinson and Shiffrin, 1976, p. 111), and acknowledged that rehearsal in everyday memory may be less often utilized. Baddeley (1976) also called into question Atkinson and Shiffrin’s assertion that semantic coding takes place in the STS.

Atkinson and Shiffrin did not expand their conceptualization of multiple modality-specific sensory registers to their conceptualization of the STS. Baddeley (1976) argued in favor of the view that modality specific short term stores existed, and presented findings from various studies supporting this view. He presented evidence suggesting recall is differentially affected by presentation modality (auditory vs. visual). For example, Murdock and Walker (1969) (as cited in Baddeley, 1976) conducted a free recall test of 20 words presented either aurally or visually at the same rate and found a more pronounced recency effect for the auditory presentation of list items.

A strong criticism of Atkinson and Shiffrin’s model is the notion that the operations of STS are essential for LTS, since both encoding to and retrieval from LTS are controlled and organized through the process of the STS (Baddeley, 1976). The strongest evidence Baddeley presented in his argument against this view was the case study of K.F., who was studied by Warrington and Shallice (1969). K.F., who had damage to the left parieto-occipital region, experienced deficits in short term memory as evidenced by impaired digit span and a limited recency effect on free recall, but exhibited surprisingly normal learning and recall. According to the Atkinson and Shiffrin model, such a syndrome should not occur because a faulty STS system would prevent the formation of a LTS memory trace as well as adequate retrieval from LTS. In response to this key limitation, Baddeley (1976) proposed a short term memory system that utilizes a parallel semantic coding system rather than the serial processing STS described in Atkinson and Shiffrin's (1968) model.

Baddeley and Hitch's Working Memory Model

Another model of short term memory proposed by Baddeley and Hitch (1974), developed as a challenge to the notion that STS, as proposed by Atkinson and Shiffrin (1968), was a unitary storage system. Instead, Baddeley and Hitch (1974) replaced the limited capacity unitary STS with a limited capacity working memory system, essentially a work space that could be allocated to both storage and control processes. Their model differed from the unitary STS in that they applied their working memory model to describe its role not only in learning and memory, but reasoning and comprehension (Baddeley & Hitch, 1974; Baddeley, 2003). Their model was also able to account for limitations a unitary STS system had in explaining neuropsychological damage that lead to deficits in STM without deficits in LTM. Working memory was conceived to have multiple components, including a phonological loop, a visuospatial sketchpad, and a central executive (Baddeley & Hitch, 1994). They described the phonological loop as an acoustic store that temporarily held speech information (for approximately 2 seconds), using a subvocal rehearsal strategy (articulatory rehearsal process) to refresh this information. Support for this construct was given by several different phenomena observed with respect to immediate recall of auditory items. The phonological similarity effect is the phenomenon in which serial recall is impaired when similar sounding items are attempted to be recalled. According to Baddeley and Hitch (1994), the phonological similarity among words appears to disrupt the memory trace of the correct item. Further evidence was provided by the fact that when other auditory stimuli such as music or speech are heard simultaneously with spoken items to be recalled, this could impair recall of the spoken items, possibly due to the addition of noise to the memory trace (Baddeley & Hitch, 1994). They also found an inverse relationship between spoken item syllable length or articulation rate and the immediate recall of those items --referred to as the word length effect (Baddeley & Hitch, 1974; Baddeley & Hitch, 1994). They posited that this word length effect is a product of the limitations of subvocal rehearsal speed and the capacity of the phonemic loop, which limits how quickly the temporary phonological memory trace can be refreshed. Studies with neuropsychological patients provided supporting evidence for the existence and structure of the phonological loop, and Baddeley (2003) points out those neuropsychological patients with select STM deficits and relatively preserved language abilities typically have lesions in the left temporoparietal area. These patients failed to show the above mentioned word length effect or word similarity effect when words were presented visually, which was interpreted as the abandoning use of a defective phonological loop of the working memory system. Other proposed functions of the phonological loop included assistance with the acquisition of language. Evidence in support of this function comes from findings suggesting that the same features that interrupt the phonological loop and recall, such as phonological similarity, word length and articulatory suppression, also interrupt foreign vocabulary acquisition (Baddeley, 2003).

The visuospatial sketchpad is the visual or spatial analog of the phonological loop (Baddeley & Hitch, 1994). Its existence has been supported by studies indicating impairments in immediate recall of spatial information when a simultaneous spatial interference task is performed by subjects. Furthermore, dissociations between visual processes, which focus on object appearance, and spatial processes, which focus on object location and direction, have been found, and appear to relate to the two visual processing paths for processing “what” information associated with object form and appearance, and a separate “where” pathway associated with object location and direction (Baddeley, 2003). Neuropsychological studies have shown evidence of a distinction between spatial and visual memory. In such experimental paradigms, testing of spatial memory is conducted using Corsi blocks, which involve the experimenter tapping a sequence of nine blocks, arranged in an array, and the participant, copying the tapping sequence, with the length of the tapping sequence increasing as the task progresses. A visual analog of this is a pattern span task in which participants are presented with a visual matrix on which half of the cells are filled, and the remaining half are empty. After the matrix is removed, the participant must attempt to recall which cells were filled. In healthy individuals, performance on the Corsi block task is interrupted more by spatial interference rather than visual interference, while the opposite pattern exists for pattern span (Baddeley, 2003). The features of the visual-spatial sketchpad were further elaborated to include two components: A visual storage component and an inner scribe, which is responsible for rehearsal and retrieval processes. Even decades after the initial proposal of a multi-component model of working memory by Baddeley and Hitch in 1974, Baddeley admits that visual-spatial working memory is a poorly integrated component of the model (Baddeley, 2003).

The central executive component of working memory is the attention control system that helps coordinate the simultaneous processes of the phonological loop and visuospatial sketchpad (Baddeley & Hitch, 1994). The main function of the central executive is to serve as an attentional controller that switches attention appropriately between the two subsystems (visualspatial sketchpad and phonological loop).

Criticisms of the Working Memory Model

Baddeley (1966) hypothesized that coding in STS was performed acoustically or phonemically and coding in the LTS was done semantically. Craik & Lockhart (1972) asserted that studies have not been entirely supportive of this distinction in coding. For example, they point to previous studies (Kroll et al., 1970; Levy, 1971; Peterson & Johnson, 1971) that suggest STS coding can be acoustic as well as articulatory and visual (even for verbal material). They also address a question that has not been properly answered with multistore models—whether STS can code semantically. The answer to this question appears to be dependent on the experimental paradigm used. Traditional STM paradigms suggest that the STS cannot semantically encode. However, Craik and Lockhart (1972) suggested that the distinction in differential coding is better explained by processing demand differences that exist in specific experimental paradigms, rather than by making STM and LTM distinctions. As an example, they pointed out that in some paradigms, semantic processing is both a possibility and advantageous to recall, whereas it is not possible in other paradigms (Craik & Lockhart, 1972). Crowder (1993) also stated that a confound exists in the proposed notion that the STS uses a phonetic coding system, and the LTS uses a semantic coding system. Specifically, this distinction makes it unclear whether coding (semantic vs. phonemic) or storage (STS vs. LTS) is more important. He proposed that to study this would require holding one factor constant (keeping the storage factor constant by studying STS or LTS only) while evaluating variations in the other factor (semantic vs. phonemic coding). This same confound affects the interpretation of deficits observed in neuropsychological patients because the distinction between coding and storage makes it difficult to determine whether their deficits in memory are due to a missing stage in processing, or due to missing machinery needed for coding (Craik & Lockhart, 1972).

Criticisms of Two Store Memory Models and Overview of Levels of Processing Theory

Crowder (1993) offers several general criticisms of two store memory models. He argues that the recency effect observed in free recall and distracter memory tasks is not adequately explained by a two store model. He argues that there is evidence that distraction does not significantly influence forgetting on initial learning and recall trials, which would suggest that long term memory plays a role in the retention of information for this initial trial, which downplays the role of short term memory in explaining performance on such distracter tasks (Crowder, 1993). He suggests that temporal distinctiveness better explains performance on such tasks. Temporal distinctiveness was conceptualized as the temporal analog to distinctiveness in perception, in which, for example, items at opposite ends of a spatial array are perceived as being more distinct than intermediate items. Crowder (1993) also challenges the idea that the recency effect is a hallmark example of forgetting in the short term memory system with evidence suggesting the presence of long term recency. He presented findings from Baddeley and Hitch (1977), who found a strong recency bias when they asked local pub patrons to recall the teams the local rugby team played that season. As further evidence, they pointed to an experiment conducted by Roediger and Crowder (1976) that found a primacy and recency effect in the recall of U.S. presidents by undergraduate students. The presidents from the most recent terms were remembered more often, as well as the first few U.S. presidents. Crowder (1993) argued that a unitary memory system was sufficient to explain this recency effect being exhibited in long term memory, without the need for a rehearsal buffer.

Craik and Lockhart (1972) presented several problems with two store memory models and questioned the utility of multistore models of memory in general. Specifically, they challenged some of the multistore notions about capacity, coding and retention in memory functioning. Two store memory models assumed the STS to be of limited storage capacity (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). According to Craik and Lockhart (1972), it is not clear what exactly this storage capacity is, or if the limitation is in storage capacity, processing rate ability, or a combination of the two. For example, they pointed to widely varying estimates of immediate memory span, which had a typical range from 5-9 items, but could also span up to 20 items in some studies. They further explained that although the multistore notion of strategy use such as chunking, can account for these large variations in immediate memory span, the existence of such processes implies the existence of a STS capable of coding both simple and complex features of input.

Finally Craik and Lockhart (1972) point out that the forgetting characteristics of each memory store should be consistent for each store across different experimental paradigms and they found this was not the case. Forgetting curves in paired associate learning tasks, which they found can extend over 20 intermediate items, differ from those of other paradigms such as free recall, which are associated with faster forgetting curves. They argued that retention is dependent on tasks demands such as study time, amount of material, test modality as well as familiarity and meaningfulness of the stimuli to the individual. They proposed that because of the limitations observed in capacity, coding and forgetting, the type and level of encoding should be studied in a fashion analogous to levels of processing involved in perception. Initial stages involve processing physical and sensory features of stimuli, and subsequent stages involve pattern recognition and extracting meaning (Craik & Lockhart, 1972). Rather than focusing on distinctions between a limited capacity STS and an unlimited capacity LTS, they replaced the STS with a limited capacity central processor that could process at several levels and in various modalities. They describe two types of processing. Type I processing maintains items in consciousness for continued processing, but does not influence the formation of a permanent memory trace. The importance of rehearsal and repetition, which was a critical component of transfer from the STS to LTS in Atkinson & Shiffrin model, was considered entirely dependent according to Craik and Lockhart's level's of processing theory, on whether the rehearsal involved deeper level or same level processing. That is, it was not the rehearsal itself that improves memory, but the depth of processing involved in the rehearsal. Type II processing that involves processing at a semantic and associative level, leads to the formation of a permanent memory trace. Type II processing was dependent on its appropriateness to the task demands and the degree to which deeper level processing was possible, although Craik & Lockhart, (1972) pointed out that processing at any level was holistic and individuals perceive items such as letters, images and sounds as wholes similar to object perception, and not as a individualized attributes.

Models of Recognition Memory

Recognition memory can be divided into two independent and parallel processes known as recollection, which encompasses specific information about the retrieved information, and familiarity, which encompasses a memory strength index for the information (Yonelinas, 2002). Considerable evidence has mounted in support of this dual process model of recognition memory.   Studies have found that decisions about the familiarity of items are faster than decisions about specific event information.   In one experiment conducted by Hintzman and Caulton (1994), participants were presented with a list of 30 mixed auditory and visual words, presented at a rate of 2 seconds per word. This was followed by one of two visual recognition trials requiring participants to make either yes/no recognition judgments or modality judgments regarding list items using a response-signal method that varies the time frame (100-2000 ms) allowed for a response. Accurate response times for above chance recognition judgments occurred faster (about 119 ms) than response times for modality judgments. In a separate series of experiments requiring participants to perform either an item or a list discrimination task, the minimum response time for list recognition (which was assumed to represent the familiarity process) was about 100 ms faster than the response time for item discrimination (which was assumed to represent the recollection process) (Hintzman, Caulton, & Levitin, 1998).

Studies have also found that as retrieval time increases, discrimination accuracy between previously learned list items (hits) and semantically related distracters (false alarms) is low initially, then increases, creating a biphasic time-accuracy response relationship (Dosher, 1984; Hintzman & Curran, 1994). This relationship is interpreted as a faster initial familiarity process that leads to the acceptance of false positive items, and a slower recognition process that allows for accurate retrieval and dismissal of incorrect responses.

Studies evaluating patterns in event-related potentials (ERPs) during recognition tests have found different temporal and spatial scalp distributions for ERPs associated with familiarity and recollection (Yonelinas, 2002). One ERP pattern known as the FN400 pattern located in the frontal brain region 300-500 ms after stimulus onset has been associated with accurate discrimination between new and old items, which is consistent with familiarity (Curran, 2000). A separate ERP pattern located in the parietal region at 400-800 ms after stimulus onset has been associated with accurate discrimination between old items and their plural forms (ex. pig vs.

pigs), which is consistent with recollection, as this task is thought to require a more detailed recollection of list items (Curran, 2000).

Several models have informed the current knowledge on recognition memory including signal detection and threshold theories. The theory of signal detection (TSD) has been found to adequately describe familiarity effects (Yonelinas, 2002). TSD assumes that there are two kinds of sensory input: A signal representing a stimulus that the subject is attempting to detect. Noise represents other stimuli associated with the environment or experimental task on any trial (Hannay, 1986). A proportion of trials in any TSD task are noise only trials.

For the rest of the trials, a fixed signal is added to whatever noise is present on that trial. Because the amount of sensory input derived from noise is assumed to vary randomly over time, its distribution is assumed to be normal with a standard deviation of 1 and a mean of zero (a standard normal distribution). The amount of sensory input provided by the presence of a signal on signal + noise trials is assumed to be fixed and thus the distribution of sensory input for a signal + noise trials is assumed to be normally distributed with a mean of d' and a standard deviation of 1 (Refer to Figures 1 and 2). D' represents the distance of the mean of the signal + noise distribution from the mean of the noise only distribution in standard score (z-score) units. If the sensory input from a signal on signal + noise trials includes a weak signal, this creates considerable overlap between the signal + noise distribution and the noise only distribution. However, if a stronger signal is added on the signal trials, the amount of sensory input from signal + noise increases and the mean of the distribution of signal + noise, d', shifts, creating less overlap of the distribution. This also occurs when the magnitude of a signal remains the same but a participant in a study is more sensitive to the signal, for instance having better auditory or visual discrimination. When the two distributions of sensory input representing noise and noise + signal trials overlaps, discriminating noise from signal becomes more ambiguous. In these overlapping distributions, *c* represents the response criterion of the individual, which is the level of sensory input at which the subject decides to affirm or deny that a signal is present. Above this level of sensory input, the subject will affirm that there is a signal present, and below this point, the subject will affirm there is only noise present. The value of *c* can be represented as a z-score, but it is independent of d' since the subject can essentially set a response criterion at any point along the distribution. Related to the response criterion *c* is a measure of response bias, which can be represented by β. β is the ratio of the height of the signal + noise curve at *c* relative to the height of the noise curve at *c*. β is the likelihood ratio of the probability that sensory input was due to signal relative to noise at the subject's response criterion point *c*. If β is greater than 1, there is a "No" response bias, and if β is less than 1, there is a "Yes" response bias. If β is equal to 1, which occurs when the subject's response criterion c is at the intersection of the noise and signal + noise curves, there is no bias in the subject's response. Unlike *c*, the value of β can change as a function of the value of d' and so is not independent of d'. For example, as the noise and signal + noise distributions overlap, the value of d' becomes smaller and the value of β increases. In contrast, *c* is theoretically independent of d'.

There are four possible stimulus/response outcomes to any signal detection task. On noise only trials, the participant can correctly say that only noise was present (a correct rejection) or incorrectly say that a signal (signal + noise) was present (a false alarm). On signal + noise trials, the participant can correctly say that a signal was present (a hit) or can incorrectly say that only noise was present (a miss) (Refer to Figure 3). The basic principles of signal detection theory can be applied to yes/no memory recognition paradigms that have similar possible stimulus response outcomes (Hannay, 1986). Overall performance on yes/no recognition tasks can be represented as the ratio of hits to false alarms using a receiver operating characteristic (ROC) procedure (Refer to Figure 4). This curve is generated by plotting the probability of hits on signal + noise trials as a function of the probability of false alarms on noise only trials at all possible levels of response criterion *c*. The separate ROC curves shown in Figure 4 represent how the curves are affected by increasing signal strength or subject sensitivity to the signal. As the signal present in signal + noise trials increases, the distance between the noise and signal + noise distributions increases, which is represented by the increasing values d' takes on when this occurs. In the ROC curve, this improved signal to noise discrimination is represented by increased curvature in the ROC curve when d' is large. So in a yes/no recognition paradigm, if a subject sets his or her criterion response at a certain level *c*, the ROC curve displays the specific probabilistic ratio of hits to false alarms associated with that decision criteria. As the response criterion is made more stringent and becomes smaller, the number of hits and false alarms becomes smaller. As the response criterion becomes more relaxed, the number of hits becomes larger, as well as the number of false alarms.

This method can be used to evaluate the degree to which recognition performance is attributable to familiarity versus recollection. ROC procedures evaluate the effect of a response criterion *c* on hits and false positive rates (Hannay, 1986; Yonelinas, 2002). In this procedure the observed ROC curve is compared to an expected ROC curve. Participants typically rate the confidence in their yes/no recognition responses. A plot of recognition hits against false positive responses associated with different confidence ratings is then created. The equation for hits [Hits $= R + (1-R) F_0$] is defined as the sum of a recollected item (R) and an un-recollected item (1-R) with a familiarity (F_0) that exceeds the participant's response criterion. The equation for false alarms (Fn) is only defined in terms of familiarity and a false alarm will occur if the familiarity of the false alarm item exceeds the response criterion (Fn) (Yonelinas, 2002).

Threshold theory has been found to adequately describe recollection (Yonelinas, 2002). A sensory threshold represents the boundary point below which an individual will never detect a stimulus, and above which, the subject will always detect the stimulus (Hannay, 1986). The threshold is assumed to be a constant value given other subject and environmental factors are held constant; although in reality, these factors are always at work, and so thresholds can vary for an individual. The related phi-gamma hypothesis describes an S-shaped relationship between stimulus magnitude and the probability that the stimulus will be detected (Hannay, 1986). Forced choice recognition procedures represent an application of threshold theory. In a forced-choice recognition paradigm, the participant is presented with the target stimulus as well as one or more alternative choices, and must correctly choose the target stimulus.

The Current State of Memory Research

The task of developing a unitary theory of human memory has been a daunting one. Currently the conceptualization that memory is best represented as having at least two components, a short term and long term component, continues to be supported empirically. However, the characteristics of these two memory components have changed. For example, due to aforementioned weaknesses in the Atkinson-Shiffrin two-store memory model, much of the recent research into characterizing STS has done so using Baddeley's working memory model. The old notion of a rather static STS has been changed to a working memory "workspace" in which information flows freely (not constrained by single direction information transfer of STS to LTM that Atkinson and Shiffrin proposed). An advantage of the working memory model is the neuroanatomical as well as neuropsychological evidence for the existence of working memory subsystems such as the phonological loop and central executive. (Curtis & D'Esposito, 2003; Levy & Goldman-Rakic, 1999; Muller & Knight, 2006; Paulesu, Frith, & Frackowiak, 1993). These subsystems have roles in other cognitive functions such as language acquisition (Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992) and behavioral control (Hofman, Gschwender, Friese, Wiers, & Schmitt, 2008) and studies have shown that the consequences of the operation (or lack of operation) of the subsystems can be predicted with Baddeley's model. For example, there is evidence that children with dyslexia develop their reading disability from a faulty phonological loop system (Baddeley, Gathercole, & Papagno, 1998; Swanson, 1999) and that the size of a child's phonological loop, operationalized as the ability for children to repeat nonwords of various lengths and complexity, may be a better predictor of the number of words learned in the first year of school than IQ (Gathercole & Baddeley, 1989). Other studies have demonstrated the role the phonological loop has in language acquisition through significant correlations between phonological loop capacity and vocabulary knowledge (Gathercole & Baddeley, 1990; Gathercole, Willis, Emslie, & Baddeley, 1992; Michas & Henry, 1994).

Craik and Lockhart's levels of processing framework for viewing how information is maintained in the short term is not necessarily mutually exclusive with Baddeley's working memory model or the existence of a memory system with two primary components. In fact, the levels of processing framework is similar to Atkinson and Shiffrin's control processes or Baddeley's central executive in that the various levels of processing (rehearsal, elaboration, etc.) could reflect intentional strategies a user may employ to help execute a learning task. The basic premise for the levels of processing framework is that increasing depth and/or elaboration of information can lead to increased recall. More recent memory models such as Baddeley's working memory model place a heavy emphasis on the active role of the learner (i.e., subject) in memory studies play in the execution of a learning task. Such active processes require the learner to organize to be learned information in such a way that it makes retrieving this information easier during recall trials.

Structure and Organization of Learning

Miller (1956) originally proposed that there are limits in our ability to process information, and that the process of chunking or grouping individual items into larger units facilitates the storage of more items. These larger units, also refered to as higher-order units are grouped subjectively by the individual within a free recall learning paradigm. Subjective organization refers generally to this organization found in an individual's output on learning trials. Subjective organization has been operationally defined as two or more items recalled together to form a subjective unit (Tulving, 1966),or more generally, as the amount of consistency in recall across trials (Sternberg & Tulving, 1977). Associations are created between items in these higher order units such that the retrieval of some items is associated with the retrieval of other items (Buschke, 1973). The units formed during learning trials are often retained during recall (Buschke, 1976; Tulving, 1962; Tulving, 1966). More importantly, the presence of higher order units is associated with increased recall of list items (Buschke, 1976; Tulving, 1962; Tulving, 1966). Repetition has been previously discussed as an important control process that helps keep information 'in line' so that it can be recalled later. However, repetition is only effective in increasing recall performance if it leads to the formation of appropriate higher order subjective organization units (Buschke, 1976; Tulving, 1966). Our processing ability is greatly increased when information is grouped or chunked into larger units. In a free recall task, words are presented faster than can be immediately recalled from working memory (or STS). Unit sizes grow larger as more individual items are organized into higher order units on each successive trial. The higher order units are retrieved as a unit and individual words within that unit are recited. In this way, the number of words recalled increases over successive trials (Bushke, 1977; Sternberg & Tulving, 1977).

Measurement and Description of Types of Subjective Organization

Subjective organization is best studied using a free recall paradigm because this experimental procedure allows subjects to freely vary their responses. The following section will review subjective organization strategies that have been identified in free recall tasks.

Priority Effect

The priority effect is the tendency in multi-trial list learning tasks for individuals to prioritize new items or items they missed on previous learning trials. Individuals are more likely to recall these items earlier in recall trials than would be expected by chance (Battig, Allen, & Jensen, 1965). The priority effect is not to be confused with the *primacy effect*, which is the tendency to recall items at the beginning of the list. Battig et al. (1965) conducted three separate experiments to demonstrate this effect. In the first experiment, six different CVC (consonant, vowel-consonant) lists were projected visually to participants who wrote as many items as they could recall after each trial for a total of 10 trials. Items were presented randomly on each presentation. The second experiment was similar to the first except stimuli included a list of 3-letter anagrams of common three-letter English words. In the third experiment, participants were presented visually with 40 common nouns and asked to record words they recalled after each learning trial for a total of six trials. The first three trials contained identical 40 words; however for trials 4-6, half of the words were replaced with new words. By comparing the recall rank of each list item for each trial, experimenters found that items that were not recalled on previous trials tended to be recalled earlier than previously correct items and this was consistent across experiments. Furthermore, as an item was recalled more often across learning trials, its recall rank tended to drop and these items tended to be recalled later.

Other evidence for a priority effect comes from part-whole transfer free-recall experiments. In this free recall paradigm introduced by Endel Tulving, individuals are presented with part of a list and given several trials to learn items on this list prior to being presented with the entire word list. An interesting result of this method is that learning a partial-list prior to learning the whole list can lead to worse recall of the whole list (Tulving, 1966). Tulving (1966) demonstrated this with two experiments that compared the recall of two groups: One group exposed to a half of a list prior to being exposed to the entire list and a second group that was exposed to an irrelevant list during the partial-list practice. When both groups were tasked with learning the entire word list, although the partial-list group had an advantage in early learning trials, this group struggled to learn the whole list and the rate of learning actually decreased across trials, while the irrelevant list group's performance steadily increased across trials. This negative part-whole transfer effect has been demonstrated in other studies (Novinski, 1969). This seemingly paradoxical finding is explained as a function of subjective organization. When exposed to the partial-list, individuals organize these words into higher order subjective units. Learning the whole list forces the original higher order subjective units to be reorganized to accommodate new list items. Learning of the whole list was therefore slowed because the individuals were unable to appropriately modify these units (Tulving, 1966).

Measurement of the priority effect has been conducted by Battig et al. (1965) who calculated a standard recall-rank score to compare the rank of newly recalled items to other items on the list. This score was determined for each item on each trial by taking the difference of its recall rank from the median recall rank. Positive ranks reflected items that were recalled before the median recall item and negative values for recall rank reflected items that were recalled after the median recall item. To account for differences in list length from trial to trial, this score was divided by the standard deviation of the total number of recall ranks on the given trial. Thus the standard recall-rank score (summed for all trials) creates an overall measure of priority in a list learning protocol.

There has been some criticism that priority effects were nothing more than artifacts of the recency effect (tendency to recall items at the end of a list first), especially in light of studies demonstrating that previously unrecalled items tended to be those at the end of a list, and therefore subject to the recency effect. Specifically, they argue that a great proportion of "new" items were also items that appeared at the end of the list, even if list presentation was randomized (Baddeley A. , 1968; Postman & Keppel, 1968;). Battig and Slaybaugh offered an interesting method of distinguishing between recency and priority effects by ensuring previously recalled items were in the first two and last two serial presentations on each presentation (to ensure that "old" items always were positioned in the prime serial positions. They found a recency effect in the first half of learning trial and a priority effect that was not confounded by recency in the second half of the learning trials.

Unfortunately, most clinical applications of free recall list learning tasks such as the California Verbal Learning Test (CVLT), Rey Auditory Verbal Learning Test (RAVLT) and

Hopkins Verbal Learning Test (HVLT) (refer to Table 2 for descriptions of these tests) require list items to be presented in the same order on each learning trial, thereby making it impossible to evaluate for a priority effect because it could be confounded by presentation order. An additional limitation is present in the SRT because the full list is read only on the first trial, while on subsequent trials, only words that were not previously recalled are presented. Recency effect and priority effect are confounded in this situation because previously missed items will always be presented on the subsequent trial, meaning that they will always be recited from STS.

Subjective Organization

Tulving (1962) initially defined subjective organization (SO) as the redundency in an individual's free recall that occurs across successive trials. In his early experiments, Tulving examined the relationships among subjective organization, multiple learning trials and accuracy in recall. In one key experiment, a list of 16 unrelated words was presented in random order to participants for a total of 16 trials. He found that the number of words recalled after each trial increased across learning trials, and the recall of words became increasingly organized across trials. In addition, increases in performance were associated with increases in subjective organization. It is well known that individuals naturally organize randomly presented words and that repeated exposure to these words, increases their natural clustering tendencies. By definition SO is a learner-defined phenomenon that cannot be determined a priori by an experimenter, making measurement of SO complicated. Several different measures of SO have been developed and will be reviewed. Refer to the Appendix for a full description of the formulas from which these measures are derived.

Unidirectional Subjective Organization Measure

This measure of SO was proposed by Tulving (1962) to represent redundant recall orders in a multi-trial recall experiment. Recalled items are represented on a matrix in which rows represent the serial position of each word *x* recalled and columns represent each word *y* recalled in the subsequent position *(x+1* position). Entries within the matrix represent frequencies of instances in which word *x* was followed by word *y*. Tulving's initial experiment applied this measure of SO to a 16-word list learning experiment of unrelated words. Two limitations of this measure have been identified. The first is that this measure of SO only considers pairs of words that are recalled in the exact same order on consecutive trials. For example, if the following represents the recall order for trial 1: A-B, this measure would only consider exact repetitions of this recall order in subsequent trials, which may underestimate SO if individuals recall a unit in reverse order (e.g., B-A). The second limitation is that the measure does not adjust for the possibility that exact repetitions could be due to chance factors rather than an intentional organizational strategy employed by the individual. Chance factors are, of course, unknown.

Intertrial Repetition Measure (ITR)

Bousfield and Bousfield (1966) proposed a measure of SO called the intertrial repetition measure (ITR) that unlike Tulving's measure of SO, accounted for the possibility that the observed SO could be due to chance. An intertrial repetition was defined as two items recalled in sequence on a single trial (trial *t*) that were also recalled in sequence on a subsequent trial (Trial *t* +1). The ITR measure was expressed as the difference of the observed ITR minus the expected ITR score [ITR= O (ITR) – E (ITR)], with the observed ITR representing the number of intertrial repetitions that occur on a pair of trials. Another difference between the ITR measure and Tulving's SO measure is that ITR is always computed for blocks of 2 trials, unlike the SO measure which is computed for the total number of learning trials. For multitrial verbal learning experiments, one will obtain multiple ITR values representing trial pairs. For example, on a six-trial list learning task, one would obtain five values for ITR representing ITR scores for trials 1 and 2, 2 and 3, 3 and 4, 4 and 5, and 5 and 6. However, this measure has the same limitation as Tulving's SO measure in that it only accounts for unidirectional repetitions, or word pairs that occur in the exact same order.

Pair Frequency (PF) or Bidirectional Intertrial Repetition Measure

The bidirectional ITR measure also referred to as pair frequency (PF) accounts for the possibility that word pairs may be recalled in the same order or in the reverse order (Anderson & Watts, 1969). The measure otherwise is similar to the unidirectional ITR score described above in that it is derived from the same general formula: O (ITR) - E(ITR). The Observed ITR score in this formula represents item pairs recalled in adjacent trials in forward or backward orders (Sternberg & Tulving, 1977). For example, if the following pair of words was identified in trial 1: A-B, the PF score would also count the same ordered pair (A-B) or backwards pair (B-A) as an intertrial repetition.

Generalized Intertrial Repetition Measure

Both the unidirectional ITR and PF measures are determined by evaluating repetition of word pairs across trials. However, these measures may underestimate the amount of SO taking place if an individual chooses to organize recall in units greater than two, such as triplet or quadruplet groupings. Pellegrino (1971) developed a more generalized ITR measure that allows for more flexibility to consider higher order units greater than two. As such, this ITR measure considers unidirectional, bidirectional and unordered units. For example, if the following recall order represents items recalled on trial 1: A-B-C, for trial 2 there are six possible recall orders that would be accepted as intertrial repetitions (A-B-C, C-B-A, B-A-C, etc.). In addition to the advantage of evaluating any-sized units with no regard to order, this ITR measure also adjusts for expected or chance values of ITR. One criticism of this formula is that when applied to unidirectional word pairs it tends to provide lower scores than the unidirectional ITR measure (Sternberg & Tulving, 1977).

Generalized Adjusted Ratio of Clustering (ARC')

Pellegrino (1971) modified Roenker, Thompson, & Brown's (1971) original adjusted ratio of clustering measure (ARC), which was used to evaluate the amount of clustering in categorized word lists, and applied it to measure subjective organization in free recall of unrelated lists. It is similar to the previously described generalized ITR measure except that it proposes an upper-bound for the value of ITR, which other measures of ITR do not have. As such, the ARC□ measure ranges from 0 to 1 where zero represents chance subjective organization and perfect subjective organization is set at one. This creates a measure with more uniformity that makes it easier to compare the level of SO between individuals in a single study and make comparisons between different experiments (Pellegrino & Battig, 1974).

Sternberg and Tulving (1977) studied the reliability and validity of all five of the above-described measures in a list learning experiment in which list length (22, 36 and 52-word lists) and list order (presentation order of these lists) were manipulated as independent variables. Individuals were randomly assigned to one of the three list-length conditions. Individuals learned three lists of a given list-length and were given eight learning trials for each list. The authors also devised a group of statistical data which was obtained by matching recall scores from participants to each statistic subject, while randomizing the actual words and order in which the words were recalled. Tulving and Sternberg (1977) found that regardless of list length, subjective organization scores were highest for measures that considered only intertrial repetitions for units of size two. That is, the unidirectional ITR measure, paired frequency (PF) measure and SO measure (for units of size two) had significantly higher scores than measures (SO and ARC) that considered units of size three or four. Alternate-form reliabilities for the different subjective organization measures were calculated using the three lists within each list-length condition. Reliability coefficients were highest for the PF measure (ranging from r = .43 to r = .71), followed by the unidirectional ITR measure (r = .45 to r = .57). Reliability coefficients were lowest for ARC measures (ranging from r = .04 to r = .35) that considered unit sizes of three or four, indicating that measurement of unit sizes larger than two is highly unreliable. In addition, inter-correlations between PF scores and SO scores (for units of size 2) were also high (ranging from r = .71 to r =.89).

To compare the empirical validity of these measures, the authors looked at the correlation between each organization measure and total list recall. Correlations between PF scores and total list recall (r = .73 to r =.86) were higher than correlations for unidirectional ITR scores and total list recall (r = .56-78) and much higher relative to correlations for ARC or SO scores and total list recall, which ranged from r = -.05 to r =.66. These results suggest that paired frequency (PF) is the best measure of subjective organization due to its consistently high reliability regardless of various task characteristics (i.e., different list lengths, alternate lists) and high reliability with other measures of subjecitve organization. Though ARC may account for greater unit sizes, it appears there is too much variability in these larger unit sizes for this measure to be reliable. Furthermore, ARC□ scores for unit sizes of two have variable reliabilities that appear to decrease as list size increases. Though Pelligrino (1971) argues that an upper bound is necessary for any measure of subjective organization in order to easily compare different unit sizes, Sternberg and Tulving (1977) demonstrate that additional corrections in the form of adding a division term that represents the maximum value of subjective organization (1/M (ITR)) creates an artificially negative correlation between ratio measures of subjective organization and recall. Sternberg and Tulving (1977) argue that a single correction in the form of a difference score that corrects for the possibility an observed word pair is due to chance is a sufficient correction. Each subjective organization measure has its inherent weaknesses; however, the PF measure appears to be the best option due to its correction for chance and its high reliability and validity.

Demographic Factors Affecting the Use of Subjective Organization

Several demographic factors including age, education and sex have been shown to influence performance on verbal memory tasks. Though the influence of these variables on the use of subjective organization in verbal list learning tasks has garnered less attention, studies generally suggest that children and older adults utilize subjective organization less than young adults (Davis, Klebe, Guinther, Cornwell, & James, 2013; Guttentag, 1985; Witte, Freund, & Sebby, 1990). A curvilinear relationship between age and subjective organization has been described in which subjective organization (as measured with pair frequency scores) increases during childhood through adolescence, then remains relatively stable in young to middle adulthood before decreasing in later life (Davis, Klebe, Guinther, Cornwell, & James, 2013). Davis et al. (2013) evaluated the use of subjective organization on the RAVLT and found that 5-

9 years olds and individuals in their 80s had lower average pair frequency scores relative to individuals 15-59 years of age. They also found that individuals 10-14 years old and individuals in their 70s had lower average pair frequency scores relative to adults 20-49 years of age. Variations in this relationship have been observed such as significantly higher subjective organization use in both young and older adults relative to young children (age 5-10 years old) (Laurence, 1966).

Closely related to the relationship between age and use of subjective organization in verbal memory is the influence that education may have on strategy use in verbal memory tasks. This is particularly true for children as their educational level is often constrained by their age. Therefore, studies evaluating the relationship between educational level, memory performance and memory organization have been cross-cultural because the confound between age and education is often less problematic in other cultures. These studies have generally found that children with at least some formal education use more mnemonic stragies and other strategies such as rehearsal relative to their same-age, non-schooled peers (Ceci, 1991; Sharp, Cole, & Lave, 1979). This has been found to be true even after controlling for other family background differences such as parental schooling, language fluency, wealth and religion (Rogoff, 1981). Among adults with varying levels of education, greater educational attainment (seven or more years of formal education) has been associated with the use of a semantic clustering strategy (Sharp, Cole, & Lave, 1979). None of these studies evaluated subjective organization specifically and instead used lists of related words to see if participants would spontaneously cluster semantically related words as a function of their educational background. It is apparent that exposure to formal schooling is related to the use of specific strategies that aid in recall.

However, this may be more important at lower levels of education (having at least elementary-level education) rather than higher levels (college level vs. high school).

It is well known that women tend to outperform men on verbal memory tasks (Herlitz, Nilsson, & Backman, 1997; Kramer, Delis, & Daniel, 1988; Krueger & Salthouse, 2010). Sex differences in strategy use have been demonstrated on list learning tasks such as the CVLT where women tend to use semantic clustering strategies in contrast to men, who tend to use a serial clustering strategy (recalling words in the same order presented to them) (Delis, Kramer, Kaplan, & Ober, 2000). Though subjective organization was not evaluated directly, a related construct measured on the CVLT, consistency across trials, was evaluated and no sex differences were found (Delis, Kramer, Kaplan, & Ober, 2000). Consistency across trials refers to the percentage of list items that are recalled on adjacent trials. In contrast, subjective organization measures take into account not only the number of list items repeated on adjacent trials, but the order in which these words are recalled.

Memory Deficit after Traumatic Brain Injury

Memory impairment after a traumatic brain injury (TBI) is a common deficit (Corrigan, Whiteneck, & Mellick, 2004; Levin, 1990) and a common complaint of individuals with TBI and their families (Corrigan, Whiteneck, & Mellick, 2004). Focal brain lesions such as contusions and intracranial hematomas can occur in the frontal and temporal lobes as a result of a TBI due to the fact that these cortical areas rest on bony protrusions on the base of the skull, and such lesions have been implicated in post-injury memory impairments (Levin, 1990). Furthermore, diffuse axonal injury (DAI), another type of primary injury observed in TBI has been associated with cognitive impairments specifically in psychomotor speed, attention, information processing and verbal memory (Lezak, Howieson, & Loring, 2004, pp. 168-169; Scheid, Walther, Guthke, Preul, & Yves, 2006). Unlike those structures involved in amnestic syndromes or Alzheimer's dementia, which generally include the hippocampus and adjacent temporal lobe areas (Lezak, Howieson, & Loring, 2004, p. 51), lesions to the prefrontal lobes that are common to TBI have been implicated in different aspects of memory impairment such as impairments in free recall (which could signify a problem with memory retrieval), strategy use in learning, working memory, memory for source, order and other contextual aspects of memory (Lezak, Howieson, & Loring, 2004, pp. 178-180; Savage, et al., 2001; Stuss, et al., 1994). In contrast, other aspects of memory such as recognition memory appear relatively preserved (Ricker et al., 2001), although sometimes recognition memory can also be impaired after severe TBI (Hannay, Levin, & Grossman, 1979).

Individuals with TBI differ in their approach to learning in verbal memory tasks. Specifically, individuals with TBI may not spontaneously organize information into meaningful categories to aid in learning and retrieval of verbal material (Vakil, 2005). However, individuals with TBI do experience improvements in memory with the aid of semantic organization if this organization is provided for them (Goldstein, Levin, Boake, & Lohrey, 1990; Levin & Goldstein, 1986). Levin and Goldstein (1986) presented an experimental list of words to a group of adults with severe TBI and a matched control group under three different list learning conditions: Unrelated words, related words presented in an unclustered format and related words presented in a clustered format. They found that the TBI group demonstrated impaired recall when presented with related words that were unclustered as compared to healthy-controls who were able to organize their recall by clustering the items by category on their own. Furthermore, the recall of the TBI group improved when related words were presented in a clustered format. They evaluated subjective organization on unrelated and related, but unclustered lists using the pair frequency (PF) measure. Whereas controls demonstrated more organized recall as evidenced by higher PF scores, the TBI group's subjective organization scores fell below chance expectation. Together, this difference in the spontaneous implementation of a learning strategy suggests that patients with TBI have difficulty applying an effortful, active learning strategy on their own and rely on passive learning strategies (Levin & Goldstein, 1986).

Other studies using PF as a measure of subjective organization in TBI samples have found that while patients with TBI may show increases in organization across learning trials (as would be expected), they demonstrate less overall organization in their learning on list learning tasks in comparison to normal controls (Brooks, 1975; Levin, Grossman, Rose, & Teasdale, 1979; O'Donnell, Radtke, Leicht, & Caesar, 1988). This is especially evident in patients with frontal lesions (Heubrock, 1999). The inability of the frontal lesion group to improve in their organization across learning trials supports the notion that this region of the brain is implicated in organized encoding/retrieval.

Recall consistency, which is the percentage of common words recalled on two consecutive trials, is a measure commonly calculated on the CVLT/CVLT-II. Higher recall consistency across trials has been associated with less severe TBI (Pluth, Hannay, Massman, & Contant, 2003). The SRT was administered to a mixed group of patients with complicated-mild, moderate and severe TBI at three and six months post injury. When the order in which words are recalled is recorded on the SRT, then several process measures that are available on the CVLT can be determined. Some of these process measures related specifically to organization in recall including primacy/recency effects and recall consistency. Consistency across trials was calculated in the same way as it is on the CVLT and is the sum of the number of words recalled on one trial that were also recalled on the subsequent trial divided by the number of words recalled on all trials, which is converted into a percentage. Patients with mild TBI were more consistent in their recall across trials, followed by moderate TBI and severe TBI. Consistency scores were also related to global outcome at three months-post injury (as measured with Disability Rating Scale scores) and more severe injury was associated with greater levels of disability. Interestingly, no significant primacy or recency effects were observed in any group. Regarding overall performance on the SRT, mild TBI was associated with the highest scores for LTS, LTR and CLTR, followed by moderate and severe TBI, respectively. Moderate and severe TBI was associated with significantly lower Delayed Recall scores on the SRT relative to mild TBI (Pluth, Hannay, Massman, & Contant, 2003). It is important to note that recall consistency is different from the consistent long term retrieval (CLTR) measure calculated from the SRT. Consistent long term retrieval represents the number of items in long term storage that are recalled consistently without any further presentation of these items. Recall consistency can also be differentiated from the pair frequency measure of subjective organization because pair frequency considers not only the consistency in the words that are recalled in adjacent trials, but the order (in either the forward or backward direction) in which the words are recalled. In addition, pair frequency adjusts for chance-expected recall order. In sum, while recall consistency evaluates what words are being recalled across trials, pair frequency evaluates what higher order units are being created by the subject.

Another study conducted by Goldstein, Levin, Boake and Lohrey (1990) evaluated whether individuals with severe closed-head injury (CHI) would benefit more from semantically-cued words on a list learning task compared to physically-cued (i.e., letters) or acoustically-cued (i.e., rhymes) words. Their results were consistent with what would be expected based on the levels of processing theory of memory encoding. That is, semantic processing increased recognition and cued recall for CHI patients compared to recall cues that tap into other more superficial processing levels (i.e., physical or acoustic cues). Other studies have also found that individuals with TBI tend to use fewer semantic clustering strategies in memory tasks (Crosson, Novack, Trenerry, & Craig, 1988; Geary, Kraus, Rubin, Pliskin, & Little, 2011; Stallings, Boake, & Sherer, 1995).

If one major memory deficit for individuals with TBI is disorganized encoding and retrieval, then TBI associated deficits in the frontal-executive system may underlie these abilities. As discussed previously, control processes are methods such as rehearsal, specific search strategies and coding procedures that are employed at will by an individual and used to facilitate learning in different situations. Functioning of the prefrontal lobes is important to one's spontaneous initiation of strategy in novel or unstructured situations.

Different regions of the prefrontal lobes have been associated with different aspects of executive control in both encoding and retrieval (Buckner, 2003; Fletcher et al., 1995). For example, imaging studies have shown that the left inferior prefrontal cortex supports semantic processing while the dorsolateral prefrontal cortex (DLPFC) supports higher-level control processes such as reorganizing information within working memory (Savage et al., 2001). In one study, eight healthy adults underwent PET scanning during a list learning task while regional cerebral blood flow (rCBF) changes in response to their learning and recall of the list items was monitored throughout the task (Savage et al., 2001). Participants were in one of three conditions: A spontaneous condition in which list items were related in four semantic categories that were randomly presented, a directed condition in which participants were explicitly informed about the categorical grouping of the words and instructed to regroup related items to help them improve their memory and an unrelated list in which words were semantically unrelated. Changes in rCBF in the left prefrontal cortex were evaluated across the three encoding conditions. The left DLPFC and left inferior prefrontal cortex (IPFC) showed increased activation relative to baseline in each condition. Furthermore, IPFC activation was higher in the directed condition relative to the unrelated condition. Increased activation in the DLPFC was associated with increased structure in encoding (directed> spontaneous>unrelated conditions) (Savage et al., 2001). This finding was relatively consistent with other studies that show that increased activation in the DLPFC is associated with tasks that require an individual to monitor, manipulate or update items in working memory (Smtih & Jonides, 1999). Individuals with prefrontal lesions are less likely to use organizational techniques including semantic clustering (Levin, 1990) and subjective organization compared to healthy controls (Stuss et al., 1994). One study found that reduced subjective organization (as measured using the paired frequency method) was associated with unilateral left frontal, unilateral right frontal and bilateral frontal lesions (of varying etiologies including infarction, tumor resection, hemorrhage and trauma) on an experimental list learning task (Stuss et al., 1994). When external support was given in the form of a blocked and categorized list, the unilateral left and bilateral frontal lesion groups did not benefit from this structure. Alexander, Stuss and Fansabedian (2003) evaluated performance on the CVLT for patients with focal frontal lesions (N=33, etiology was an acute event including infarction, resection for benign tumor, contusion or hemorrhage), non-frontal lesions (N=11) and healthy controls (N=12). Frontal lesion patients were divided into subgroups based on lesion site and the study authors found that the frontal lesion group overall had significantly impaired subjective organization (as measured using PF scores) relative to the control group and non-frontal lesion group. Specifically, lesions to the posterior left dorsolateral frontal area and posterior medial frontal region were associated with lower subjective organization scores relative to the control group (Alexander, Stuss, & Fansabedian, 2003). As a result, the frontal lesion group had impaired total learning across the five learning trials primarily due to poor use of semantic organization and subjective organization strategy. Other studies have also found that lesions to the DLPFC are associated with organization deficits in learning and memory for lists of words and pictures (Gershberg & Shimanura, 1991) using the paired frequency measure.

Similar findings for reduced strategy use have been observed in mild TBI (mTBI) samples. In one study, strategy use on the CVLT-II was compared between a healthy control group (N=15 females) and a mTBI group (N=19 females). Mild TBI was defined by having a period of loss of consciousness (LOC) of less than 30 minutes, post-traumatic amnesia (PTA) duration of less than 24 hours, and /or Glasgow Coma Scale (GCS) score $\geq$ 13. The authors found that relative to the control group, the mTBI group used less semantic clustering (Geary, Kraus, Rubin, Pliskin, & Little, 2011). However, subjective clustering (as measured with PF scores) was not a significant predictor of learning rate in either the mTBI group or control group. In some ways, this finding is not surprising as subjective organization is typically observed in list learning tasks with extended trials and is often not seen until the latter half of list learning trials. Since the CVLT-II only has five learning trials, it is possible the subjective organization strategy would not have been an appropriate one for such a task.

In summary, of the studies that have evaluated subjective organization of learning in TBI samples (Alexander, Stuss and Fansabedian, 2003; Geary, Kraus, Rubin, Pliskin, & Little, 2011;

Gershberg & Shimanura, 1991; Heubrock, 1999; Levin & Goldstein, 1986; O'Donnell, Radtke, Leicht, & Caesar, 1988; Pluth, Hannay, Massman, & Contant, 2003; Stuss et al., 1994), all but one (Pluth, Hannay, Massman, & Contant, 2003) used PF scores to measure subjective organization. Several studies (Alexander, Stuss and Fansabedian, 2003; Geary, Kraus, Rubin, Pliskin, & Little, 2011; Gershberg & Shimanura, 1991; Heubrock, 1999; Stuss et al., 1994) found that lower PF scores are associated with frontal lesions (usually bilateral or left frontal) and that reduced organization is associated with decreased learning in multitrial list learning tasks. The majority of these studies used the CVLT/CVLT-II, RAVLT or an experimental memory task. None of these studies has used PF to evaluate subjective organization on the SRT in TBI patients within the first year of injury.

In addition to impairments in the use of memory strategies, general impairments in learning across trials, free recall and cued recall have been observed in individuals with TBI. Increases in rCBF in the left angular/supramarginal gyri have been observed in individuals after severe TBI (as compared to increased rCBF to the left middle frontal gyrus, left medial cerebellum/vermis and left angular/supramarginal gyrus of the control group). This difference in activation corresponded to between group differences in learning across trials, retrieval and cued recall (Ricker et al., 2001). The authors posit that the increased cerebral blood flow observed in more posterior brain regions of the TBI group signifies that these individuals may have used a phonological or semantic-based strategy rather than an executive-organizational based recall strategy. Other studies find that when differences in initial learning between TBI patients and control groups are controlled, individuals with TBI have a more rapid rate of forgetting on the CVLT from trial 5 to short-delay free recall, but relatively equivalent rates of forgetting from short delay to long delay recall (Vanderploeg, Crowell, & Curtiss, 2001).

Poor performance on consistent long term retrieval (CLTR) and delayed recall on the 6-trial selective reminding test (SRT) has been associated with indicators of decreased white matter integrity within the perforant pathway using diffusion tensor imaging (DTI) (Christidi et al., 2011).This suggests that though degradation of white matter tracks in the frontal cortex is considered a hallmark injury of TBI, degradation of white matter in other areas such as the perforant pathway which includes connections from the cortex to the entorhinal cortex and dentate gyrus and hippocampus occurs which can adversely affect memory performance.

Regarding recognition performance, individuals with TBI usually demonstrate intact recognition capabilities, however some impairments specifically in the ability to discriminate between target items and false-positive items have been observed. Furthermore, this deficit has been observed specifically with lesions to the left posterior dorsolateral frontal region (Alexander, Stuss, & Fansabedian, 2003). Increases in rCBF in the right and left middle frontal gyri for individuals with TBI and healthy controls during yes/no recognition trials have also been found (Ricker et al., 2001).

Several methodological issues in studies evaluating the relationship between TBI and memory impairment have been described (Vakil, 2005): 1. Selection bias, 2. A lack of matched control groups, and 3. Small sample sizes. Many study samples include only patients who have participated in formal rehabilitation programs, who often do not represent the larger TBI population. Vakil (2005) encourages researchers to use patient-control groups, such as spinal cord injury samples, rather than healthy control groups since patient control groups share a history of significant trauma, possible emotional effects and hospitalization experiences that a TBI sample would likely experience, and a healthy sample would not. However, this may not always be a feasible or practical option. In lieu of a patient-control group, patient samples that have a range of TBI severity levels should be represented in the overall TBI sample, as the injury severity can be used a covariate in statistical analyses. More serious TBI generally is expected to result in poorer performance on memory as well as other tasks (Vakil, 2005). If the pattern of performance of patients with more severe TBI is not consistent with expectations for particular measures of various aspects of memory, it brings into question the appropriateness of the sample, its characteristics including severity of injury, the validity of the measures and even the theorized underlying model of memory. Finally, sample sizes in a recent review of 68 studies on memory and TBI were typically small, with the majority of studies (N=58 studies) having sample sizes below 50 (Vakil, 2005). The large variances and low statistical power resulting from these small sample sizes can limit the interpretation of study findings.

The Selective Reminding Procedure

The Selective Reminding procedure is a unique list learning procedure that requires respondents to learn a list of words across multiple auditory presentation trials (Buschke, 1973). Rather than reciting the entire word list to the learner on each learning trial, the full word list is only read once to the respondent (unless the learner does not recall any words on a trial). On subsequent trials, only words that were not recalled by the respondent on the previous trial are repeated to the respondent. Consequently, respondents are selectively reminded of only words which they failed to recall on each previous learning trial. This procedure is repeated until either the respondent attempts all learning trials, or until the respondent meets the criterion number of reciting the entire list on a criterion number of consecutive trials, which in Buschke's (1973) procedure was two consecutive trials (Buschke, 1973). This basic procedure has been employed with variations by researchers in list length, number of learning trials, number of criterion trials for ending the procedure, rate of presentation of list items, characteristics of the list words (nouns, adjectives, verbs, etc.), the degree of association among the words (unrelated vs. semantically related), and the time between learning and a delayed recall trial as well as the presence of other types of recall formats such as cued recall and recognition recall depending on the specific test that is administered. For example, Bushcke (1973) administered a selective reminding test that included 20 animal names to be learned across 12 trials. In later experiments, Buschke (1976) administered another form of the test by presenting a list of 40 unrelated words individually on 3x5 inch cards. Buschke and Fuld's (1974) selective reminding procedure involved reading aloud a list of 10 animals to participants. Later, modifications were made by Levin (Levin et al., 1982; Peters & Levin, 1977) which included using a list of 12 unrelated words that were spoken aloud to participants, the addition of a visual cued recall trial that followed immediately after the learning trials, followed by a multiple choice recognition trial and then a 30 minute delayed free recall trial.

Generally, the process involved in learning trials and delayed free recall trials of the selective reminding procedure is thought to involve the two memory systems- Short Term Storage (STS) and Long Term Storage (LTS) of the separate stores memory model. Furthermore, the selective reminding procedure provides a method to distinguish between different items recalled from STS and LTS (Buschke & Fuld, 1974). Items recalled after the initial presentation are thought to be stored and retrieved during free recall from LTS. These items continue to be recalled without further presentation, and thus they are thought to be represented in LTS. Using a scored SRT protocol represented in Figure 5, if the word *county,* which is presented initially on trial one, is recalled immediately after that initial presentation, it is not presented on the second trial. In the protocol displayed in Figure 5, *county* is recalled on trial two and it is assumed, because there was no further presentation of the word after the first trial, to have been stored in LTS after the initial trial and retrieved on the second trial from LTS. Though there is a lapse in the subject's recall of *county* on trial seven, this is believed to be a function of faulty retrieval from LTS, rather than a problem with storage since items in LTS are assumed to be encoded and stored sufficiently. Evidence for this is given by using a unique selective reminding format known as a restricted reminding procedure in which an individual learning a word list is reminded of only items not recalled on previous trials until each item is recalled once. The subject is not reminded again of that word even if it is not retrieved on a subsequent trial (Buschke & Fuld, 1974). Using the restricted reminding paradigm, Buschke (1973) found that the participant was able to spontaneously retrieve items without further presentation of the word, even after many consecutive retrieval failures. This indicated the word was encoded in LTS, and available for eventual retrieval. Retrieval from LTS can be characterized as either consistent retrieval in which there are no lapses in recall and indicates list learning, or random retrieval in which recall of list items is inconsistent across trials and indicates item learning (Buschke & Fuld, 1974). Random retrieval from LTS remains relatively constant across trials (Buschke & Fuld, 1974). The apparent learning over trials evident in verbal learning tests is due to the transfer of items from this random item store to a list store that represents consistently recalled items without reminders (Buschke & Fuld, 1974; Fuld & Buschke, 1976), and is a measure of what percentage of the list is truly learned (Buschke, 1973). Furthermore, consistent recall of an item is an indication of either organized encoding or organized retrieval or both. In contrast to LTS, items recalled from STS are those that can only be recalled after direct presentation. Several measures can be calculated from the Selective Reminding Test including Short Term Storage (STS), Long Term Storage (LTS), Short Term Retrieval (STR), Long Term Retrieval (LTR), Consistent Long

Term Retrieval (CLTR), and Random Long Term Retrieval (RLTR) (Buschke, 1973) which are described in Table 1. According to Fuld & Buschke (1976), retrieval can be measured as recall from storage after a single presentation of the items. Items that are able to recalled spontaneously after they are not recalled on one or more trials are still considered to be retrieved from LTS. After an initial presentation, storage can be estimated by totaling the number of different items recalled across trials. Traditional recall procedures in which words are repeated prior to each recall trial make it difficult to determine if the respondent is recalling items just presented (i.e., "dumping" from STS) or if items are being retrieved from LTS, and there is also no way to distinguish already learned list items from items that have not been learned (Buschke, 1973). The advantage of using a selective reminding procedure is that it allows one to determine portions of the list that have truly been learned (that is retrieved from LTS) since the respondent must attempt to recall words without further presentation of items (Buschke, 1973). By limiting the number of presentations given to respondents, the respondent's own subjective organization in learning and retrieval is more apparent since it is being minimally influenced by frequent repetitions of items in a predetermined order (Buschke, 1973).

Beatty, Krull, Wilbanks, Blanco, Hames and Paul (1996) evaluated the validity of these measures by assessing the probability of word recall based on the memory status of that word (i.e., whether the word is considered to be a part of STM vs. RLTR vs. CLTR, etc.) in a group of adults diagnosed with multiple sclerosis and a small control group. Based on the two-store memory model, they predicted words recalled from CLTR should have the highest probability of being recalled after a 30 minute long delay compared to words recalled from RLTR. Words recalled from STM would have the lowest probability of being recalled. Though the probabilities for spontaneous recall of words in LTS theoretically could differ according to whether the word is recalled consistently or randomly, it is assumed that all words that can be recalled without further presentation are stored in LTS. Therefore, in a delayed recognition test, the predicted probability of recognizing words from LTS, including both CLTR and RLTR words, would be close to 100%, while the recognition of words stored in STS would be much lower (Beatty et al., 1996). As predicted, word recall after a 30 minute delay was highest for words consistently retrieved from LTS, followed by randomly retrieved words from LTS and lowest for words retrieved from STS. As predicted, recognition of words after the 30 minute delay was highest for words in LTS, regardless of whether those words were retrieved consistently or randomly, and lowest for words from STS. This provides evidence in support of the predictive validity of these constructs, which are assumed to be measured in the SRT (Beatty et al., 1996).

Normative Studies

Larrabee, Trahan, Curtiss and Levin (1988) conducted a normative study on Form I of Levin et al.'s (1982) version of the Verbal Selective Reminding Test (VSRT) that includes 12 trials of 12 unrelated words, immediately followed by a cued recall, which was then followed by a multiple choice visual recognition trial, that required participants to distinguish between the correct word and three other foil words (phonemic, semantic and unrelated foil presented simultaneously on a 2x4 in. index card). After this, a spontaneous recall trial was administered after a 30 minute delay period. They conducted their normative study on 271 healthy adults ranging from 18 to 91 years of age. To test for age, gender and education effects, a subset of 80 men and 80 women were matched for age and education. Age and gender were both predictors of SRT performance, although age was a stronger predictor (Larrabee et al., 1988). Specifically age contributed as significant predictor above gender for the following measures: Increasing age was generally associated with decreases in scores of total recall, LTR, LTS, CLTS and increases in STR, RLTR, as well as reminders by examiner and intrusions. Incremental r^2 values ranged from .084 for intrusions to .253 for CLTR (Larrabee et al., 1988). Women outperformed men on measures of Total recall, LTS, LTR, CLTR and delayed recall. Men had higher scores on measures of STR, RLTR and Reminders by examiner. R^2 values ranged from .027 for total recall, to .043 for delayed recall. A significant interaction between gender and education was found for total recall, LTR, STR, LTS and needing reminders that appeared to be driven by education effects observed in females only (Larrabee et al., 1988). The authors created separate gender corrected raw scores and norms for comparison were based on 10 age groups that were created.

Another normative study conducted by Ruff, Light and Quayhagen (1989) included 392 healthy, primarily Non-Hispanic White (95% of sample) adults ranging in age from 16-70 years old, with a roughly equivalent male to female ratio. Like Larrabee et al. (1988), Ruff et al. (1989) used Levin’s Form I SRT, with a few modifications. Ruff et al.’s (1982) criterion for test completion was spontaneous recall of all 12 list items on two consecutive trials, as opposed to Larrabee et al.’s (1988) criterion of correct recall of all list items on three consecutive trials. Also, they chose a one hour free recall delay, rather than the 30 minute delay in Larrabee et al. (1988). Ruff et al. (1989) divided their sample into 15 year age increments to test the effect of age on performance, while Larrabee et al. (1988) used 10 year age increments. They found gender differences in test completion, and women were more likely to have successfully recalled all 12 words compared to men (In the entire sample, 56% met test completion criteria). Furthermore, at all percentile levels, women overall had superior recall compared to men on each learning trial. In addition, women had significantly higher total recall and CLTR scores, although the authors caution against interpreting this too closely because the overall effect size of association between gender and these scores was small (ω^2 =.04). Test-retest reliability was assessed six months after the first administration of the SRT using Levin et al.'s (1982) alternate form (Form II) of the test. Measures of total recall and total CLTR were highly correlated between Form I and II of the SRT (r =.73 and r =.66 respectively), and again, women were more likely to have successfully recalled all 12 words (Ruff et al., 1989). One hour delayed recall scores indicate those who were able to finish the SRT recalled on average about 10 words, while those who did not finish the SRT recalled on average about 7 words after the delay. There were no gender differences in delayed recall. Only minor education effects were observed and participants with 13-15 years of education performed slightly better than those with educational levels above or below this range (Ruff, Light, & Quayhagen, 1989). Furthermore, minimal age effects were observed, but the pattern followed that of Larrabee et al. (1988) and increasing age was associated with lower scores for both LTS, CLTR. Successful performance on the SRT was also not significantly influenced by verbal intelligence or auditory attention (Ruff et al., 1989). Overall, the authors concluded that SRT performance results should be compared to the combined normative data of both finishers and non-finishers that are gender stratified (Ruff et al., 1989). Despite differences in procedure and demographic stratification, both Ruff et al. (1989) and Larrabee et al. (1988) found few education effects. Studies differed in the relative contribution of age and gender to SRT performance. Ruff et al. (1989) seemed to find that gender effects were more prominent than age effects, while Larrabee et al. (1988) seemed to find the opposite was true. This difference may be accounted for by demographic differences in their samples. Ruff et al. (1989) included more equivalent male and female ratios for each age group, and had at least 15 males and 15 females represented in each of their four age groups. The Larrabee et al. (1988) sample had a more disproportionate gender distribution for their seven age groups. Specifically, the ratios of men to women in the 50-59 age group (2:22) and 80-91 age group (4:23) were especially small.

Interestingly, in Ruff et al.'s (1989) normative study for the SRT, many participants used clustering strategies to recall words. Furthermore, those who were able to finish the test used organizational strategies based on temporal association between words (ex. Primacy and recency) or semantic associations between words (ex. agriculture-related word clusters). Those who were unable to finish the test by the twelfth trial utilized fewer meaningful clusters (Ruff et al., 1989).

Alternate Forms of the SRT

Four alternate forms of the SRT have been created, each matched in terms of list length, word length and word frequency (Hannay & Levin, 1985). They tested the validity of the assumed equivalence of these four alternate forms by administering the four forms to participants in one week intervals using a Latin square design which permitted the determination of learning to learn effects and order effects as well as relative difficulty of the test forms. They found that Form 1 was significantly more difficult than the remaining three forms of the test, as indicated by significantly lower total recall, LTS, LTR, CLTR scores and higher STR scores compared to other Forms 2-4 (Hannay & Levin, 1985). They also found a significant test administration effect such that performance (specifically in terms of total recall, LTR, CLTR, RLTR) on the initial SRT administration was worse compared to performance on the second through fourth administrations. Specific test order effects were only found for cued recall. SRT presentation order of Form 4, 3, 2, then 1 resulted in fewer correct cued-recall responses than the remaining three test presentation orders (1234, 2341, and 3412). Test-retest reliability ranged from r =.484 - .654, which was significant, but considered low by the authors when compared to accepted reliability coefficients of other psychological tests.

A shorter, six-trial version of the SRT (Larrabee, Trahan, & Levin, 2000) allows for the estimation of the same measures of short and long term storage and retrieval with the advantage of a shortened administration time, which is important for clinical applications of the test. Larrabee, Trahan and Levin (2000) used normative data for the 12-trial Form I version of the SRT. These original protocols were rescored as if the learning trials ended after six trials instead of 12 trials. Measures of LTS, CLTR, RLTR, LTR, STS and STR were calculated based on recall after only 6 trials and compared to these same measures calculated from the 12 trial SRT. Since delayed recall data are thus only available for the 12 trial version of the SRT, delayed recall differences between the 6-trial and 12-trial versions could not directly be compared. However, recall after the first 6 trials predicted over half of the variance in delayed recall (Larrabee, Trahan, & Levin, 2000). As expected, high correlations (ranging from r =.916-.958) were observed between the 6 and 12-trial versions of the SRT for measures of total recall, LTS, LTR, and CLTR. Table 2 presents the details related to the most commonly used verbal learning and memory tests.

Present Problem

The current study evaluated the use of subjective organization on the SRT in individuals with complicated-mild, moderate and severe TBI at three and six months post injury. Pair frequency (PF) scores were used to measure subjective organization given the evidence of its high reliability and validity. The association between injury severity (Best Day 1 Glasgow Coma Scale Score and CT scan classification) and PF scores at each time point will also be evaluated, with past research suggesting that more severe TBI is associated with lower subjective organization scores. Prior research suggests that increased organization in learning corresponds to increased memory recall and retrieval. Therefore, higher PF scores, which would indicate more organized learning, were expected to be associated with higher total recall, LTS and 30 minute delayed recall scores on the SRT at each time point. Furthermore, because consistent recall of an item is an indication of either organized encoding, organized retrieval or both, higher PF scores should be associated with greater CLTR scores on the SRT.

Hypotheses

Hypothesis 1: Severity of injury will be related to PF scores when demographic variables are taken into account with a higher severity of injury being related to lower PF scores at each of three and six months post injury.

Hypothesis 2: In a subgroup of patients who completed the SRT at both three and six months post injury, these patients will obtain significantly higher PF scores at six months post injury relative to PF scores at three months post injury when severity of injury and demographic variables are taken into account.

Hypothesis 3: There will be a significant positive correlation between PF scores and performance on the SRT (higher scores on total recall, LTS, CLTR and 30 minute delayed recall) at three and six months post injury.

Method

Participants

Data were obtained retrospectively from a database of 160 patients with complicated mild, moderate or severe TBI consecutively admitted to the Neurosurgery Intensive Care Unit of Ben Taub General Hospital, Harris County Hospital District, a level one Trauma Center in Houston, TX. Patients were administered a neuropsychological battery at 3, 6, 12 and/or 24 months post injury. All patients were enrolled in NIH-NINDS funded grants during the years 1987-2004. In the current investigation, patients were excluded if they were administered the Spanish-version of the Selective Reminding Test (n= 30 at three months and n = 17 at six months). Patients were also excluded (n = 6 at three months; n = 4 for six months) if test protocols were incomplete or if the patient had difficulty with the test precluding the test from being administered correctly (e.g., perseveration, not understanding instructions, etc.). Due to the comparatively small number of Asian patients in the sample, these patients (n =3) were excluded from analyses and ethnic group comparisons on outcome measures included Non-Hispanic Whites, Hispanics and Blacks. The remaining sample consisted of 121 patients tested at three months post injury and 87 patients tested at six months post injury. The research protocol for grants under which patients were studied included a two week window around the 90 day (3 month) and 180 day (6 month) post injury time points. Approximately 79% of the three month sample was tested within this two-week window. Ninety-five percent of the three month sample was tested within 30 days before or after the three month post injury time point and 99% were tested within 45 days of the three month post injury time point (one patient was tested 48 days after the three month post injury date). Approximately 74% of the six month sample was tested within a two week window of 180 days post injury. Ninety percent of the six month sample was tested within 30 days before or after the six month post injury time point and 95% of the sample was tested within 45 days of the six month post injury time point. The demographic breakdown of the three and six month post injury samples are displayed in Table 3. The three month sample was comprised of approximately 83 % men and 17% women and the six month sample was comprised of approximately 81% men and 19% women. These ratios are consistent with male-female ratios in other studies with TBI samples (McCauley, Hannay, & Swank, 2001; Robertson et al., 1999). The ethnic makeup of the sample at three months post injury was 44% Non-Hispanic White, 23% Hispanic and 33% Black while the ethnic makeup of the sample at six months post injury was approximately 47% Non-Hispanic White, 23% Hispanic and 30% Black. Patients at both time points had an average educational level of approximately 11 years. A one-way ANOVA revealed educational differences among the different ethnic groups at three months, $F(2,117) = 9.034$, $p < .001$ and six months post injury, $F(2,83) = 8.067$, $p = .001$. The average educational level for Non-Hispanic Whites was 12.34 years, compared to Hispanics, (average = 10.14 years) and Blacks (average = 10.79 years) at three months post injury. At six months post injury, Non-Hispanic Whites had on average 12.56 years of education, Hispanics had 10.40 years of education and Blacks had 10.68 years of education. Post-hoc testing with Bonferroni corrections for multiple comparisons revealed that Non-Hispanic Whites had a higher educational level compared to both Hispanics and Blacks at three and six months post injury. The most common mechanism of injury at both time points was the motor vehicle accident followed by assaults and pedestrian vs. motor vehicle accidents. Patients ranged in age from 15 to 78 years of age with an average age of 31 years old. Best

Day 1 GCS scores ranged from 3 to 15 at three and six months post injury, with approximately 40% of patients at three months post injury having scores in the severe range (GCS 3-8), 31% had GCS scores in the moderate range (GCS 9-12) and 27% had scores in the mild range (GCS 13-15). The percentage of patients with severe TBI (GCS 3-8) (approximately 51%) was slightly higher at six months post injury relative to three months. The Galveston Orientation and Amnesia Test (GOAT), which assesses length of post traumatic amnesia (PTA), was administered to all patients at each time point. The average GOAT score for three and six months post injury was approximately 85 and 87 respectively. At three months post injury, 15.8 % of the sample (n=19) was not fully oriented (as indicated by a GOAT score less than 75). Furthermore, 15% of the sample (n =18) had a maximum GOAT score of 100. At six months post injury, only 8.1% of the sample (n =7) was not fully oriented (as indicated by a GOAT score less than 75). Furthermore, 16.3% of the sample (n=14) had a maximum GOAT score of 100. One way ANOVA and Chi-square tests revealed no significant differences on demographic factors at three and six months post injury.

A subset of individuals who completed the SRT at both three and six months (N= 73) were included in a separate analysis evaluating the second hypothesis. Table 5 displays the demographic makeup of this group. This subsample was similar in age, education and injury severity to the larger samples tested at each time point. The subsample had an average three month post injury GOAT score of 83.69 (SD = 17.96) and an average six month GOAT score of 86.79 (SD = 86.79). The approximate three point increase in GOAT scores from three to six months post injury was significant, $t(71) = 2.806$, $p = .006$. In addition, mechanism of injury mirrored that of the larger sample with MVA, assaults and pedestrian vs. MVA representing the most common sources of injuries.

Measures of Injury Severity

The **Best Day 1 Glasgow Coma Scale (GCS)** (Teasdale & Jennett, 1974) was used as one measure of brain injury severity. The GCS measures three components of consciousness: Eye opening, best verbal response and best motor response (Jennett & Teasdale, 1981). Scores for the three components are summed to yield a single total score. Total scores on the GCS can range from 3, indicating no eye, verbal or motor response, to 15 indicating full consciousness in patients who are alive. Severe head injuries may be roughly associated with GCS scores of 3-8. Moderate injuries are roughly associated with GCS scores of 9-12 and mild injuries with GCS scores of 13-15 (Fischer, Hannay, Loring, & Lezak, 2004). The Best Day 1 GCS, which is the highest GCS score obtained on a patient within the first 24 hours after admission to the hospital was used as a measure of injury severity.

The **Worst CT Scan** (using the Marshall CT Classification system**)** (Marshall, et al., 1991) was used as a measure of injury severity. The worst CT scan represents the scan showing the largest amount of intracranial pathology (hematoma thickness, midline shift, etc.) among serial CT examinations and has been shown to be related to injury severity and outcome (Servadei et al., 2000). The Marshall Classification system classifies TBI patients into one of six classes according to the severity and type of abnormality seen on a CT brain scan. The guiding characteristics of the system are 1) the presence or absence of mass lesions, 2) the presence or absence of intracranial abnormalities, 3) CT signs of raised intracranial pressure (status of basal cisterns, shift), and 4) planned evacuation of mass lesions (Chawda, Hildebrand, Pape, & Giannoudis, 2004). The categories for the system are Diffuse Injury I (D1), Diffuse Injury II

(D2), Diffuse Injury III (D3), Diffuse Injury IV (D4), evacuated mass lesion (M1) and non-evacuated mass lesion (M2) (Marshall, et al., 1991). Previous studies have formed the following four groupings to enter Marshall CT scan into prediction models: Diffuse I and II, Diffuse III and IV, Evacuated Mass lesion, Non-evacuated mass lesion (Contant, Valadka, Gopinath, Hannay, & Robertson, 2001). Alternatively, groupings that combine evacuated and non-evacuated mass lesions into one category have also been entered into models (Robertson et al., 1999). In the present study, the second method was used to enter the worst Marshall CT scan and Marshall categories were separated into three groups: Diffuse I/II (D1/D2), Diffuse III/ IV (D3/D4), Evacuated mass lesion/Non-evacuated mass lesion (M1/M2).

Memory Performance Measure

The Selective Reminding Test (Hannay & Levin, 1986 version) was administered to patients. Form 2 of the SRT was used which includes the following l2 list items in order of initial presentation (*shine, disagree, fat, wealthy, drunk, pin, grass, moon, prepare, prize, duck, leaf*). The SRT was administered according to standardized procedure which includes reading the entire 12-word list to the patient on the first trial and recording the order of recall responses. On each subsequent trial, only words that were not recalled by the patient on the previous trial are read aloud to the patient, hence they are selectively reminded of only list items that were omitted from their response on the previous trial. The order in which they recall list items is recorded for each trial. This procedure is continued until the patient completes 12 learning trials or until the patient meets a criterion of two consecutive perfectly recalled trials with no reminders. Following the 12 learning trials, a cued-recall trial is given in which the first two letters (or three letters in the case that more than one word shared the first three letters) are presented visually on an index card to the patient as cues. Of note, due to the original construction of the list items (Levin et al., 1982; Peters & Levin, 1977), one list item from Form 1 (bee) was omitted from the cued-recall trial because the word would be identified with the two-letter cue. Therefore, one word from each form *(pin, tin* and *egg* for Forms 2, 3 and 4, respectively) was omitted, leaving a total of eleven cued-recall items. A multiple choice recognition trial immediately follows the cued-recall trial. Four words are presented visually on an index card consisting of a target list item, a synonym, homonym and unrelated word and the patient is asked to select the target list item. A final free-recall trial is given 30 minutes after the multiple choice recognition trials are completed and the patient completes nonverbal tasks in the interim to prevent confounding effects of other verbal tasks. Performance measures obtained from the SRT include total recall, short term storage (STS), long term storage (LTS), short term retrieval (STR), long term retrieval (LTR), consistent long term retrieval (CLTR), random long term retrieval (RLTR), cued recall, multiple choice recognition, delayed total recall and delayed multiple choice recognition scores. A summary of these measures is provided in Table 1. For patients who meet criterion performance prior the trial 12, total recall, LTS, CLTR and LTR are assumed to be 12 and this is imputed into remaining trials because of the assumption that the patient would have continued to recall all list items consistently for the remainder of the test.

Measure of Subjective Organization

Pair frequency (PF) measures (Anderson & Watts, 1969) were used to calculate subjective organization on the SRT. PF is determined as a function of the number of observed item pairs in either forward or reverse order that are recalled on adjacent recall trials (also known as an inter-trial repetition) minus the number of item pairs that would be expected by chance.

Because inter-trial repetitions are calculated according to trial pairs, 11 PF scores are possible for the SRT (e.g., trials 1 and 2, 2 and 3,….11 and 12). The sum of pair frequency scores across all trials completed by the patient is variable for two reasons: 1.) Individuals differ in their use of subjective organization. 2) Standard administration of the SRT allows the examiner to end testing once the patient reaches a criterion performance of two consecutive perfect recall trials without any reminders of words. For example, if a patient reaches criterion performance on trials 7 and 8, their PF scores cannot be determined for the combination of adjacent trial responses for trials 8 and 9, 9 and 10, 10 and 11, 11 and 12 because the individual did not have to recall any more words. In calculating the usual measures of SRT performance (total recall, LTS, LTR, CLTR, etc.), it is assumed that for all trials beyond the criterion, the patient continues to recall all 12 words from the list. In order to calculate a PF total score for individuals who meet criterion performance prior to the last trial (trial 12), a similar assumption is warranted. In this case, the PF score for the last trial combination in reaching criterion would be given to all subsequent trial combinations for which actual responses did not have to be provided. Given the scenario described above in which an individual reaches criterion performance on trial 8, the PF score for adjacent trials 7-8 would be used for adjacent trial combinations 8-9, 9-10, 10-11 and 11-12. As with the traditional SRT measures, there is no way to determine if this measure over or under-estimates the pair frequency total, had the individual made responses on all 12 trials, however it makes an analogous assumption. This imputation was only applied to individuals who completed the SRT to criterion (n= 11 at three months post injury and n= 18 at six months post injury) prior to the 12th trial.

A second pair frequency (PF) score can be calculated that will represent the total pair frequency scores for the first five recall trials of the SRT. All patients were able to complete at least the first five trials of the SRT before meeting criterion performance. This also allows for comparison of pair frequency scores with other list learning tests that have five learning trials such as the CVLT or RAVLT (although differences in list length may prevent direct comparisons of PF scores on other list learning tasks). All pair frequency scores were calculated using the Scoring Options for Recall Tests, Version 2.0 (SORT 2.0) program (Elie & Payne, 1999).

Data Screening Procedures

All variables were evaluated for possible errors and missing cases. Missing data were coded as 99 or -99 where appropriate and were excluded in listwise fashion in all analyses so that patients with valid data from all predictors entered into models were included in analyses. All SRT protocols were reviewed for accuracy and data excluded if there were indications of invalid scores such as confounding between the SRT and other list learning tasks that may have been given by other clinicians near the time point of interest.

Data Analysis

Hypothesis 1: A hierarchical linear regression model was used to test the hypothesis that high severity of injury would be related to lower PF scores at both three and six months post injury. Predictor variables included measures of injury severity (Best Day 1 GCS total scores and Worst CT Scan Marshall classification category) which were entered first into the model followed by demographic variables (age, gender, ethnicity and education). Best Day 1 GCS total scores were entered as continuous variables. Worst CT scan Marshall classification was collapsed into the following groups: Diffuse I/II (D1/D2), Diffuse III/ IV (D2/D4), Evacuated Mass lesion/Non-evacuated lesion (M1/M2). These three categories were then dummy coded into two separate variables (using M1/M2 as the reference category) so that they could be entered into regression models. For demographic variables, age and years of education were entered as continuous variables. Gender and ethnicity were dummy coded with males and Non-Hispanic Whites set as reference groups respectively for each variable and entered into models.

Criterion variables included total pair frequency (PF) scores for all 12 SRT trials as well as total PF scores for the first five SRT trials. Separate models were run for three months post injury and six month post injury follow up points.

Hypothesis 2: In a subset of patients who completed the SRT at both three and six months post injury, a repeated measures Analysis of Covariance (ANCOVA) was used to test the hypothesis that patients would obtain significantly higher PF scores at six months post injury relative to scores at three months post injury when severity of injury (Best Day 1 GCS scores) and demographic variables (age and gender) were taken into account. Total PF scores for 12 trials and total PF scores for the first 5 trials at three and six months post injury were entered as repeated within-subjects factors. Severity of injury (Best Day 1 GCS score), age and gender were entered as covariates.

Hypothesis 3: Bivariate correlations between the pair frequency scores (total PF scores for 12 trials and total PF scores for the first 5 trials) and measures of SRT performance (total recall, LTS, CLTR, cued recall, multiple choice recognition, 30 minute delayed free recall, 30 minute delayed multiple choice recognition) were evaluated to test the hypothesis that PF scores would be positively associated with performance on the SRT. Correlations were evaluated separately for each time point.

Results

Description of SRT Performance

Performance measures on the SRT for the three and six month post injury samples are displayed in Table 4. A one way ANOVA revealed there were no significant differences between patients tested at three months vs. six months on STR, total intrusions, multiple choice recognition, 30 minute recall, 30 minute multiple choice recognition and pair frequency totals for all 12 SRT trials. Patients tested at six months scored 12 points higher on total recall compared to those tested at three months post injury. Patients tested at six months post injury also had higher LTS and CLTR scores relative to those tested at three months post injury. Patients tested at six months post injury recalled approximately two more list items on cued recall relative to patients tested at three months post injury. Patients at six months post injury had approximately one more inter-trial repetition (indicated by five trial PF scores) across the first five SRT trials relative to patients tested at three months post injury. Though Non-Hispanics Whites had a higher education level relative to Hispanics and Blacks, there was no significant correlation between years of education and pair frequency total scores at three months post injury $r(120) = .068$, $p = .46$ or six months post injury, $r(86) = .100$, $p = .36$. Likewise, there was no significant correlation between years of education and five trial pair frequency scores at three months post injury $r(120) = .034$, $p = .72$ and a small, but significant correlation at six months post injury was observed, $r(86) = .22$, $p = .039$.

Performance indicators on the SRT for the subsample tested at both three and six months post injury are displayed in Table 6. Dependent sample *t*-tests revealed that the subsample overall performed significantly better at six months post injury relative to three months on several SRT performance measures including total recall, LTS, CLTR, STR, cued recall, delayed recall, immediate and delayed recognition, although these differences were small. Fewer intrusions were observed at six months relative to three months post injury. Furthermore, five trial PF scores were higher at six months relative to three months in the sample. In contrast, no differences in 12 trial PF scores were observed between the two time points.

Prediction of 12 Trial Pair Frequency Scores

Preliminary analyses of the distribution of total PF scores for all 12 trials showed that the data was positively skewed for PF scores at both three and six post injury (displayed in Figure 6). The distribution of PF total scores at three months post injury was non-normal with skewness of 1.79 (SE = .22) and kurtosis of 4.01 (SE = .44). The distribution of PF scores at six months post injury was also non-normal with skewness of 1.37 (SE = .26) and kurtosis of 1.76 (SE. = .51). Though the average PF total score was 8.26 and 9.59 at three and six months respectively, scores as a whole tended to be very low. Negative PF scores indicate that the observed number of inter-trial repetitions was smaller than the expected number of inter-trial repetitions given the number of words recalled by the patient. Regression of these PF scores onto the predictors produced residuals that were also positively skewed violating the regression assumption that errors are normally distributed (Field, 2009). Therefore, PF total scores at three months post-injury were transformed to their natural logarithmic (ln) equivalents in order to create a more normal distribution of standardized residuals in the regression analyses. The natural log transformation was chosen (as opposed to a square root transformation) because this transformation had the greatest effect in creating normally distributed errors in regression analyses when multiple transformations were attempted.

To demonstrate the utility of using the log transformed total PF scores, a hierarchical regression model with all planned predictors (BD1 GCS, worst CT scan, age, gender, ethnicity, education level) was run using the untransformed total PF scores at three months. Using +/- 1 as a cutoff for values of skewness and kurtosis, the distribution of residuals for the regression model with untransformed total PF scores for three months post injury had a skewness of 1.68 (SE = .23) and kurtosis of 4.63 (SE = .45). The log transformation normalized the distribution of errors as evidenced by skewness of -.014 (SE = .23) and kurtosis of -.532 (SE = .45) for the distribution of residuals for the log transformed data. Figure 8 displays comparisons in residual statistics for the regression analysis with untransformed PF scores vs. log-transformed PF scores. The transformed PF total scores will be notated as lnPF to indicate the natural log transformation. Due to the presence of negative and zero values of PF scores present in the data, a constant (*c*) was added to raw PF total scores to ensure all PF values were positive because the log of any number zero or less is undefined. A constant of two was added to raw scores because the lowest PF total score was -1.69. For the log transformation of the three-month data, the log (ln) of the raw PF scores plus the constant [$\ln(Y+c)$] was computed for each individual.

Three month PF scores were transformed to natural log equivalents so that the model parameter *b* could be interpreted as a reflection of the percentage change in PF scores due to a unit of change in a particular predictor. A literal interpretation of the resulting regression statistics from the log transformed outcome variable is that every unit of change in a predictor variable *X* corresponds to a change of *b* in the $\ln(Y+c)$ (holding all other predictors constant) in which *b* is the unstandardized b-weight of the predictor, *X* is the raw PF score and *c* is the constant added to each score. However, due to the properties of natural logarithms, *b* can also be interpreted as an *approximate* percent change in Y (untransformed) that is associated with a unit of change in *X* given the following assumptions:

$$\log_e Y = lnY = Ye^b \cong 100(e^b - 1)$$

A square root transformation was applied to six month PF total scores to create a normal distribution of standardized residuals in the regression analyses for six month data. A square root transformation of the six-month data was applied as this transformation had the greatest effect in normalizing residuals in regression models. The same constant of two was added to raw PF scores at six months post injury to ensure all PF values were positive prior to applying the square root transformation. The square root of the raw PF scores plus the constant [$\sqrt{}$(Y+ *c*)] was computed for each individual. To demonstrate the utility of the square root transformation, a regression model with all planned predictors (BD1 GCS, worst CT scan, age, gender, ethnicity, education level) was fit to the data using the untransformed total PF scores. Figure 10 compares the distribution of the standardized residuals and normal probability plots for this regression model in the untransformed data and square root transformed data to illustrate the validity of the square root transformation. The distribution of errors in the untransformed data was non-normal as indicated by a skewness of 1.20 (SE = .26) and kurtosis of 2.023 (SE = .52). The square root transformation of the six month data normalized the distribution of errors as evidenced by skewness value of .41 (SE = .26) and kurtosis of .20 (SE = .52). The square root transformed PF total scores will be notated as √PF total scores to indicate this transformation. In all analyses of PF total scores, only the criterion variable (pair frequency scores) was transformed.

Hypothesis 1: *Severity of injury will be related to PF scores when demographic variables are taken into account with a higher severity of injury being related to lower PF scores at each of three and six months post-injury.*

Predicting Pair Frequency Total Scores at Three Months Post Injury

Hierarchical multiple regression analysis was used to evaluate the effect of severity of injury on PF total scores at three months post injury. As stated previously, a log (ln) transformation of PF total scores was conducted after adding a constant of two to each PF total score. The lnPF total score was then entered into the model as the criterion variable. All predictors were entered into the model untransformed using a Block entry method and the entry order of predictors was guided by theoretical considerations. Best Day 1 GCS total scores and worst CT scan Marshall categories were entered initially into the model as injury severity measures. Age was entered in a second block while gender, ethnicity and years of education were entered into the model in the third, fourth and fifth block, respectively.

The results of the initial hierarchical regression analysis for lnPF scores are displayed in Table 7. Injury severity measures (Best Day 1 GCS and Worst CT scan classification) entered in the first block accounted for 8% of the variance in lnPF scores at three months post injury, $R^2 = .084$, F (3, 112) = 3.424, $p = .02$ and were a significant predictor of lnPF scores. Age was entered in the second block and was a marginally significant predictor of lnPF scores above injury severity (R^2_{change} = .075, F_{change} (1, 111) = 9.915, p =.002), accounting for an additional 7.5% of the variance in lnPF scores. Sex was entered in the third block and increased the explained variance in lnPF scores by 2.4%, R^2_{change} = .054, F_{change} (1, 110) = 3.289, p =.072. Ethnicity was entered in the fourth block and failed to increase the predictive power of the model, R^2_{change} = .000, F_{change} (2, 108) = 0.584, p =.560. Education level was entered in the fifth block and also failed to increase the predictive power of the model R^2_{change} = .001, F_{change} (1, 107) = .132, p =.718. Due to the possibility that ethnicity and education could be redundant predictors, another model was run in which education was added to the model in the fourth block, before ethnicity which was added in the fifth block. Adding education in the fourth block before ethnicity did not account for additional variance above injury severity, age and sex, R^2_{change} = .000, F_{change} (1, 109) = .003, p =.955. Ethnicity added in the fifth step after education did not account for additional variance in explained PF scores over education, R^2_{change} = .01, F_{change} (2, 107) = .643, p =.528. In short, education accounted for less than 1% of the variance in total lnPF scores regardless of when it was entered in the model relative to ethnicity. After all predictors were entered into the initial model, Best Day 1 GCS scores and age emerged as significant predictors of lnPF scores, while worst CT scan classification, age, ethnicity and years of education failed to contribute uniquely to the variance in lnPF scores.

A final regression model was fit to the data that did not include Worst CT scan classification, sex, ethnicity or years of education because they were not significant predictors of lnPF scores in the initial model. The results of the final regression model are displayed in Table 8. Together, Best Day 1 GCS total scores and age accounted for about 14% of the variance in lnPF scores at three months post injury, R^2 = .144, $F(2, 115)$ = 9.639, p <.001. Best Day 1

GCS total score emerged as an independent predictor of lnPF scores. Specifically, a one point increase in Best Day 1 GCS total scores was associated with an increase of .079 in lnPF scores across 12 trials of the SRT. This is also approximately equivalent to an 8.2% increase in total PF scores (untransformed) for every one point increase in Best Day 1 GCS scores when other predictors are held constant. A one year increase in age was associated with a decrease of .018 points in lnPF scores across 12 trials, which is approximately equal to a 1.8% decrease in PF scores for every year increase in age when other predictors are held constant.

This model was tested to ensure it met the assumptions of multiple regression analysis including the absence of multicollinearity, assumption of normally distributed errors, independence of errors and homoscedasticity. The absence of collinearity between predictors was determined by the criterion of tolerance levels greater than 0.2 and VIF statistics of less than 10, (Field, 2009). Tolerance levels and VIF statistics were within appropriate ranges and thus this assumption was met. Normally distributed errors and homoscedasticity were confirmed by plotting standardized residuals on a normal probability plot and histogram and plotting standardized predicted lnPF scores against standardized residuals (Figure 9). The assumption of independence of errors was tested using the Durbin-Watson test, which tests the null hypothesis that there are no serial correlations between residuals in the regression model. The observed statistic for serial correlation between errors was $d = 2.094$, which was above the critical d upper limit of 1.582 (p=.01) (Savin & White, 1977) indicating residuals were uncorrelated with each other.

Another model was fit to the data with predictors from the final model (Best Day 1 GCS total scores and age as predictors) in addition to the number of days post injury at testing to evaluate if variations in the three month timeframe accounted for any variance in pair frequency scores. Days post injury at testing did not account for a significant portion of variance in lnPF scores explained above the original model, R^2_{change} = .000, F_{change} (1,112) < .001, *p* =.983.

Predicting Pair Frequency Total Scores at Six Months Post Injury

Hierarchical multiple regression analysis was used to evaluate the effect of severity of injury on PF total scores at six months post injury. After attempting both log and square root transformation to the data, a square root transformation was applied to total PF scores as it had the greatest effect in normalizing residuals in regression analyses. The square root transformation was applied to total PF scores after adding a constant of two to each raw score and entered into the model as the criterion variable. All predictors were entered into the model untransformed using a Block entry method. Best Day 1 GCS total scores and Worst CT scan Marshall categories were entered together in the first block. Age was entered in a second block and gender, ethnicity and years of education were entered into the model in the third, fourth and fifth block, respectively.

The results of the initial hierarchical multiple regression analysis for six month √PF total scores are displayed in Table 9. Injury severity measures (Best Day 1 GCS and Worst CT scan classification) accounted for 6.8% of the variance in √PF total scores at six months post injury, R^2 = .068, *F* (3, 81) = 1.969, *p* = .125. This model fit with only injury severity variables was not a significantly better predictor of √PF total scores relative to the mean estimation of √PF total scores. Age was entered in the second block and increased the amount of √PF total scores explained by 6.6%, R^2_{change} = .066, F_{change} (1, 80) = 6.117, *p* =.016. Sex was entered in the third block and failed to increase the predictive power of the model. Ethnicity was entered in the fourth block and failed to increase the predictive power of the model, R^2_{change} = .025, F_{change} (2, 77) = 1.197, *p* =.308. Years of education was entered in the fifth block and did not increase the predictive power of the model, R^2_{change} = .000, F_{change} (1, 76) = 0.015, *p* =.902. Due to the possibility that ethnicity and education could be redundant predictors, another model was run in which education was added to the model in the fourth block, before ethnicity, which was added in the fifth block. Adding education in the fourth block before ethnicity did not account for additional variance above injury severity, age and sex, R^2_{change} = .005, F_{change} (1, 78) = 0.472, *p* =.494. Ethnicity added in the fifth step after education did not account for additional variance in explained PF scores over education, R^2_{change} = .020, F_{change} (2, 76) = 0.954, *p* =.390. In short, education accounted for less than 1% of the variance in total lnPF scores regardless of when it was entered in the model relative to ethnicity. After all predictors were entered into the initial model, Best Day 1 GCS score and age emerged as significant predictors of √PF total scores while Worst CT scan classification, sex, ethnicity and education level failed to contribute uniquely to the explained variance in √PF total scores.

A final multiple regression model was fit to the data that excluded Worst CT scan classification, sex, ethnicity and years of education because they were not significant predictors of √PF total scores in the initial model. Table 10 displays the final regression model. Predictors in the final model included Best Day 1 GCS total score and age as predictors and accounted for about 14% of the variance in √PF total scores at six months post injury, R^2 = .138, *F*(2, 83) = 6.621, *p* = .002. Specifically, the model predicts √PF total scores increase by .122 with every one point increase in Best Day 1 GCS scores when other predictors are controlled. A one year increase in age was associated with a decrease in √PF total scores of .027 controlling for other predictors in the model.

The final model was tested to ensure it met the assumptions of multiple regression analysis including the absence of multicollinearity, assumption of normally distributed errors, independence of errors and homoscedasticity. The absence of collinearity between predictors was determined by the criterion of tolerance levels greater than 0.2 and VIF statistics of less than 10, (Field, 2009). Tolerance levels and VIF statistics were within appropriate ranges and thus this assumption was met. Normally distributed errors were confirmed by plotting standardized residuals on a normal probability plot and histogram. Homoscedasticity was evaluated by plotting standardized predicted √PF scores against standardized residuals (Figure 11). The presence of independent errors was tested using the Durbin-Watson test. The observed statistic for serial correlation between errors was d =2.152, which was above the critical d upper limit of 1.553 (p =.01) (Savin & White, 1977) indicating residuals were uncorrelated with each other.

Another model was fit to the data that included predictors from the final model (Best Day 1 GCS total scores and age as predictors) in addition to the number of days tested post-injury. Number of days tested post injury was not a significant predictor of six month PF scores, R^2_{change} = .001, F_{change} (1, 78) = .121, p =.729.

Predicting Pair Frequency Scores for the First Five Trials of the SRT

Hierarchical multiple regression analysis was used to test the hypothesis that increased injury severity would be associated with lower PF scores for the first five trials of the SRT. The distributions of three and six month post injury PF scores for the first five SRT trials are displayed in Figure 7. Though the average PF score for the first five SRT trials was 1.88 and 2.94 for three and six months, respectively, the majority of patients had very low scores with the exception of a few patients who performed well and had high PF scores causing a positively skewed distribution of PF scores for the first five SRT trials. Skewness of 1.94 (SE = .22) and kurtosis of 5.15 was observed for data at three months post injury. Skewness of 1.57 (SE = .26) and kurtosis of 3.16 (SE = .51) was observed for data at six months post injury. Regression models using raw PF scores created skewed distributions of residuals. Therefore raw PF scores were transformed in order to normalize the distribution of residuals in regression analyses and satisfy the regression assumption of normally distributed errors. A natural log transformation was chosen for five trial PF scores at three months after trying multiple transformations (log and square root) and determining that the log transformation had the greatest effect in normalizing the residuals in regression analyses. The distribution of errors for the initial hierarchical regression model with all planned predictors (Best Day 1 GCS, Worst CT Scan classification, age, sex, ethnicity and education level) with the untransformed three month first five trial data was non-normal with a skewness of 1.55 (SE = .23) and kurtosis of 4.30 (SE = .45). The natural log transformation normalized the distribution of errors as evidenced by skewness value of .46 (SE = .23) and kurtosis of .14 (SE = .47) for residuals of the log transformed data. PF scores were transformed by first adding a constant (*c*) of four to all raw five trial PF total scores as this was the smallest whole number that would give the lowest PF score of -3.325 a positive value. The logarithmic transformation is notated as *five trial lnPF scores* to indicate the natural log transformation of PF totals for the first five SRT trials.

A square root transformation was chosen for five trial PF scores at six months after trying multiple transformations (log and square root) and determining that the square root transformation had the greatest effect in normalizing the residuals in regression analyses. In the initial regression model with all predictors (Best Day 1 GCS, Worst CT scan, age, sex, ethnicity and level of education), the distribution of errors in the model with untransformed six month data was non-normal with a skewness of 1.32 (SE = .26) and kurtosis of 3.52 (SE = .52). The square root transformation increased the normalization of the distribution of errors as evidenced by skewness value of .58 (SE = .26) and kurtosis of 1.85 (SE = .52). The square root transformed PF scores are notated as *five trial √PF scores*. Figures 12 and 13 display the comparison between the distribution of residuals for the initial regression analysis for untransformed PF scores and transformed PF scores for three and six month data, respectively.

Predicting Pair Frequency Scores for the First Five SRT Trials at Three Months Post Injury

Hierarchical multiple regression analysis was used to evaluate the effect of severity of injury on five trial lnPF scores at three months post injury. Five trial lnPF score was entered as the criterion variable into the model. All predictors were entered into the model untransformed using a Block entry method. Best Day 1 GCS total scores and Worst CT scan Marshall categories were entered together in the model in the first block as injury severity measures. Age was entered in a second block while gender, ethnicity and education level were entered into the model in the third, fourth and fifth block, respectively. The results of the regression model are displayed in Table 11. Best Day 1 GCS total score and worst CT scan Marshall categories entered in the first block accounted for about 7% of the variance in five trial lnPF scores at three months post injury, $R^2 = .070$ F (3, 112) = 2.830, p =.042. Age, entered in the second block increased the explained variance in five trial lnPF scores by approximately 12%, $R^2_{change} = .122$, F_{change} (1, 111) = 16.845, p <.001. Sex was entered in the third block and did not significantly increase the explained variance in five trial lnPF scores, R^2_{change} = .122, F_{change} (1, 111) = 16.845, *p* <.001. Ethnicity was entered in the fourth block and increased the explained variance in five trial lnPF scores by 4.5%, R^2_{change} = .045, F_{change} (2, 108) = 3.244, *p* =.043. Education level was entered in the model in the fifth block and did not significantly increase the predictive power of the model, R^2_{change} = .002, F_{change} (1, 107) = .216, *p* =.643. Due to the possibility that ethnicity and education could be redundant predictors, another model was run in which education was added to the model in the fourth block, before ethnicity, which was added in the fifth block. Adding education in the fourth block before ethnicity did not account for additional variance above injury severity, age and sex, R^2_{change} = .001, F_{change} (1, 109) = 0.153, *p* =.696. Ethnicity added in the fifth step after education accounted for additional variance in explained PF scores over education, R^2_{change} = .046, F_{change} (2, 107) = 3.249, *p* =.043. In short, education accounted for less than 1% of the variance in total lnPF scores regardless of where it was entered in the model relative to ethnicity. After all predictors were entered into the initial model, Best Day 1 GCS, Worst CT scan classification, age and ethnicity emerged as significant independent predictors of five trial lnPF scores at three months post injury.

A final regression model was fit to the data that excluded sex and education level because they were not significant predictors of five trial lnPF scores in the initial model. Predictors in this final model included Best Day 1 GCS total scores, Worst CT scan classification, age and ethnicity as predictors (refer to Table 12) and accounted for approximately 25% of the total variance in five trial lnPF scores at three months post injury. R^2 = .245, F(6, 116) = 5.964, *p* = .031. Specifically, a one point increase in Best Day 1 GCS total scores was associated with an increase in five trial lnPF scores of .034. This is approximately equivalent to a 3.4% increase raw PF total scores across the first five SRT trials for every one point increase in Best Day 1 GCS score. A Diffuse I or Diffuse II worst CT classification relative to a Mass lesion classification was associated with a decrease in five trial lnPF scores of .216, which is approximately equivalent to a 19.43% decrease in raw PF total scores for the first five SRT trials. In contrast, having a worst CT Marshall classification of Diffuse III or IV relative to a mass lesion injury was not associated with PF scores for the first five SRT trials. Increasing age was associated with lower PF scores for the first five trials. The model predicted a decrease of .011 in lnPF scores for every one year increase in age, which is an approximate 1% decrease in raw PF scores across the first five trials of the SRT for every one year increase in age. Ethnicity was a significant predictor of five trial PF scores at three months post injury and the model predicted Hispanics had lower five trial PF scores relative to Non-Hispanic Whites. Specifically, Hispanics scored .23 points lower on lnPF scores relative to Non-Hispanic Whites, which is approximately equivalent to 20.6% lower raw PF scores relative to Non-Hispanic Whites. Blacks did not have significantly different five trial PF scores relative to Non-Hispanic Whites.

This model was tested to ensure it met the assumptions of linear regression analyses including the absence of multicollinearity, assumption of normally distributed errors, independence of errors and homoscedasticity. The absence of collinearity between predictors was determined by the criterion of tolerance levels greater than 0.2 and VIF statistics of less than 10, (Field, 2009). Tolerance levels and VIF statistics were within appropriate ranges and thus this assumption was met. The presence of normally distributed errors was evaluated by plotting standardized residuals on a normal probability plot and histogram. The plot of standardized predicted 12 trial lnPF total scores against standardized residuals suggests that mild heteroscedasticity is present in the model even after transformation. The assumption of independence of errors was tested using the Durbin-Watson test, which tests the null hypothesis that there are no serial correlations between residuals in the regression model. The observed statistic for serial correlation between errors was d =2.036, which was above the critical d upper limit of 1.67 (p =.01) (Savin & White, 1977) indicating that residuals were uncorrelated with each other. Residual statistics for the final model are displayed in Figure 14.

Another model was fit to the data using predictors from the final model (Best Day 1 GCS total scores, worst CT scan classification, age and ethnicity) entered in the first block. The number of days tested post injury was entered into the model in the second block to evaluate whether this would emerge as a significant predictor. Number of days tested post injury did not account for a significant portion of explained variance in five trial lnPF scores above the original model, $R^2_{change} < .001$, F_{change} (1, 107) = .023, p = .934.

Predicting Pair Frequency Scores for First Five SRT Trials at Six Months Post Injury

Hierarchical multiple regression analysis was used to evaluate the effect of severity of injury on √PF total scores for the first five SRT trials at six months post injury. Five trial √PF score was entered as the criterion variable into the model. All predictors were entered into the model untransformed using a block entry method. Best Day 1 GCS total scores and Worst CT scan Marshall categories were entered together in the first block as injury severity measures. Age was entered in a second block while gender, ethnicity and years of education were entered into the model in the third, fourth and fifth block, respectively. The results of the regression model are displayed in Table 13. Best Day 1 GCS total score and Worst CT scan Marshall categories entered in the first block accounted for about 4% of the variance in five trial √PF scores at six months post injury, although injury severity alone was not a significantly better predictor of five trial √PF scores relative to the mean estimation of five trial √PF scores, R^2 = .039, F (3, 81) = 1.103, p =.353. Age was entered in the second block and increased the explained variance in five trial √PF scores by about 5%, R^2_{change} = .049, F_{change} (1, 80) = .4.309, p =.041. Sex was entered in the third block and increased the explained variance in five trial √PF by approximately 4% although this was only marginally significant, R^2_{change} = .042, F_{change} (1, 79) = 3.775, p =.056. Ethnicity was entered in the fourth block and failed to increase the predictive power of the model, R^2_{change} = .028, F_{change} (1, 77) = 1.293, p =.280. Education level was entered in the model in the fifth block and also failed to increase the predictive power of the model, R^2_{change} = .023, F_{change} (1, 76) = 2.143, p =.147. Due to the possibility that ethnicity and education could be redundant predictors, another model was run in which education was added to the model in the fourth block, before ethnicity, which was added in the fifth block. Adding education in the fourth block before ethnicity added an additional 4% of explained variance above injury severity, age and sex, R^2_{change} = .041, F_{change} (1, 78) = 3.902, p =.052, which was marginally significant. Ethnicity added in the fifth step after education did not account for a significant portion of additional variance above education, R^2_{change} = .010, F_{change} (2, 76) = 0.460, p =.633. In this model, it appears that the variance attributed to ethnicity can actually be accounted for by education level. After all predictors were entered into the final block of the initial model, only Best Day 1 GCS and age emerged as significant predictors of √PF scores for the first five trials of the SRT at six months post injury.

A final model was fit to the data that included Best Day 1 GCS, age as well as education because after the removal of other predictors such as worst CT scan and ethnicity, education was a marginally significant predictor of five trial √PF scores. Results of the final model are displayed in Table 14. This final model accounted for approximately 15% of the variance in five trial √PF scores at six months post injury. R^2 = .015, $F(3,81) = 4.785$, p =.004. Specifically, increasing age was associated with lower five trial √PF scores with a decrease in five trial √PF scores of .01 points associated with every year increase in age. In this model, there was a trend for higher Best Day 1 GCS scores to be associated with an increase of .035 in five trial √PF scores however this did not reach statistical significance likely due to the reduced sample size at this time point.

This model was tested to ensure that it met the assumptions of linear regression analyses including the absence of multicollinearity, assumption of normally distributed errors, independence of errors and homoscedasticity. The absence of collinearity between predictors was determined by the criterion of tolerance levels greater than 0.2 and VIF statistics of less than 10, (Field, 2009). Tolerance levels and VIF statistics were within appropriate ranges and thus this assumption was met. The presence of normally distributed errors was confirmed by plotting standardized residuals on a normal probability plot and histogram. The plot of standardized predicted five trial √PF scores against standardized residuals from the model was analyzed to ensure homoscedasticity of residuals. The assumption of independence of errors was tested using the Durbin-Watson test, which tests the null hypothesis that there are no serial correlations between residuals in the regression model. The observed statistic for serial correlation between errors was d = 1.954, which was above the critical d upper limit of 1.58 (p=.01) (Savin & White, 1977) indicating that residuals were not correlated with each other. Regression residual statistics are displayed in Figure 15.

Another model was fit to the data from the final model entered in the first block. Number of days tested post injury was entered into the model in the second block to evaluate whether it was a significant predictor of √PF scores across the first five SRT trials at six months post injury. Number of days tested post injury did not account for a significant portion of explained variance in five trial √PF scores above the original model, $R^2_{change} = .008$, $F_{change}\ (1, 77) = .690$, $p = .409$.

Hypothesis 2: In a subset of patients who completed the SRT at both 3 and 6 months post injury, will obtain significantly higher pair frequency scores at six months post injury relative to score sat three months post injury when severity of injury (Best Day 1 GCS scores) and demographic variables are taken into account.

Changes in Total Pair Frequency Scores from Three to Six Months Post Injury

A repeated measures Analysis of Covariance (ANCOVA) was used to test the hypothesis that patients will obtain significantly higher PF scores at six months post injury relative to scores at three months post injury when severity of injury (Best Day 1 GCS scores) and demographic variables (age and gender) are taken into account. Time (three month vs. six month post injury follow up) was entered as the within-subjects factor and gender entered as a fixed factor. Age and Best Day 1 GCS total scores were entered as continuous variable covariates. Education information was available for only 74% of the sample and thus was excluded as a covariate due to the reduction in sample size for the analysis. Furthermore, ethnicity was excluded from the analyses as there were an unacceptable (>5% of cases) percentage of cells with expected cell counts of five or less for combinations of gender and ethnic groups. Prior to analyses, a one way ANOVA with gender as predictor and age and Best Day 1 GCS scores as dependent variables was conducted to ensure that age and GCS scores did not differ according to sex (to ensure that the covariate and fixed factors were independent). Age did not differ significantly between males and females, $F(1, 71) = .020$, $p = .887$. Best Day 1 GCS total scores also did not differ significantly between males and females, $F(1, 70) = .1.240$, $p = .269$. The assumption of sphericity in a repeated measures design refers to the assumption that the differences in each pair of within-subject factor levels have an equal variance (e.g., scores at time 1 and 2 have equal variance compared to scores at time 2 and 3 and score at time 1 and 3). This assumption is met when there are only two levels of a within-subjects factor. Thus, with only two time points as within subject factors in the current investigation, the sphericity assumption was met.

Age was entered as a covariate and held constant at 31.92 years in the model, while Best Day 1 GCS score was held constant at 9.35 in the model. The results of the repeated measures ANCOVA indicate there was no significant effect of follow-up point on PF total scores and PF scores overall were not significantly higher at six months post injury compared to scores at three months post injury when age and injury severity are controlled, $F(1, 68) = 2.44$, $p = .123$. Figure 16 displays the means and 95% confidence intervals for PF total scores at each time point. When age and Best Day 1 GCS scores were controlled, there was no main effect of sex on PF total scores, indicating that men and women did not differ in overall PF scores across time points $F(1, 68) = .733$, $p = .395$. Sex differences in total PF scores are displayed in Figure 17. The interaction between sex and follow up point was insignificant when age and Best Day 1 GCS score are controlled, indicating that men and women experienced similar changes in PF scores from three to six months post injury $F(1, 68) = .370$, $p = .545$. Sex differences in PF total scores at each time point are displayed in Figure 18.

Changes in Five Trial Pair Frequency Scores from Three to Six Months Post Injury

A repeated measures Analysis of Covariance (ANCOVA) was used to test the hypothesis that patients will obtain significantly higher PF scores for the first five SRT trials at six months post injury relative to scores at three months post injury when severity of injury (Best Day 1 GCS scores) and demographic variables (age and gender) are taken into account. Five trial PF scores at each time point (three month vs. six month post injury follow up) was entered as the within-subjects factor and gender entered as a fixed factor. Age and Best Day 1 GCS total scores were entered as continuous variable covariates. Education information was available for only 74% of the sample and thus was excluded as a covariate due to the reduction in sample size for the analysis as was ethnicity due to low cell counts of certain gender x ethnicity combinations.

Age was entered as a covariate and held constant at 31.92 years in the model, while Best Day 1 GCS score was held constant at 9.35 in the model. The results of the repeated measures ANCOVA indicate that when age and injury severity are held constant, there was no significant main effect of follow-up point on PF scores in the first five SRT trials $F(1, 68) = 1.784$, $p = .186$. Patients did not score significantly higher on five trial PF scores at six months compared to three months when injury severity and age are controlled. Average five trial PF scores at both time points are displayed in Figure 19. There was no main effect of sex on five trial PF scores, indicating that men and women did not differ in overall PF scores when age and Best Day 1 GCS scores were controlled, $F(1, 68) = .934$, $p = .337$. Average five trial PF scores for males and females are displayed in Figure 20. The interaction between sex and follow up point was insignificant when age and Best Day 1 GCS score were controlled, indicating that men and women experienced similar changes in PF scores from three to six months post injury, $F(1, 68) =$

.003, $p = .960$. Sex differences in five trial PF scores at each time point are displayed in Figure 21.

Hypothesis 3: Paired frequency scores will be positively associated with performance on the SRT. Correlations were evaluated separately for each time point.

Associations Between Pair Frequency Scores and Selective Reminding Test Performance

To evaluate the relationship between pair frequency and performance on the SRT, separate bivariate correlations between pair frequency total scores and SRT performance measures at three and six months were performed. The correlations between pair frequency total scores and SRT performance measures are displayed in Table 15. There was a strong positive correlation between pair frequency total scores and total recall, long term storage, consistent long term retrieval and cued recall (ranging from $r = .544$ - .800) at both three and six months. Furthermore, pair frequency total scores were also positively associated with measures of delayed recall at both time points (ranging from $r = .567$ to $r = .633$). Higher pair frequency totals were also associated with lower short term retrieval scores at both time points ($r = -.465$ to -.592), indicating that higher pair frequency scores were associated with recall from long term storage rather than short term storage. Higher pair frequency totals were also modestly associated with fewer total intrusions at both time points (ranging from $r = -.216$ to $r = -.261$). . Regarding recognition memory, pair frequency total scores were moderately associated with both immediate (ranging from $r = .348$ to $r = .400$) and delayed multiple choice recognition (ranging from $r = .383$ to $r = .425$) at both time points.

A similar pattern of relationships between pair frequency scores for the first five SRT trials and overall SRT performance was also observed and correlations are displayed in Table 16. There was a strong positive correlation between five trial pair frequency scores and total recall, long term storage, consistent long term retrieval and cued recall (ranging from r = .419 to r = .723) at both three and six months. Five trial pair frequency scores were positively correlated with measures of delayed recall at both time points (ranging from r = .475 to r = .544). Higher five trial pair frequency scores were also associated with lower short term retrieval scores at both time points (r = -.429 to r = -.551), indicating that higher pair frequency scores in the first five SRT trials were associated with recall from long term storage rather than short term storage. Higher five trial pair frequency scores were also modestly associated with fewer total intrusions at three months post injury. A similar relationship between five trial pair frequency scores and intrusions was observed at six months post injury, however this did not reach statistical significance. Regarding recognition memory, five trial pair frequency scores were moderately associated with both immediate (ranging from r = .290 to r = .331) and delayed multiple choice recognition (ranging from r = .347 to r = .349) at both time points.

Discussion

Relationship Between Injury Severity and Subjective Organization at Three and Six Months Post Injury

The current study evaluated the relationship between injury severity and the use of subjective organization on the SRT in the first six months post injury in a TBI sample. It was hypothesized that more severe injuries would be associated with less subjective organization as determined by the pair frequency measure. This hypothesis was generally supported as less severe injury (as indicated by higher Best Day 1 GCS scores) was associated with both higher total pair frequency scores and higher pair frequency scores for the first five SRT trials at three and six months post injury. Among injury severity injury measures entered into models (Worst CT Scan classification and Best Day 1 GCS score), the Best Day 1 GCS was the most consistent predictor in all models (regardless of time point or specific pair frequency measure). Pair frequency is a chance adjusted measure of the number of inter-trial repetitions present in adjacent trials (in either forward or backward order). Thus, injury severity was associated with more inter-trial repetitions on the SRT and increased recall organization. Because pair frequency scores were transformed for these analyses, a direct interpretation of the magnitude of these associations is not feasible. However, the log-transformations allow the slope of individual predictors (b) to be interpreted as a percent change in total pair frequency scores (rather than the log of this variable) for a unit change in a particular predictor. Thus, when age was held constant, Best Day 1 GCS scores were associated with an approximate 8% and 3% increase in total pair frequency scores and five trial pair frequency scores, respectively at three months post injury (when other predictors in the model are held constant).

Interestingly, among severity of injury variables, Worst CT scan classification was predictive only of first five trial pair frequency scores at a single time point (three months post injury) and not in the expected direction. Additionally, Worst CT scan classification was not related to pair frequency scores at six months post injury. Diffuse Injury I and II CT classifications relative to Mass Lesion I and II injuries were associated with lower five trial pair frequency scores at three months (holding other predictors constant) suggesting that lower levels of diffuse injury were associated with lower pair frequency scores compared to large evacuated and un-evacuated mass lesions (Mass Lesion I and II classifications). This finding is inconsistent with the prediction of more severe injury being associated with less organization. To remind the reader, Diffuse Injury I and II Marshall classifications include diffuse TBI with either no intracranial abnormalities present on CT scan in the case of Diffuse I, or diffuse injuries in which there are either cisterns present, a small midline shift (<5 mm) or lesion (<25 cc) in the case of Diffuse II. In contrast, Mass Lesion I and II classifications correspond to lesions >25cc that are surgically evacuated (Mass Lesion I) or not surgically evacuated (Mass Lesion II). One potential explanation for this finding is that though Diffuse Injury I/II classifications are less severe than Diffuse III/IV, it is possible that patients with Diffuse III/IV classifications who completed the SRT at three months post injury represented individuals who had better recovery and were less severe on the continuum compared to patients with the same Diffuse Injury III/IV classification who were not included in the study.

Previous research has largely centered around the relationship between location of injury (e.g., left vs. right hemisphere lesion or frontal vs. non-frontal injuries) and subjective organization, rather than severity of injury, which was evaluated in the current study. Therefore, it is difficult to compare the results of this study with those previous studies that generally find that less subjective organization is associated with left hemisphere insults, lesions within the frontal lobe and a double-effect of very poor organization noted in individuals with left frontal injuries (Alexander, Stuss, & Fansabedian, 2003; Geary, Kraus, Rubin, Pliskin, & Little, 2011; Gershberg & Shimanura, 1991; Gershberg & Shimamura, 1995). More detailed study of the relationship between lesion location and subjective organization use is planned for future studies. Comparisons with previous studies based on localization and lateralization is not possible in the current investigation. However, pair frequency scores in the first five trials at three months were lower for individuals with Diffuse I/II injury relative to Mass Lesion I/II which could signify that the more widespread effect characteristic of diffuse injuries was associated with poorer memory organization compared to individuals with more focal lesions, which is consistent with past findings described above of bilateral injury being associated with poor subjective organization. Findings from the current investigation agree with previous studies that have found that patients with mild TBI were more consistent in their recall across trials relative to moderate and severe TBI patients on the SRT (Pluth, Hannay, Massman, & Contant, 2003).

Among demographic variables, age emerged as a consistently significant predictor of both total pair frequency scores and first five trial pair frequency scores, with older age being associated with a 2% and 1% decrease in total pair frequency scores and five trial pair frequency scores, respectively at three months post injury when other predictors are held constant. Older age was also associated with lower pair frequency scores at six months post injury (both five trial and total scores). The contribution of other demographic factors to predicting subjective organization was variable. At three months post injury, ethnicity emerged as significant predictor of first five trial pair frequency scores and Hispanics had lower pair frequency scores compared to Non-Hispanic Whites. Previous research has not speculated on ethnic differences in subjective organization, thus these findings are interesting. Overall, Hispanics and Blacks had a lower (on average about 2 years less) educational level compared to Non-Hispanic Whites. Education emerged as a significant predictor of pair frequency scores for the first five SRT trials at six months post injury only, with higher educational level being associated with increased subjective organization, which is consistent with previous literature (Sharp, Cole, & Lave, 1979). Otherwise, reversing the order of entry for ethnicity and education did not reveal that education was the driving force for ethnic group differences in pair frequency scores for other models and after putting education before ethnicity in models, education accounted for little variance in pair frequency scores, while ethnicity accounted for additional variance beyond education in these other models.

Overall, more variance in pair frequency scores was accounted for at three months post injury relative to six months post injury, which is likely a function of how impaired the three month sample was relative to the six month sample. Injury severity appeared to influence pair frequency scores more at three than six months post injury. Patients in the study were not a rehabilitation sample and were recruited and followed from the acute care hospital. Therefore, they represent a sample of patients with more severe injuries relative to typical TBI samples recruited from rehabilitation programs (where they would have to meet certain requirements for admission which often select out the worse cases).

Pair frequency totals for the first five SRT trials were analyzed separately for several reasons. No data imputation was required for any of the first five trials because all patients completed at least five trials to before meeting criterion performance. Furthermore, while the SRT has twelve total learning trials, many other list learning tasks such as the RAVLT and CVLT have only five learning trials. Thus, studies evaluating subjective organization on these word lists could be compared with findings from the current study. For example, in a previous study comparing pair frequency scores on the RAVLT among a TBI group (reported loss of consciousness ranging from 1 hour to 112 days and tested nine months to 7 years following their injury), healthy control group and learning disability group, both clinical groups evidenced less subjective organization compared to the control group, but no distinction could be made between the TBI and learning disability group (O'Donnell, Radtke, Leicht, & Caesar, 1988). The summed pair frequency scores for the RAVLT, which is a 15-word list with five learning trials, in the TBI group was 3.3 in comparison to the pair frequency totals in the current study for the first five SRT trials (which has 12 words) at three months (1.98) and six months (3.17) post injury. The pair frequency total in O'Donnell et al. (1988) corresponds to about three intertrial repetitions across the five trials of the RAVLT which appears low, however this appears generally consistent with five trial pair frequency totals observed on the SRT, which is a 12-word list, in our sample. Levin and Goldstein (1986) found in their study comprised of a adults with severe closed head injury tested approximately 2-9 years post injury that the average pair frequency score across four learning trials of an experimental 18-word list learning task was -.03, indicating below chance performance. This is considerably lower than the average pair frequency for the first five trials in our subsample (.396 and .634 for three and six months respectively), although differences in list length and number of trials may account for discrepancy.

Regarding age, prior studies have found that older adults use less subjective organization relative to younger adults (Davis, Klebe, Guinther, Cornwell, & James, 2013; Guttentag, 1985; Witte, Freund, & Sebby, 1990). This is consistent with the negative relationship predicted between age and pair frequency scores in our statistical models.

Changes in Pair Frequency Use From Three and Six Months Post Injury

The current study also evaluated changes in pair frequency scores in a subgroup of patients who completed the SRT at both three and six months post injury. It was hypothesized that patients would have higher pair frequency scores at six months post injury relative to three months post injury when Best Day 1 GCS, age and gender are controlled. There was no significant change in either pair frequency total scores or five trial pair frequency scores from three to six months post injury when these factors were controlled. There were no gender differences or interactions between gender and age or gender and Best Day 1 GCS scores that may have influenced this effect which is telling in light of research showing a performance advantage on verbal list learning tasks for women compared to men (Herlitz, Nilsson, & Backman, 1997; Kramer, Delis, & Daniel, 1988; Krueger & Salthouse, 2010). Pair frequency scores for the first five SRT trials did appear to be influenced by Best Day 1 GCS and age being entered as covariates. Overall, pair frequency scores increased by approximately 1 point (refer to Table 6) from three to six months post injury without consideration of any covariates, indicating on average that patients gained one more inter-trial repetition over the three month time period between follow up evaluations. However, when these scores were adjusted for Best Day 1 GCS score and age, this change became non-significant, indicating that age and injury severity accounted for some of the variance in the small change in five trial pair frequency scores observed from three to six months post injury. Though no significant increases in either five trial or 12 trial pair frequency scores were observed when gender, age and injury severity were considered, the sample size was small (N = 73), which may have precluded differences from being observed. The lack of findings may not be unexpected given the fact that even at six months post injury, these patients were still in the early stages of recovery. Previous studies have been cross-sectional in nature rather than longitudinal and this is the first study to the author's knowledge to evaluate subjective organization in a TBI sample across different time points.

Relationship Between Subjective Organization and Recall

As predicted, in the current investigation, pair frequency scores were positively related to indicators of long term storage and recall. Positive correlations were observed between pair frequency scores and total recall, long term storage and consistent long term retrieval. Other studies with healthy adult subjects have also found strong correlations between measures of subjective organization and total recall on list learning tasks (r =.47-66) (Witte, Freund, & Sebby, 1990) and picture memory tasks (r = .354 - .635) (Laurence, 1966). This suggests that though individuals with TBI may show decreased subjective organization (this is suggested by the literature reviewed), the relationship between subjective organization and retrieval and storage in long term memory is the same for this group as it is for healthy individuals. A moderate negative relationship was observed between pair frequency scores and recall from short term storage. Bushke (1973) himself in describing the selective reminding procedure noted that subjective organization increases along with CLTR and LTR while being associated with lower

STR scores. The same pattern was observed in the current study in a group of individuals with TBI. This suggests that subjective organization is a function of long term memory rather than short term memory. It also provides evidence to the theory that subjective organization is an example of a control process that may facilitate the encoding of unrelated verbal material.

Study Strengths and Limitations

This study had several strengths including the inclusion of two time points in the prediction of subjective organization use, the inclusion of longitudinal data, ethnic diversity in the sample and the inclusion of different injury severity variables in models. Much of the recovery following TBI occurs within the first year post injury. When evaluating neuropsychological performance within this critical time frame, it is important to look at more than one time point as performance can change drastically within the first year due to spontaneous recovery of functioning. The inclusion of a subsample who completed a neuropsychological evaluation at both three and six months post injury also allowed for direct comparison of change in subjective organization over time with minimal within-subject differences because baseline characteristics (age at injury, education level, ethnicity, injury severity, etc.) did not change. The current study was ethnically diverse and allowed for the comparison of three major ethnic groups (Non-Hispanic Whites, Hispanics and Blacks) on subjective organization use on the SRT. In the current investigation, differences between ethnic minority and majority groups (Hispanics vs. Non-Hispanic Whites) contributed a small, but significant prediction to subjective organization scores on the first five SRT trials at three months post injury. Finally, the inclusion of multiple injury severity measures including Best Day 1 GCS scores and Worst CT Scan classification offer different methods of assessing injury severity. The Best Day 1 GCS relies on levels of consciousness as a measure of altered brain functioning and would be expected to be less influenced by intoxication, anesthesia, tranquilizers, etc. than a GCS score taken upon admission to the emergency room. The Worst CT scan classification relies on identification of anomalies in anatomical brain structures and reduces the possibility that initial scans may not fully account for progressively worsening injuries (e.g., swelling, increases in intracranial pressure, etc). Finally, the current study evaluated pair frequency scores for all 12 SRT trials as well as the first five SRT trials. The five trial pair frequency totals offer a measure of pair frequency free of potential ceiling effects that the 12 trial pair frequency scores may have contained due to imputation of data for patients who completed the test prior to trial 12. The five trial pair frequency scores also offer a method to roughly compare subjective organization in our sample to other studies that utilized the same measure on a five trial list learning task.

Weaknesses of the current study include the relatively small sample size, particularly at six months post injury and in the subsample of patients tested at both three and six months post injury and may have played a role in non-significant findings observed within the subsample. Still, sample sizes in the current study were larger than those for other studies with TBI or neurological samples, which typically include small groups of heterogeneous TBI or head injury patients with samples sizes for these groups ranging from 8 to 44 patients (Alexander, Stuss, & Fansabedian, 2003; Geary, Kraus, Rubin, Pliskin, & Little, 2011; Gershberg & Shimamura, 1995; Levin & Goldstein, 1986). Another limitation is the exclusion of intrusion errors and repetitions in the calculation of pair frequency scores, which could potentially under or over-estimate the pair frequency scores for patients. For example, if a patient consistently recalled one list item followed by a non-list intrusion, this would be excluded from the calculation of pair frequency scores. While it is customary for an experimenter to record the order in which patients recall list items (see Figure 5 for an example), intrusion errors are typically not included in this order and are recorded separately for each trial. Other studies also exclude intrusions and repetitions from the calculation of subjective organization (Geary, Kraus, Rubin, Pliskin, & Little, 2011; Gershberg & Shimanura, 1991; Gershberg & Shimamura, 1995; Stuss, et al., 1994). The CVLT-II, which provides the same measure of pair frequency used in the current study, does not count intrusions or repetitions in calculations of pair frequency (Delis, Kramer, Kaplan, & Ober, 2000). Other studies simply did not report how they handled intrusions and repetitions in test protocols (Alexander, Stuss, & Fansabedian, 2003; Heubrock, 1999; Levin & Goldstein, 1986; O'Donnell, Radtke, Leicht, & Caesar, 1988). A similar under or over-estimation of subjective organization in the data is possible given that some patients' pair frequency scores were imputed if they completed the SRT prior to trial 12 (11 patients at three months post injury and 18 patients at six months post injury had scores imputed). Specifically, the pair frequency score for the last two trials completed by these patients was imputed for remaining trials as this is typically done for SRT performance measures such as total recall, LTR, STR, CLTR, etc. on the SRT when an individual reaches criterion performance. However, no imputation was necessary for five trial pair frequency scores because none of the patients reached criterion performance. Another limitation of the current study was the exclusion of a control group. Many previous studies, although containing very small patient samples, compare this group to a healthy control group offering a relative determination of how 'impaired' patients are in their subjective organization. While the current study lacked a control group, it did have a range of injury severity levels.

Future Directions

Future studies evaluating the use of subjective organization in individuals with TBI should continue to utilize longitudinal designs, particularly when evaluating changes in subjective organization in the first year or two after injury. The current study focused on the two earliest time points available, three and six months post injury primarily due to concerns about loss of subjects and very small sample sizes at later time points (12 and 24 months post injury). However, one might expect significant changes in subjective organization at twelve months post injury relative to three months post injury. Future studies should control for injury severity and demographic factors that influence memory performance and organization in memory (such as age and gender) in addition to utilizing control groups consisting of patient-control groups. In his review of literature pertaining to memory functioning after TBI, Vakil (2005) recommends that researchers use patient-control groups (i.e., spinal cord injury patients) and larger sample sizes. Finally, if subjective organization is one control process used by individuals to aid in recall, it would be important to determine if one's use of this strategy correlates with other indicators of executive functioning and similar strategy use such as the use of subjective or semantic organization on phonemic and semantic fluency tasks. Such an investigation was outside of the scope of the current study. We plan to study the relationship between subjective organization and other indicators of executive functioning in future studies.

Appendix

Formulas for Subjective Organization Measures

Bousfield and Bousfield (1966) Intertrial Repetition Measure (ITR):

$$\mathrm{ITR} = \mathrm{O(ITR)} - \mathrm{E(ITR)} \text{where } \mathrm{E(ITR)} = \frac{c(c-1)}{hk}$$

PF	Pair frequency or bidirectional inter-trial item repetition
O (IRT)	Number of pairs of items recalled in the same order on trials t and t+1 in adjacent output position.
E (IRT)	Expected number of pairs of items
c	Number of common items recalled in trials t and t+1
h	Number of items recalled on trial t
k	Number of items recalled on trial t+1

Pair Frequency or Bidirectional ITR (Anderson & Watts, 1969)

$$\mathrm{PF} = \mathrm{O(ITR2)} - \mathrm{E(ITR)} \text{ with } \mathrm{O(ITR2)} = \frac{-2c(c-1)}{hk}$$

PF	Pair frequency or bidirectional inter-trial item repetition
O (IRT2)	Number of pairs of items recalled on trials t and t+1 in adjacent output position in either of two possible orders
E (IRT2)	Expected number of pairs of items
c	Number of common items recalled in trials t and t+1
h	Number of items recalled on trial t
k	Number of items recalled on trial t+1

Generalized ITR Measure (Pellegrino, 1971):

$$\mathrm{ITR} = \mathrm{O(ITR)} - \mathrm{E(ITR)}$$

Subjective Organization Measure (Tulving, 1962):

$$\mathrm{SO2} = \frac{\sum_{IJ} (n_{ij} + n_{ji}) \log(n_{ij} + n_{ji})}{2 \sum_i n_i \log(n_i)}$$

Adjusted ratio clustering (ARC') (Pellegrino, 1971):

$$\mathrm{ARC}' = \frac{\mathrm{O\ (ITR) - E\ (ITR)}}{\mathrm{M\ (ITR) - E\ (ITR)}} \text{ where } \mathrm{E(ITR)} = \frac{(N-X+1)!(A)(M-X+1-R)}{N!}$$

M	Number of items recalled on trial t
N	Number of items recalled on trial t+1
R	Number of units of size X from trials t that have one or more items not recalled on trial t+1
A	Variable parameter dependent on specific order within the subjective organization unit. Ex. A=1 for unidirectional, 2 for bidirectional and X! for unordered units
X	Size of the subjective organization unit
k	Number of items recalled on trial t+1

References

Alexander, M. P., Stuss, D. T., & Fansabedian, N. (2003). California verbal learning test: Performance by patients with focal frontal and non-frontal lesions. *Brain, 126*, 1493-1503.

Anderson, R. C., & Watts, G. H. (1969). Bidirectional associations in multi-trial free recall. *Psychonomic Science, 15*(6), 288-289.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. (K. W. Spence, & J. T. Spence, Eds.) *The Psychology of Learning and Motivation: Advances in Research and Theory, 2,* pp. 89-195.

Baddeley, A. (1968). Prior recall of newly learned items and the recency effect in free recall. *Canadian Journal of Psychology, 22*(3), 157-163.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews, 4,* 829-839.

Baddeley, A. D. (1976). *The Psychology of Memory.* New York: Basic Books, Inc.

Baddeley, A., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 8, pp. 47-87). New York: Academic Press.

Baddeley, A., & Hitch, G. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485-493.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158-173.

Battig, W., Allen, M., & Jensen, A. R. (1965). Priority of free recall of newly learned items. *Journal of Verbal Learning and Verbal Behavior, 4*, 175-179.

Beatty, W. W., Krull, K. R., Wilbanks, S. L., Blanco, C. R., Hames, K. A., & Paul, R. H. (1996). Further validation of constructs from the Selective Reminding Test. *Journal of Clinical and Experimental Neuropsychology, 18*(1), 52-55.

Benedict, R. H., Schretlen, G. 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.

Bousfield, A., & Bousfield, W. (1966). Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports, 19*, 935-942.

Brooks, D. N. (1975). Long and short term memory in head injured patients. *Cortex*, *11*, 309-340.

Buckner, R. L. (2003). Functional-anatomic correlates of control processes in memory. *The Journal of Neuroscience, 23*(10), 3999-4004.

Buschke, H. (1973). Selective reminding for analysis of memory and learning. *Journal of Verbal Learning and Verbal Behavior*, *12*, 543-550.

Buschke, H. (1976). Learning is organized by chunking. *Journal of Verbal Learning and Verbal Behavior, 15*(3), 313-324.

Buschke, H., & Fuld, P. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology, 24*, 1019-1025.

Bushke, H. (1977). Two-dimensional recall: Immediate identification of clusters in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior, 16*(2), 201-215.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? *Developmental Psychology, 27*(5), 703-722.

Chawda, M., Hildebrand, F., Pape, H., & Giannoudis, P. (2004). Predicting outcome after multiple trauma: which scoring system? *Injury, 35*, pp. 347-358.

Christidi, F., Bigler, E. D., McCauley, S. R., Schnelle, K., Merkley, T. L., Mors, M. B., . . . Wilde, E. A. (2011). Diffusion tensor imaging of the perforant pathway zone and its relation to memory function in patients with severe traumatic brain injury. *Journal of Neurotrauma, 28*, 711-725.

Corrigan, J. D., Whiteneck, G., & Mellick, D. (2004). Perceived needs following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 19*(3), 205-216.

Craik, F. I. (1970). Fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior, 9*, 143-148.

Craik, F. I., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Crosson, B., Novack, T. A., Trenerry, M. R., & Craig, P. L. (1988). California verbal learning test (CVLT) performance in severely head-injured and neurologically normal adult males. *Journal of Clinical and Experimental Neuropsychology, 10*(6), 754-768.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory and Cognition, 21*(2), 142-145.

Curran, T. (2000). Brain potentionals of recollection and familiarity. *Memory and Cognition, 28*(6), 923-938.

Curtis, C., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7*(9), 415-423.

Davis, H., Klebe, K., Guinther, P. S., Cornwell, E., & James, L. (2013). Subjective organization, verbal learning, and forgetting across the life span: From 5 to 89. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process, 39*(1), 1-26.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test-Second Edition (CVLT-II).* San Antonio, TX: Psychological Corporation.

Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology, 16*, 519-555.

Elie, C., & Payne, D. (1999). Scoring options for recall tests (SORT), version 2.0: Updated to include false memory analyses, nominal group creation, and phenomenological measures. *Behavior Research Methods, Instruments, and Computers, 31*(2), 380-383.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd edition ed.). London: Sage.

Fischer, J., Hannay, J., Loring, D., & Lezak, M. (2004). Observational Methods, Rating Scales, and Inventories. In M. Lezak, D. Howieson, & D. Loring, *Neuropsychological Assessment* (4th edition ed., pp. 719-721). New York, New York: Oxford University Press.

Fletcher, P. C., Frith, C. E., Grasby, P. M., Shallice, T., Frackowiak, R. S., & Dolan, R. J. (1995). Brain systems for encoding and retrieval of auditory-verbal memory: An in vivo study in humans. *Brain, 118*, 401-416.

Fuld, P. A., & Buschke, H. (1976). Stages of retrieval in verbal learning. *Journal of Verbal Learning and Verbal Behavior, 15*, 401-410.

Gathercole, S., & Baddeley, A. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200-213.

Gathercole, S., & Baddeley, A. (1990). The role of phonological memory in vocabulary acquision: A study of young children learning new names. *British Journal of Psychology, 81*, 439-454.

Gathercole, S., Willis, C., Emslie, H., & Baddeley, A. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*(5), 887-898.

Geary, E. K., Kraus, M. F., Rubin, L. H., Pliskin, N. H., & Little, D. M. (2011). Verbal learning strategy following mild traumatic brain injury. *Journal of the International Neurospychological Society, 17*, 709-719.

Gershberg, F. B., & Shimanura, A. P. (1991). The role of the frontal lobes in free recall. *Society for Neuroscience Abstracts, 17*, 136.

Gershberg, F., & Shimamura, A. P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia, 13*(10), 1305-1333.

Glanzer, M., & Cunitz, A. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351-360.

Goldstein, F. C., Levin, H. S., Boake, C., & Lohrey, J. H. (1990). Facilitation of memory performance through induced semantic processing in survivors of severe closed-head injury. *Journal of Clinical and Experimental Neuropsychology, 12,* 286-300.

Guttentag, R. (1985). Memory and aging: Implications for theories of memory development during childhood. *Developmental Review, 5*, 56-82.

Hannay, H. J., Levin, H. S., & Grossman, R. G. (1979). Impaired recognition memory after head injury. *Cortex, 15*(2), 269-283.

Hannay, J. H. (1986). *Experimental Techniques in Human Neuropsychology.* New York: Oxford University Press.

Hannay, J. H., & Levin, H. S. (1985). Selective reminding test: an examination of the equivalence of four forms. *Journal of Clinical and Experimental Neuropsychology, 7*(3), 251-263.

Heeger, D. (n.d.). *Signal Detection Theory*. Retrieved 2013, from http://www.cns.nyu.edu/~david/handouts/sdt/sdt.html

Herlitz, A., Nilsson, L., & Backman, L. (1997). Gender differences in episodic memory. *Memory and Cognition*, *25*(6), 801-811.

Heubrock, D. (1999). Subjective organization of verbal memory and learning in adolescents with brain damage. *Child Neuropsychology, 5*(1), 24-33.

Hintzman, D. L., & Caulton, D. A. (n.d.). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory and Language, 37*(1), 1-23.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language*, *33*, 1-18.

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory and Cognition, 26*(3), 449-462.

Hofman, W., Gschwender, T., Friese, M., Wiers, R., & Schmitt, M. (2008). Workimg memory capacity and self-regulatory behavior: Toward an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology*, *95*(4), 962-977.

Jacobs, B., Beems, T., Stulemeijer, M., Van Vugt, A., Van der Vliet, T., Borm, G. F., & Vos, P. E. (2010, April). Outcome prediction in mild traumatic brain injury: age and clinical variables are stronger predictors than CT abnormalities. *Journal of Neurotrauma*, *27*, 655-668.

James, W. (1890). *The Principles of Psychology* (Vol. 1). New York: Henry Holt and Company. Retrieved from http://babel.hathitrust.org/cgi/pt?view=2up;size=100;id=nyp.33433070248038;page=root;seq=6;orient=0

Jennett, B. (2002). The glasgow coma scale: History and current practice. *Trauma*, *4*, 91-103.

Jennett, B., & Bond, M. (1975). Assessment of outcome after severe brain damage. *Lancet*, *i*, pp. 480-484.

Jennett, B., & Teasdale, G. (1981). Assessment of Impaired Consciousness. In B. Jennett, & G. Teasdale, *Management of head injuries* (pp. 77-84). Philadelphia: F.A. Davis.

Kahana, M., & Wingfield, A. (2000). A functional relation between learning and organization in free recall. *Psychonomic Bulletin and Review*, *7*(3), 516-521.

Kramer, J., Delis, D., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology*, *44*(6), 907-915.

Kraus, J. F., & McArthur, D. L. (1996). Epidemiologic Aspects of Brain Injury. *Neurologic Clinics*, *14*(2), pp. 435-451.

Krueger, L., & Salthouse, T. (2010). Differences in acquisition, not retention, largely contribute to sex differences in multitrial word recall performance. *Personality and Individual Differences, 49*, 768-772.

Larrabee, G. J., Trahan, D. E., & Levin, H. S. (2000). Normative data for a six trial administration of the Verbal Selective Reminding Test. *The Clinical Neuropsychologist, 14*(1), 110-118.

Larrabee, G. J., Trahan, D. E., Curtiss, G., & Levin, H. S. (1988). Normative data for the Verbal Selective Reminding Test. *Neuropsychology, 2*, 173-182.

Laurence, M. (1966). Age differences in performance and subjective organization in the free recall learning of pictorial material. *Canadian Journal of Psychology*, *20*(4), 388-399.

Levin, H. S. (1990). Memory deficit after closed-head injury. *Journal of Clinical and Experimental Neuropsychology*, *12*(1), 129-153.

Levin, H., & Goldstein, F. C. (1986). Organization of verbal memory after severe closed-head injury. *Journal of Clinical and Experimental Neuropsychology, 8*, 643-656.

Levin, H., Grossman, R. G., Rose, J. E., & Teasdale, G. (1979). Long term neuropsychological outcome of closed head injury. *Journal of Neurosurgery, 50*, 412-422.

Levy, R., & Goldman-Rakic, P. S. (1999). Association of storage and processing functions in the dorsolateral prefrontal cortex of the nonhuman primate. *The Journal of Neuroscience, 19*(12), 5149-5158.

Lewis, M. Q. (1971). Short-term memory items in repeated free recall. *Journal of Verbal Learning and Verbal Behavior, 10*, 190-193.

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

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

Marshall, L. F., Marshall, S., Klauber, M., Clark, M. v., Eisenburg, H., Jane, J. A., . . . Foulkes, M. (1991). A new classification of head injury based on computerized tomography. *Journal of Neurosurgery, 75*(1s), S14-S20.

McCauley, S., Hannay, H., & Swank, P. (2001). Use of the disability rating scale recovery curve as a predictor of psychosocial outcome following closed-head injury. *Journal of the International Neuropsychological Society, 7*(4), 457-467.

Michas, I., & Henry, L. (1994). The link between phonological memory and vocabulary acquision. *British Journal of Developmental Psychology, 12*, 147-163.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Muller, N., & Knight, R. T. (2006). The functional neuroanatomy of working memory: Contributions of human brain lesion studies. *Neuroscience, 139*, 51-58.

Novinski, L. (1969). Part-whole and whole-part free recall learning. *Journal of Verbal Learning and Verbal Behavior*, *8*(1), 152-154.

O'Donnell, J. P., Radtke, R. C., Leicht, D. J., & Caesar, R. (1988). Encoding and retrieval processes in learning-disabled, head-injured, and nondisabled young adults. *The Journal of General Psychology*, *115*(4), 355-368.

Oliver, R., Bjoertomt, O., Greenwood, R., & Rothwell, J. (2008). 'Noisy patients'-can signal detection theory help? *Nature Clinical Practice Neurology*, *4*, 306-316.

Ornstein, P., Naus, M., & Liberty, C. (1975). Rehearsal and organizational processes in children's memory. *Child Development*, *46*, 818-830.

Paulesu, E., Frith, C., & Frackowiak, R. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*, 342-345.

Pellegrino, J. W. (1971). A general measure of organization in free recall for variable unit size and internal sequential consistency. *Behavior Research Methods and Instrumentation, 3*(5), 241-246.

Pellegrino, J. W., & Battig, W. F. (1974). Relationships among higher order organizational measures and free recall. *Journal of Experimental Psychology*, *102*(3), 463-472.

Peters, B., & Levin, H. (1977). Memory enhancement after physostigmine treatment in the amnesic syndrome. *Archives of Neurology, 34*, 215-219. Retrieved from http://archneur.ama-assn.org/

Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*(3), 193-198.

Pluth, S., Hannay, H. J., Massman, P. J., & Contant, C. F. (2003). Selective reminding test: Novel measures of performance in a CHI population. *Poster Presented at the Thirtieth Annual International Neuropsychological Society Conference.* Toronto, Canada.

Postman, L., & Keppel, G. (1968). Conditions determing the priority of new items in free recall. *Journal of Verbal Learning and Verbal Behavior, 7*(1), 260-263.

Ricker, J. H., Muller, R.-A., Zafonte, R. D., Black, K. M., Millis, S. R., & Chugani, H. (2001). Verbal recall and recognition following traumatic brain injury: A [O-15]-water positron emission tomography study. *Journal of Clinical and Experimental Neuropsychology, 23*(2), 196-206.

Robertson, C., Valadka, A., Hannay, J., Contant, C., Gopinath, S., Cormio, M., . . . Grossman, R. (1999). Prevention of secondary ischemic insults after severe head injury. *Critical Care Medicine, 27*(10), 2086-2095.

Roenker, D., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures of these estimation of clustering in free recall. *Psychological Bulletin, 76*(1), 45-48.

Rogoff, B. (1981). Schooling's influence on memory test performance. *Child Development, 52,* 260-267.

Ruff, R., Light, R., & Quayhagen, M. (1989). Selective reminding tests: A normative study of verbal learning in adults. *Journal of Clinical and Experimental Neuropsychology, 11,* 539-550.

Savage, C. R., Deckersbach, T., Heckers, S., Wagner, A. D., Schacter, D. L., Alpert, N. M., . . . Rauch, S. L. (2001). Prefrontal regions supporting spontaneous and directed application of verbal learning strategies evidence from PET. *Brain, 124*, 219-231.

Savin, N., & White, K. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica, 45*, 1989-1996.

Scheid, R., Walther, K., Guthke, T., Preul, C., & Yves, C. v. (2006). Cognitive sequelae of diffuse axonal injury. *Archives of Neurology, 63*, 418-424.

Servadei, F., Nasi, M., Giuliani, G., Cremonini, A., Cenni, P., Zappi, D., & Taylor, G. (2000). CT prognostic factors in acute subdural haematomas: the value of the 'worst' CT scan. *British Journal of Neurosurgery, 14*(2), 110-116.

Sharp, D., Cole, M., & Lave, C. (1979). Studies of memory and the learning of classifications. *Monographs of the Society for Research in Child Development, 44*(1-2), pp. 35-50.

Smtih, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*, 1657-1661.

Stallings, G., Boake, C., & Sherer, M. (1995). Comparison of the california verbal learning test and the Rey Auditory Verbal Learning Test in head-injured patients. *Journal of Clinical and Experimental Neuropsychology, 17*(5), 706-712.

Sternberg, R., & Tulving, E. (1977). The measurement of subjective organization in free recall. *Psychological Bulletin, 84*(3), 539-556.

Stuss, D. T., Alexander, M. P., Palumbo, C. L., Buckle, L., Sayer, L., & Pogue, J. (1994). Organizational strategies of patients with unilateral or bilateral frontal lobe injury in word list learning tasks. *Neuropsychology*, 355-373.

Swanson, H. L. (1999). Reading comprehension and working memory in learning disabled readers: Is the phonological loop more important than the executive system? *Journal of Experimental Child Psychology, 72*(1), 1-31.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness. *Lancet, 304*(7872), 81-84.

Teasdale, G., & Jennett, B. (1976). Assessment and prognosis of coma after head injury. *Acta Neurochirurgica, 34*, 45-55.

Teasdale, G., & Murray, L. (2000). Revisiting the glasgow coma scale and coma score. *Intensive Care Medicine, 26*, 153-154.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review, 69*(4), 344-354.

Tulving, E. (1966). Subjective organization and effects of repetition in mutli-trial free recall learning. *Journal of Verbal Learning and Verbal Behavior, 5*(2), 193-197.

Vakil, E. (2005). The effect of moderate to severe traumatic brain injury (tbi) on different aspects of memory: A selective review. *Journal of Clinical and Experimental Neuropsychology, 27*(8), 977-1021.

Vanderploeg, R. D., Crowell, T., & Curtiss, G. (2001). Verbal learning and memory deficits in traumatic brain injury encoding, consolidation, and retrieval. *Journal of Clinical and Experimental Neuropsychology, 23*(2), 185-195.

Waugh, N., & Norman, D. (1965). Primary memory. *Psychological Review, 72*, 89-104.

Witte, K., Freund, J., & Sebby, R. (1990). Age differences in free recall and subjective organization. *Psychology and Agining, 5*(2), 307-309.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517. doi:10.1006/jmla.2002.2864

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5*, 418-441.

Table 1. Definitions of Measures Obtained from SRT (Buschke, 1974a)

Recall	Total number of items recalled on each trial
Short Term Storage (STS)	Item stored after immediate presentation of that item
Long Term Storage (LTS)	Recall without further presentation. Occurs on trials that precede items recalled without presentation and shows item learning. An item recalled on two consecutive trials without further presentation after the initial presentation is assumed to enter LTS after the first presentation.
Short Term Retrieval (STR)	Items recalled only after direct presentation. Indicates recall of items not transferred to LTS. STR = Recall – LTR on individual trials.
Long Term Retrieval (LTR)	Recall of words from LTS and indicates list learning.
Consistent Long Term Retrieval (CLTR)	Number of items recalled consistently on subsequent trials without further presentation of the items after LTS has been achieved.
Random Long Term Retrieval (RLTR)	Inconsistent recall of words from LTS. RLTR= LTR- CLTR

Table 2. Comparison of Common Verbal Learning and Memory Tests

	SRT	CVLT-II	RAVLT	HVLT-R
Age range for administration	18-91	16-89	6-89	13-80+
Number of Words on List	12	16	15	12
Inter-stimulus interval	2 sec	1-2 sec	1 sec.	2 sec.
Semantic Categorization	Unrelated	4 semantic categories: (Standard Form List A: Furniture, vegetables, ways of traveling and animals)	Unrelated nouns	3 semantic categories (Form I: animals, precious stones, human dwellings)
Number of learning trials	12	5	5	3
Modality	Auditory	Auditory	Auditory	Auditory
Delayed Recall Length (in minutes)	30	20	20	20-25
Delayed Free Recall Modality	Auditory	Auditory	Auditory	Auditory
Cued Recall	First 2-3 letters of each word given	Semantic category cues given (“Tell me all the words on the list that are animals”)	None	None
Cued Recall Modality	Visual	Auditory	N/A	N/A
Recognition	Multiple Choice (list word, synonym,	Yes/No delayed recognition Forced choice	Yes/No delayed recognition	Yes/No delayed recognition (12 target words,

	homonym, unrelated distracter)	recognition (target and foil)	(50 words with list A and B items, semantically related and phonemically related items)	6 related distracters, 6 unrelated distracters)
Recognition Modality	Visual	Auditory	Visual *auditory if Ss has visual/literacy problems	Auditory
Interference Task and Modality	None	List B given after List A learning trials (vegetables, animals, instruments, house parts for Standard Form) Auditory	List B (15 words) given after List A learning trials Auditory	None
Measures of Subjective Organization	None	Semantic, serial and subjective clustering measures, primacy-recency effects	None	None
Total Number of Forms	4 (Forms 1-4)	2 (Standard and Alternate Form)	Geffen et al. (1994b) and Majdan et al. (1996) offer comparable alternate test forms	6
Test Procedure	12 selective reminding learning trials—visual cued recall, multiple choice recognition, 30	5 List A learning trials—1 List B interference trial, free recall of List A, cued (semantic cue) recall of list A, 20 minute delay,	5 List A learning trials, 1 List B interference trial, free recall of List A, 20 minute delay,	3 learning trials, 20-25 minute delay, free recall of list, yes/no recognition

	minute delay, delayed free recall	free recall of List A, cued recall of list A, yes/no recog. Of list A items, optional forced choice recognition	free recall of List A, recognition list (visual or auditory)	

*Note: SRT: Selective Reminding Test (*Hannay & Levin, 1985 version, adapted from Levin et al., 1982; Peters & Levin, 1977) *CVLT-II: California Verbal Learning Test-II* (Delis, Kramer, Kaplan, & Ober, 2000)*; RAVLT: Rey Auditory-Verbal Learning Test; HVLT-R: Hopkins Verbal Learning Test-Revised* (Benedict, Schretlen, & Brandt, 1998).

Table 3. Sample Demographics

Three Month Post Injury Sample (N=121)				Six Month Post Injury Sample (N =87)		
Gender	*n*	%		*n*	%	
Male	100	82.6		70	80.5	
Female	21	17.4		17	19.5	
Ethnicity	*n*	%		*n*	%	
Non-Hispanic White	53	43.8		41	47.1	
Hispanic	28	23.1		20	23.0	
Black	40	33.1		26	29.9	
Mechanism of Injury	*n*	%		*n*	%	
[a]MVA	51	42.1		42	48.3	
Assault	28	23.1		17	19.5	
Pedestrian vs. MVA	11	9.1		7	8.0	
Fall/Jump	9	7.4		3	3.4	
Gunshot wound	7	5.8		3	3.4	
Bicycle	4	3.3		2	2.3	
Motorcycle	3	2.5		6	6.9	
Other	5	4.1		4	4.5	
Unknown	3	2.5		3	3.4	
[b]**Worst CT scan**	*n*	%		*n*	%	
D1/D2	65	54.6		41	47.7	
D3/D4	20	16.8		20	23.3	
M1/M2	34	28.6		25	29.1	
Best Day 1 GCS Total	*n*	%		*n*	%	
GCS 3-5	5	4.2		5	5.8	
GCS 6-8	44	37.3		39	45.3	
GCS 9-12	37	31.4		23	26.7	
GCS 13-15	32	27.1		19	22.1	
[c]GOAT Score	*n*	%		n	%	
GOAT 0-39	5	4.2		4	4.7	
GOAT 40-74	14	11.6		3	3.5	
GOAT 75 -100	101	84.2		79	91.9	
	Mean	(SD)	Range	Mean	(SD)	Range
Age (years)	31.68	14.21	15-78	31.54	14.24	15-78
Education (years)	11.33	2.56	3-20	11.51	2.51	3-20
GOAT score	85.03	16.98	17-100	86.92	16.33	12-100
Days tested post injury	97.40	14.75	50-138	192.33	16.00	175-266

Note: [a] MVA = motor vehicle accident [b] D1/D2= Diffuse I and II Marshall CT Classification, D3/D4 = Diffuse III and IV Marshall CT classification, M1/M2 = Mass I and II Marshal CT classification. [c]GOAT = Galveston Orientation and Amnesia Test.

Table 4. SRT Performance at Three and Six Months Post Injury

Selective Reminding Test Performance Measures						
	3 Months Post Injury			**6 Months Post Injury**		
	Mean	(SD)	Range	Mean	(SD)	Range
Total Recall (12 trials)**	83.74	29.56	3-137	95.34	31.98	23-141
[a]LTS Total*	73.64	39.06	0-136	87.66	43.02	0-140
[b]CLTR Total**	39.10	37.59	0-136	54.81	44.24	0-140
[c]STR Total	20.18	12.57	0-50	17.26	13.65	0-59
[d]Total Intrusions	5.77	5.99	0-32	4.64	5.56	0-24
Cued Recall**	6.80	3.31	0-11	8.15	3.25	0-11
Recognition	10.55	2.58	0-12	10.96	2.31	0-12
30 Minute Recall	6.45	3.74	0-12	7.46	3.92	0-12
30 Minute Recognition	10.45	2.47	0-12	10.89	2.27	3-12
[e]12 Trial PF Total	8.26	9.31	-0.06– 49.26	9.59	10.10	-1.70 – 44.91
[f] 5 Trial PF Total*	1.88	2.69	-2.04 – 14.43	2.94	3.60	-3.33 - 17.74

Note: [a]LTS = Long Term Storage [b]CLTR = Consistent Long Term Retrieval [c]STR = Short Term Retrieval [d]Total number of non-list intrusions across all 12 trials. [e]Total pair frequency scores for 12 SRT trials [f]Total pair frequency scores for the first five SRT trials.

* significant mean difference between three and six month samples at p<.05, ** significant mean difference between three and six month samples at p<.01.

Table 5. Demographics for Subsample Tested at Both Three and Six Months Post Injury (N = 73)

Gender	*n*		%
Male	57		78.1
Female	16		21.9
Ethnicity	*n*		%
Non-Hispanic White	33		45.2
Hispanic	18		24.7
African American	22		30.1
Mechanism of Injury	*n*		%
[a] MVA	36		49.3
Assault	15		20.5
Pedestrian vs. MVA	7		9.6
Fall/Jump	3		4.1
Gunshot wound	2		2.7
Bicycle	2		2.7
Motorcycle	2		2.7
Other	4		5.5
Unknown	2		2.7
[b]**Worst CT scan**	*n*		%
D1/D2	37		51.4
D3/D4	15		20.8
M1/M2	20		27.8
[c] **3 Month GOAT Score**	*n*		%
GOAT 0-39	4		5.6
GOAT 40-74	7		9.7
GOAT 75 -100	61		84.7
6 Month GOAT Score	*n*		%
GOAT 0-39	3		4.2
GOAT 40-74	3		4.2
GOAT 75 -100	66		91.7
Best Day 1 GCS total	*n*		%
GCS 3-5	4		5.6
GCS 6-8	33		45.8
GCS 9-12	19		26.4
GCS 13-15	16		22.2
	Mean	(SD)	Range
Age (years)	31.84	14.99	15-78
Education (years)	11.40	2.54	3-20

Note: [a] MVA = motor vehicle accident [b] D1/D2= Diffuse I and II Marshall CT Classification, D3/D4 = Diffuse III and IV Marshall CT classification, M1/M2 = Mass I and II Marshal CT classification. [c] GOAT = Galveston Orientation and Amnesia Test.

Table 6. SRT Performance for Subsample at Three and Six Months Post Injury

Selective Reminding Test Performance Measures						
	3 Months Post Injury			**6 Months Post Injury**		
	Mean	(SD)	Range	Mean	(SD)	Range
Total Recall (12 trials)***	81.21	29.87	3-133	97.39	30.45	23-141
[a]LTS Total***	69.67	38.95	0-134	89.54	41.44	9-140
[b]CLTR Total***	36.27	37.19	0-124	57.55	44.44	0-140
[c]STR Total**	21.65	12.52	1-50	17.54	13.80	1-59
[d]Total Intrusions*	6.42	6.67	0-32	4.64	5.57	0-24
Cued Recall***	6.51	3.41	0-11	8.43	3.02	1-11
Recognition***	10.52	2.58	1-12	11.10	2.07	2-12
30 Minute Recall***	6.04	3.88	0-12	7.59	3.78	0-12
30 Minute Recognition***	10.46	2.42	3-12	11.03	2.08	3-12
[e]12 Trial PF Total	8.41	10.14	-0.56-49.26	10.24	11.55	-1.70-54.22
[f]5 Trial PF Total**	1.98	2.82	-0.67-14.43	3.17	3.86	-3.33-17.74

Note: [a]LTS = Long Term Storage [b]CLTR = Consistent Long Term Retrieval [c]STR = Short Term Retrieval [d]Total number of non-list intrusions across all 12 trials. [e]Total pair frequency scores for 12 SRT trials [f]Total pair frequency scores for the first five SRT trials.

* significant mean difference for three and six months post injury at p<.05, ** significant mean difference for three and six months post injury at p<.01., *** significant mean difference for three and six months post injury at p<.001

Table 7. Initial Regression Model Predicting lnPF Total Scores at Three Months Post Injury

Block 1	**R^2 = .084**	**Adj. R^2 = .059,**	**p = .02**	
	B	β	t	*p*
[a]BD1 GCS	0.078	.291	3.145	.002
[b]<u>Worst CT Scan</u>				
D1/D2 vs. M1/M2	-0.250	-.136	-1.283	.202
D3/D4 vs. M1/M2	-0.169	-.070	-0.667	.506
Block 2	**R^2 = .159**	**Adj. R^2 = .129,**	**R^2_{change} = .075,**	**p =.002**
	B	β	t	*p*
BD1 GCS	0.087	.326	3.634	<.001
<u>Worst CT Scan</u>				
D1/D2 vs. M1/M2	-0.282	-.153	-1.498	.137
D3/D4 vs. M1/M2	-0.262	-.108	-1.069	.287
Age	-0.018	-.278	-3.149	.002
Block 3	**R^2 = .184**	**Adj. R^2 = .146,**	**R^2_{change} = .024,**	**p =.072**
	B	β	t	*p*
BD1 GCS	0.090	.338	3.791	<.001
<u>Worst CT Scan</u>				
D1/D2 vs. M1/M2	-0.299	-.162	-1.601	.112
D3/D4 vs. M1/M2	-0.286	-.118	-1.173	.243
Age	-0.018	-.276	-3.162	.002
[c]Female vs. Male	0.381	.157	1.814	.072
Block 4	**R^2 = .192**	**Adj. R^2 = .140,**	**R^2_{change} = .009,**	**p =.560**
	B	β	t	*p*
BD1 GCS	0.094	0.353	3.897	<.001
<u>Worst CT Scan</u>				
D1/D2 vs. M1/M2	-0.300	-.163	-1.567	.120
D3/D4 vs. M1/M2	-0.289	-.119	-1.179	.241
Age	-0.019	-.294	-3.272	.001
Female vs. Male	0.336	.138	1.537	.127
[d]<u>Ethnicity</u>				
Hispanic vs. Non-Hispanic White	-0.218	-.099	-1.016	.312
Black vs. Non-Hispanic White	-0.141	-.073	-0.735	.464
Block 5	**R^2 = .193**	**Adj. R^2 = .133,**	**R^2_{change} = .001,**	**p =.718**
	B	β	t	*p*
BD1 GCS	0.095	.358	3.894	<.001

Worst CT Scan				
D1/D2 vs. M1/M2	-0.294	-.160	-1.520	.132
D3/D4 vs. M1/M2	-0.296	-.122	-1.197	.234
Age	-0.019	-.295	-3.266	.001
Female vs. Male	0.351	.144	1.571	.119
Ethnicity				
Hispanic vs. Non-Hispanic White	-0.246	-.112	-1.075	.285
Black vs. Non-Hispanic White	-0.160	-.082	-0.801	.425
Years of Education	-0.013	-.035	-0.363	.718

Note: [a] BD1 GCS = Best Day 1GCS total score [b]Worst CT scan category was dummy coded into two separate variables using M1/M2 as a reference category. [c]Sex was dummy coded using males as a reference category. [d]Ethnicity was dummy coded into two separate variables using Caucasians as a reference category.

*p<.05, **p<.01.

Table 8. Final Linear Regression Model Predicting lnPF Total Scores at 3 Months Post Injury

Block 1	**R^2 = .144**	**Adj. R^2 = .129, *p* <.001**		
	B	β	t	*p*
[a]BD1 GCS	0.079	.294	3.384**	.001
Age	-0.018	-.275	-3.170**	.002

Note: [a] BD1 GCS = Best Day 1GCS total score

*p<.05, **p<.01.

Table 9. Initial Hierarchical Multiple Regression for √PF Total Scores at Six Months Post Injury

Block 1 **R^2 = .068** **Adj. R^2 = .033, p = .125**				
	B	β	t	*p*
[a]BD1 GCS	0.103	.254	2.359*	.021
[b]Worst CT Scan				
D1/D2 vs. M1/M2	0.060	.021	0.168	.867
D3/D4 vs. M1/M2	0.198	.060	0.474	.637
Block 2 **R^2 = .134** **Adj. R^2 = .091** **R^2 Change = .066, p = .016**				
	B	β	t	*p*
BD1 GCS	0.122	.30	2.829**	.006
Worst CT Scan				
D1/D2 vs. M1/M2	-0.018	-.006	-0.052	.959
D3/D4 vs. M1/M2	0.001	.000	0.002	.998
Age	-0.026	-.266	-2.473*	.016
Block 3 **R^2 = .168** **Adj. R^2 = .115** **R^2 Change = .034, p = .078**				
	B	β	t	*p*
BD1 GCS	0.132	.327	3.088**	.003
Worst CT Scan				
D1/D2 vs. M1/M2	-0.070	-.025	-0.204	.839
D3/D4 vs. M1/M2	-0.034	-.010	-0.084	.933
Age	-0.027	-.273	-2.569*	.012
[c]Female vs. Male	0.652	.186	1.789	.078
Block 4 **R^2 = .193** **Adj. R^2 = .120** **R^2 Change = .025 , p = .308**				
	B	B	t	*p*
BD1 GCS	0.133	.329	3.034**	.003
Worst CT Scan				
D1/D2 vs. M1/M2	-0.045	-.016	-0.130	.897
D3/D4 vs. M1/M2	-0.056	-.017	-0.136	.892
Age	-0.031	-.313	-2.769**	.007
Female vs. Male	0.493	.140	1.292	.200
[d]Ethnicity				
Hispanic vs. Non-Hispanic White	-0.557	-.165	-1.399	.166
Black vs. Non-Hispanic White	-0.390	-.127	-1.103	.273
Block 5 **R^2 = .193** **Adj. R^2 = .108** **R^2 Change = .000, p = .902**				
	B	β	t	*p*
BD1 GCS	0.132	.326	2.907**	.005
Worst CT Scan				

D1/D2 vs. M1/M2	-0.047	-.017	-0.134	.894
D3/D4 vs. M1/M2	-0.055	-.017	-0.134	.894
Age	-.031	-.312	-2.736**	.008
Female vs. Male	0.483	.138	1.234	.221
Ethnicity				
Hispanic vs. Non-Hispanic White	-0.539	-.160	-1.262	.211
Black vs. Non-Hispanic White	-0.378	-.123	-1.024	.309
Years of Education	0.008	.015	0.123	.902

Note: [a] BD1 GCS = Best Day 1GCS total score [b] Worst CT scan category was dummy coded into two separate variables using M1/M2 as a reference category. [c] Sex was dummy coded using males as a reference category. [d]Ethnicity was dummy coded into two separate variables using Non-Hispanic Whites as a reference category.

*p<.05, **p<.01.

Table 10. Final Regression Model of √ PF Total Scores at Six Months Post Injury

Block 1	**R^2 = .138**	**Adj. R^2 = .117, *p* = .002**			
		B	β	t	*p*
[a] BD1 GCS		0.122	.300	2.899**	.005
Age		-0.027	-.274	-2.649*	.010

Note: [a] BD1 GCS = Best Day 1GCS total score

*p<.05, **p<.01.

Table 11. Initial Hierarchical Regression Model for First Five Trial lnPF Scores at Three Months Post Injury

Block 1 R^2 **= .070** Adj. R^2 **= .046, *p* = .042**				
	B	β	t	*p*
[a]BD1 GCS	0.025	.219	2.351*	.020
[b]Worst CT Scan				
D1/D2 vs. M1/M2	-0.179	-.230	-2.150*	.034
D3/D4 vs. M1/M2	-0.082	-.080	-0.759	.450
Block 2 R^2 **= .193** Adj. R^2 **= .164** R^2 **Change = .122, *p* < .001**				
	B	β	t	*p*
BD1 GCS	0.030	.264	3.003	.003
Worst CT Scan				
D1/D2 vs. M1/M2	-0.196	-.252	-2.511	.013
D3/D4 vs. M1/M2	-0.132	-.129	-1.301	.196
Age	-0.10	-.355	-4.104	<.001
Block 3 R^2 **= .201** Adj. R^2 **= .165** R^2 **Change = .009, *p* = .281**				
	B	β	t	*p*
BD1 GCS	0.031	.271	3.075	.003
Worst CT Scan				
D1/D2 vs. M1/M2	-0.200	-.257	-2.563	.012
D3/D4 vs. M1/M2	-0.138	-.135	-1.357	.178
Age	-0.010	-.354	-4.096	<.001
[c]Female vs. Male	0.095	.093	1.083	.281
Block 4 R^2 **= .247** Adj. R^2 **= .198** R^2 **Change = .045, *p* = .043**				
	B	β	t	*p*
BD1 GCS	0.034	.304	3.476**	.001
Worst CT Scan				
D1/D2 vs. M1/M2	-0.220	-.283	-2.813**	.006
D3/D4 vs. M1/M2	-0.152	-.148	-1.512	.133
Age	-0.011	-.402	-4.631**	<.001
Female vs. Male	0.074	.072	0.827	.410
[d]Ethnicity				
Hispanic vs. Non-Hispanic White	-0.222	-.238	-2.533*	.013
Black vs. Non-Hispanic White	-0.058	-.071	-0.742	.460
Block 5 R^2 **= .248** Adj. R^2 **= .192** R^2 **Change = .002, *p* = .643**				
	B	β	t	*p*
BD1 GCS	0.035	.310	3.495	.001

Worst CT Scan				
D1/D2 vs. M1/M2	-0.217	-.279	-2.747	.007
D3/D4 vs. M1/M2	-0.155	-.151	-1.536	.128
Age	-0.011	-.403	-4.623	<.001
Female vs. Male	0.082	.080	0.896	.372
Ethnicity				
Hispanic vs. Non-Hispanic White	-0.237	-.255	-2.529	.013
Black vs. Non-Hispanic White	-0.068	-.083	-0.835	.406
Years of Education	-0.007	-.044	-0.465	.643

Note: [a] BD1 GCS = Best Day 1GCS total score [b]Worst CT scan category was dummy coded into two separate variables using M1/M2 as a reference category. [c]Sex was dummy coded using males as a reference category. [d]Ethnicity was dummy coded into two separate variables using Non-Hispanic Whites as a reference category.

*p<.05, **p<.01.

Table 12. Final Regression Model for Five Trial lnPF Scores at Three Months Post Injury

Block 1	**R^2 = .245**	**Adj. R^2 = .204, p = .031**			
		B	β	t	*p*
[a]BD1 GCS		0.034	.302	3.486*	.001
[b]Worst CT Scan					
D1/D2 vs. M1/M2		-0.216	-.278	-2.790*	.006
D3/D4 vs. M1/M2		-0.146	-.142	-1.473	.144
Age		-0.011	-.405	-4.704**	<.001
[c]Ethnicity					
Hispanic vs. Non-Hispanic White		-0.231	-.248	-2.672*	.009
Black vs. Non-Hispanic White		-0.077	-.095	-1.032	.304

Note: [a] BD1 GCS = Best Day 1GCS total score [b]Worst CT scan category was dummy coded into two separate variables using M1/M2 as a reference category. [c]Ethnicity was dummy coded into two separate variables using Non-Hispanic Whites as a reference category.

*p<.01, **p<.001.

Table 13. Initial Regression Model for Five Trial √PF Scores at Six Months Post Injury

Block 1	**R^2 = .039**	**Adj. R^2 = .004 ,**	***p* = .353**	
	B	β	t	*p*
[a] BD1 GCS	0.036	.198	1.811	.075
[b]Worst CT Scan				
D1/D2 vs. M1/M2	-0.020	-.015	-0.121	.904
D3/D4 vs. M1/M2	0.016	.011	0.085	.933
Block 2	**R^2 = .088**	**Adj. R^2 = .043**	**R^2 Change = .049, *p* = .041**	
	B	β	t	*p*
BD1 GCS	0.043	.238	2.184*	.032
Worst CT Scan				
D1/D2 vs. M1/M2	-0.050	-0.039	-0.310	.757
D3/D4 vs. M1/M2	-0.060	-0.040	-0.315	.753
Age	-0.010	-0.229	-2.076*	.041
Block 3	**R^2 = .130**	**Adj. R^2 = .075**	**R^2 Change = .042, *p* = .056**	
	B	β	t	*p*
BD1 GCS	0.049	.267	2.469*	.016
Worst CT Scan				
D1/D2 vs. M1/M2	-0.076	-.060	-0.480	.633
D3/D4 vs. M1/M2	-0.078	-.052	-0.414	.680
Age	-0.010	-.237	-2.179*	.032
[c] Female vs. Male	0.326	.206	1.943	.056
Block 4	**R^2 = .158**	**Adj. R^2 = .082**	**R^2 Change = .028, *p* = .280**	
	B	B	t	*p*
BD1 GCS	0.049	.269	2.433*	.017
Worst CT Scan				
D1/D2 vs. M1/M2	-0.064	-.050	-0.402	.689
D3/D4 vs. M1/M2	-0.088	-.059	-0.467	.642
Age	-0.012	-.279	-2.418*	.018
Female vs. Male	0.249	.158	1.425	.158
[d]Ethnicity				
Hispanic vs. Non-Hispanic White	-0.265	-.175	-1.452	.150
Black vs. Non-Hispanic White	-0.187	-.135	-1.150	.254
Block 5	**R^2 = .181**	**Adj. R^2 = .095**	**R^2 Change = .023, *p* = .147**	
	B	β	t	*p*
BD1 GCS	0.042	.232	2.053*	.043

<u>Worst CT Scan</u>				
D1/D2 vs. M1/M2	-0.072	-.057	-0.460	.647
D3/D4 vs. M1/M2	-0.085	-.057	-0.454	.651
Age	-0.012	-.267	-2.327*	.023
Female vs. Male	0.198	.125	1.115	.268
<u>Ethnicity</u>				
Hispanic vs. Non-Hispanic White	-0.168	-.111	-0.867	.389
Black vs. Non-Hispanic White	-0.121	-.088	-0.726	.470
Years of Education	0.044	.030	1.464	.147

Note: [a] BD1 GCS = Best Day 1GCS total score [b]Worst CT scan category was dummy coded into two separate variables using M1/M2 as a reference category. [c]Sex was dummy coded using males as a reference category. [d]Ethnicity was dummy coded into two separate variables using Non-Hispanic Whites as a reference category.

*p<.05, **p<.01.

Table 14. Final Regression Model for Five Trial √PF Scores at Six Months Post Injury

Block 1	**R^2 = .151**	**Adj. R^2 = .119**	***p* = .004**	
	B	β	t	*p*
[a]BD1 GCS	0.035	.190	1.799	.076
Age	-0.010	-.231	-2.222*	.029
Education level (in years)	0.065	.257	2.464*	.016

Note: [a] BD1 GCS = Best Day 1GCS total score [b]Sex was dummy coded using males as a reference category.

*p<.05 **p<.01

Table 15. Correlations Between Total PF Scores and SRT Performance Measures

	3 Month Total PF Score	**6 Month Total PF Score**
[a] Total PF	-	-
Total Recall	.724***	.664***
[b]LTS	.691***	.603***
[c]CLTR	.800***	.685***
[d]STR	-.592***	-.465***
[e]Intrusions	-.216**	-.261**
Cued Recall	.570***	.540***
[f]MC Recognition	.400***	.348***
[g]30 Min. Recall	.633***	.567***
[h]30 Min. Recognition	.425***	.383***

Note: [a] Total pair frequency scores for all 12 SRT trials [b]LTS = Long Term Storage [c]CLTR = Consistent Long Term Retrieval [d]STR = Short Term Retrieval [e]Total number of non-list intrusions across all 12 trials. [f] Multiple Choice Recognition(immediate) [g] 30 Minute Delayed Recall [h] 30 Minute Delayed Multiple Choice Recognition

* p<.05, ** p<.01, ***p<.001

Table 16. Correlations Between First Five Trial Pair Frequency Scores and SRT Performance Measures

	3 Month Five Trial PF Score	**6 Month Five Trial PF Score**
[a] 5 trial PF	-	-
Total Recall	.676***	.585***
[b]LTS	.651***	.544***
[c]CLTR	.723***	.596***
[d]STR	-.551***	-.429***
[e]Intrusions	-.194*	-.196
Cued Recall	.538***	.419***
[f]MC Recognition	.331***	.290**
[g]30 Min. Recall	.544***	.475***
[h]30 Min. Recognition	.349***	.347**

Note: [a] Total pair frequency scores for the first five SRT trials [b]LTS = Long Term Storage [c]CLTR = Consistent Long Term Retrieval [d]STR = Short Term Retrieval [e]Total number of non-list intrusions across all 12 trials. [f] Multiple Choice Recognition(immediate) [f] 30 Minute Delayed Recall [g] 30 Minute Delayed Multiple Choice Recognition [h]Total pair frequency scores for 12 SRT trials.

* p<.05, ** p<.01, ***p<.001

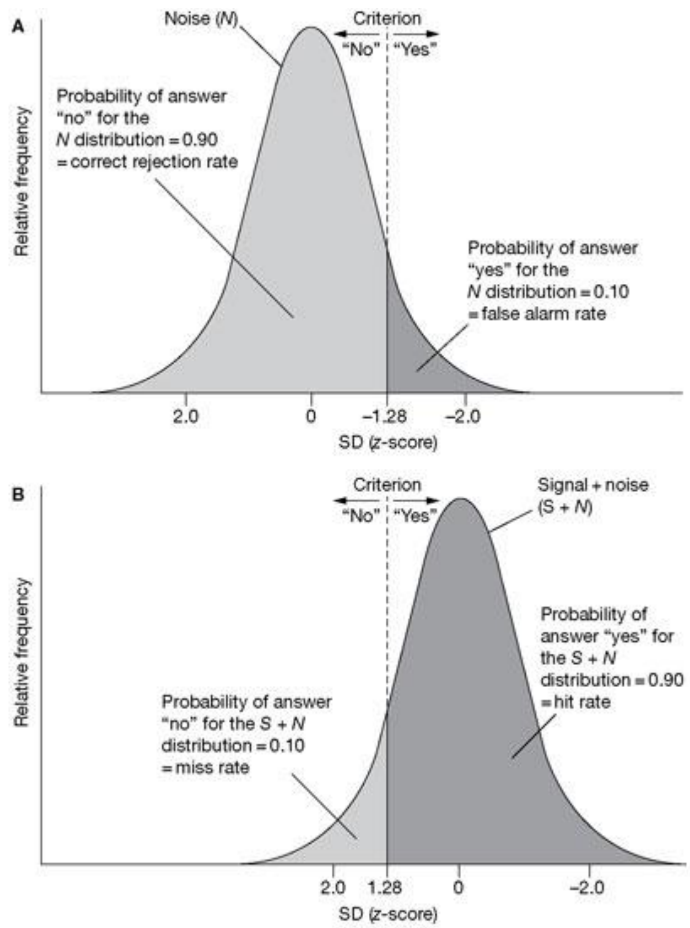


Figure 1. Distribution of sensory input on noise and signal + noise trials. Reprinted from Noisy patients'-can signal detection theory help?" by R. Oliver, O. Bjoertomt, R. Greenwood and J. Rothwell, 2008, *Nature Clinical Practice Neurology*, *4*, p. 308. Reprinted with permission.

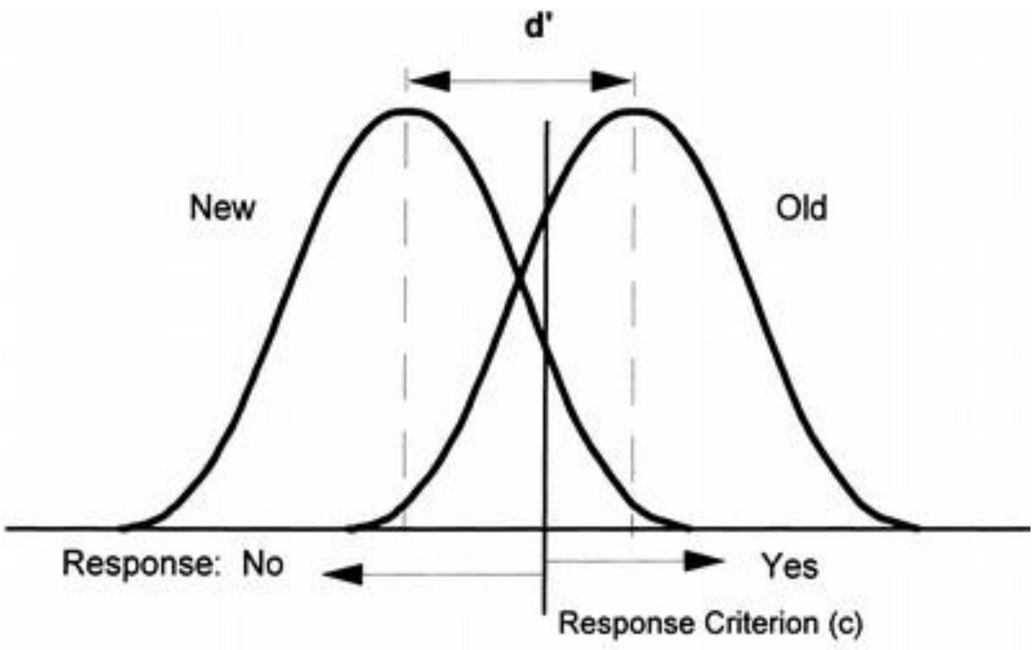


Figure 2. Distributions of signal + noise (new) and noise (old) trials drawn overlapping on the same X-axis). Reprinted from "Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection" by A. P. Yonelinas, I. Dobbins, M. D. Szymanski, H. S. Dhaliwal and L. King, 1996, *Consciousness and Cognition, 5*, p. 421. Reprinted with permission.

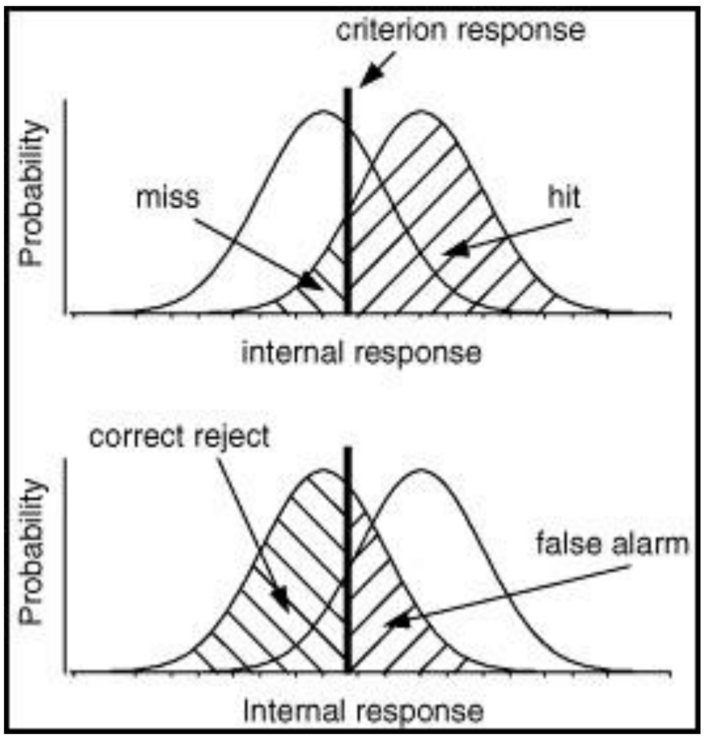


Figure 3. Relationship between response types and sensory input distributions. Reprinted from *Signal Detection Theory* by D. Heeger, 2003, Retrieved from http://www.cns.nyu.edu/~david/handouts/sdt/sdt.html. Reprinted with permission.

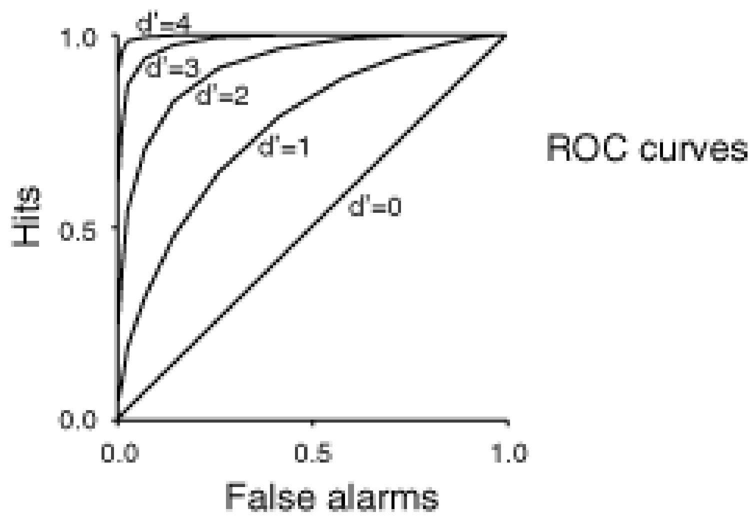


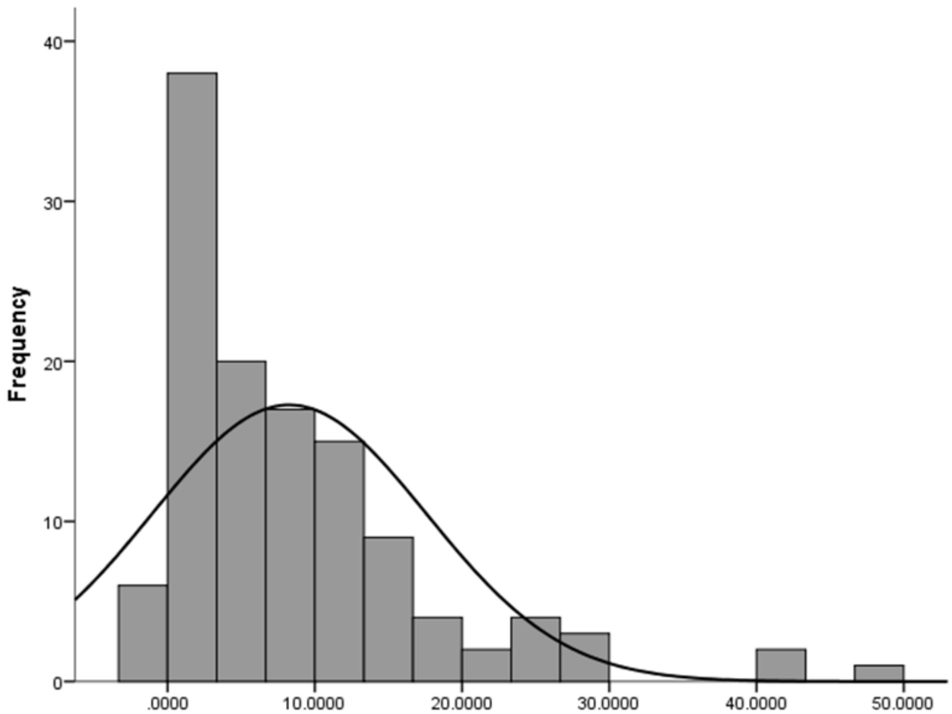
Figure 4. Receiver operating characteristic (ROC) curves. Reprinted from *Signal Detection Theory* by D. Heeger, 2003, Retrieved from http://www.cns.nyu.edu/~david/handouts/sdt/sdt.html. Reprinted with permission.

SELECTIVE REMINDING

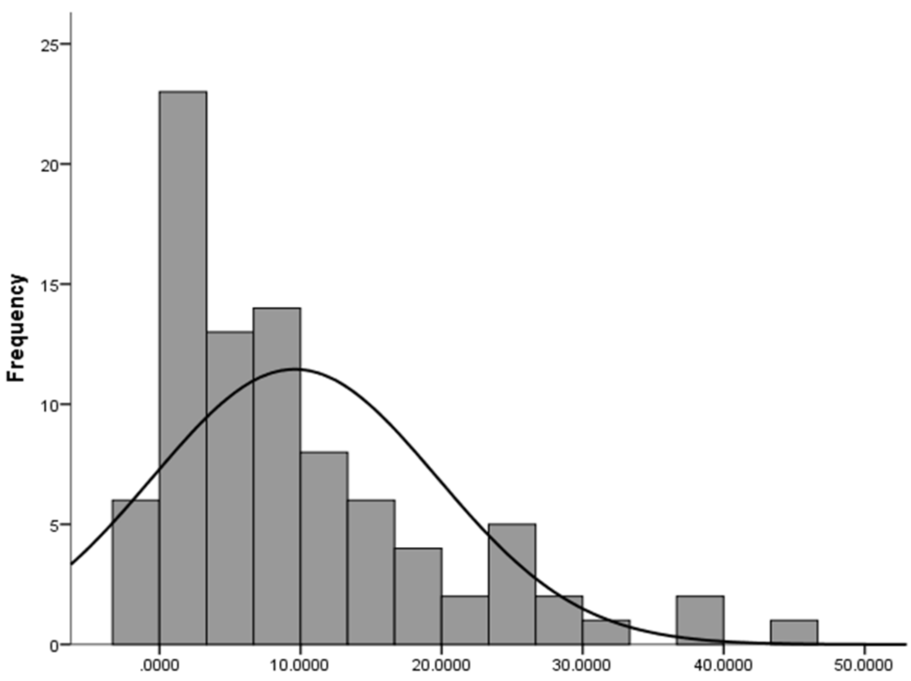
	1	2	3	4	5	6	7	8	9	10	11	12
Bowl	1		2	1	5	3	11	3	10	3	12	1
Passion	2	5	6	6	8	2	2	8	6	9	7	6
Dawn		4	9	7	9	9	3	7	7	6	8	5
Judgement	5	2	5	8	7	7	4	6	5	11	6	7
Grant			1	3	4	5	7	5	3	5	11	2
Bee		6	7	10	10	10	6	12	1	2	10	10
Plane	6	1	8	9	11	8	10	10	8	10	9	9
County	4	8	11	11	2	11		1	9	7	3	11
Choice			3		1	6	8	2	2	8	4	12
Seed	3	7	10	4	12	1	5	9	12	1	2	8
Wool		3	12	5	6		1	11	11	12	1	4
Meal			4	2	3	4	9	4	4	4	5	3
Total Recall	6	8	12	11	12	11	11	12	12	12	12	12
LTR	5	8	11	11	12	11	11	12	12	12	12	12
STR	1	0	1	0	0	0	0	0	0	0	0	0
LTS	5	8	11	11	12	12	12	12	12	12	12	12
CLTR	4	6	9	9	10	10	11	12	12	12	12	12
Random LTR	1	2	2	2	2	1	0	0	0	0	0	0
Presentations	12	6	4	0	1	0	1	1	0	0	0	0

LTR = LONG TERM RETRIEVAL
STR = SHORT TERM RECALL
LTS = LONG TERM STORAGE
CLTR = CONSISTENT LONG TERM RETRIEVAL
RANDOM LTR = RANDOM LONG TERM RETRIEVAL

Figure 5. Sample SRT Protocol using Levin et al. (1986) Form 1

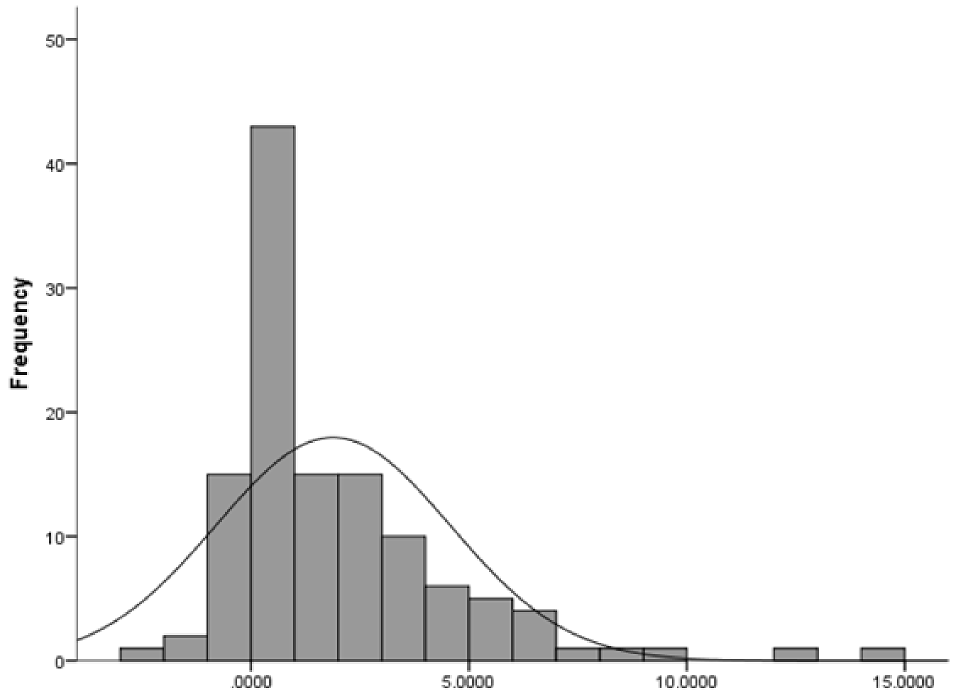


Distribution of PF total scores at three months post injury

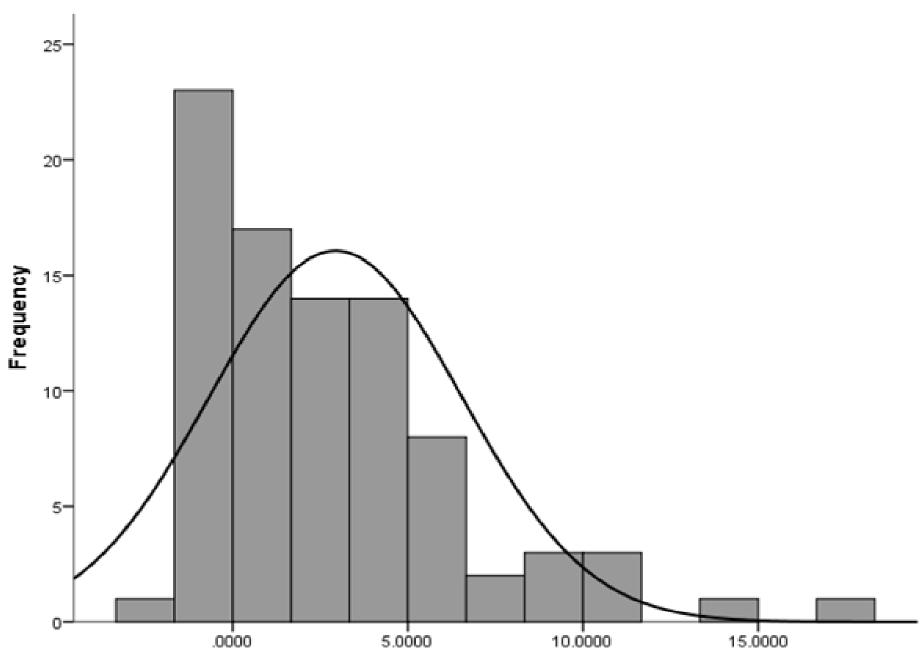


Distribution of PF total scores at six months post injury

Figure 6. Comparison of the distributions of total PF scores at three and six months post injury



Distribution of five trial PF scores at three months post injury



Distribution of five trial PF scores at six months post injury

Figure 7. Comparison of the distributions of five trial PF scores at three and six months post injury

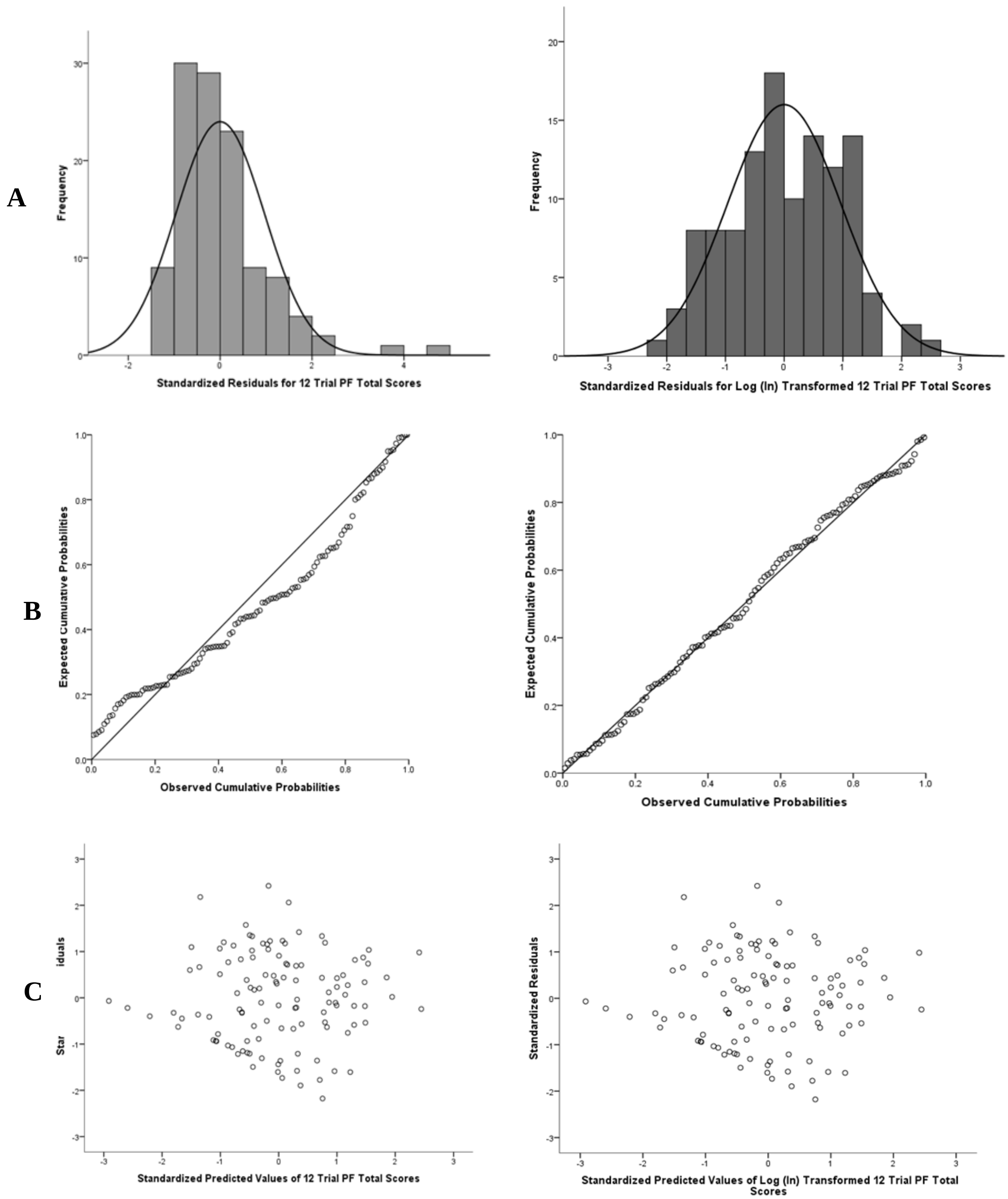


Figure 8. Comparison of residual statistics on initial hierarchical regression of raw 12 trial PF scores to log (ln) Transformed scores at three months post injury. A) Distribution of regression standardized residuals from ln PF scores. B) Normal probability plots for standardized residuals. C) Scatterplot of predicted lnPF scores vs. standardized residuals.

A

B

C

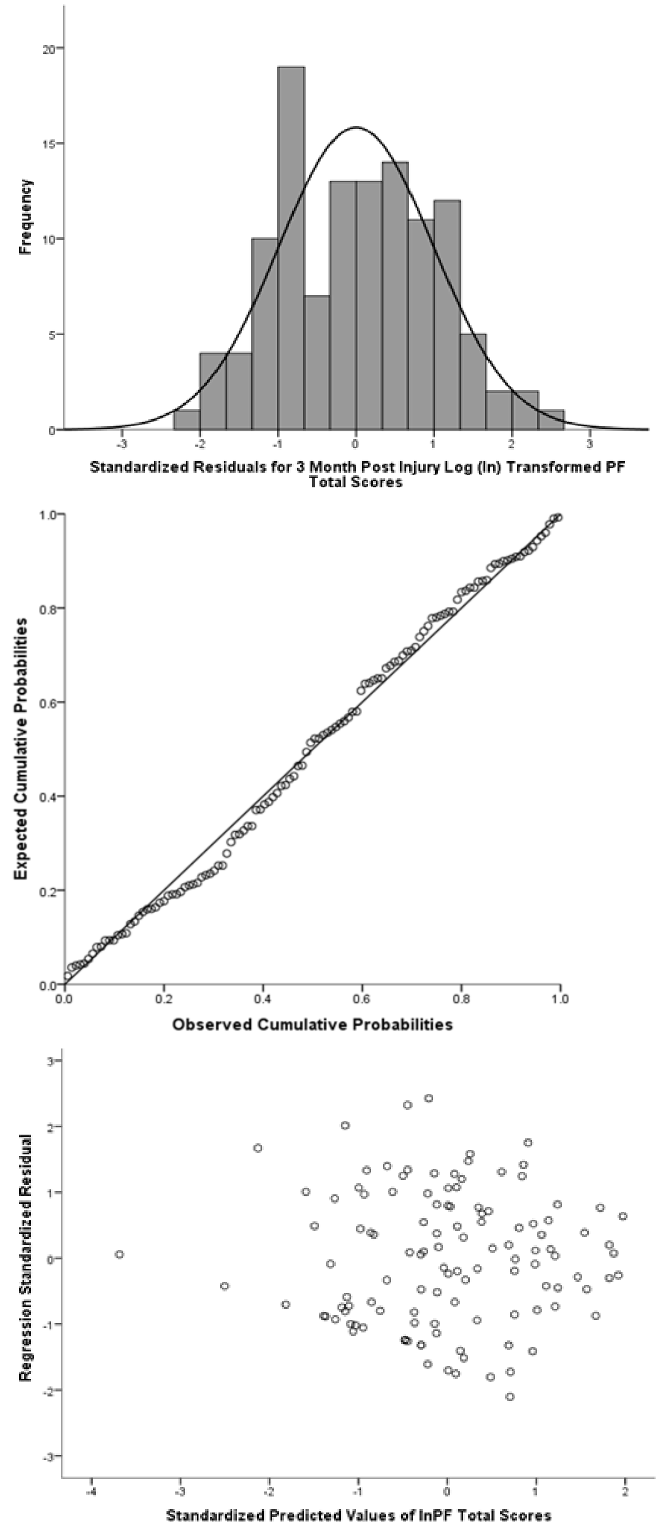


Figure 9. Residuals for final regression model of prediction of lnPF total scores at three months post injury. A) Distribution of regression standardized residuals. B) Normal probability plot for observed vs. expected residuals. C) Plot of standardize predicted values of lnPF scores vs. standardized residuals.

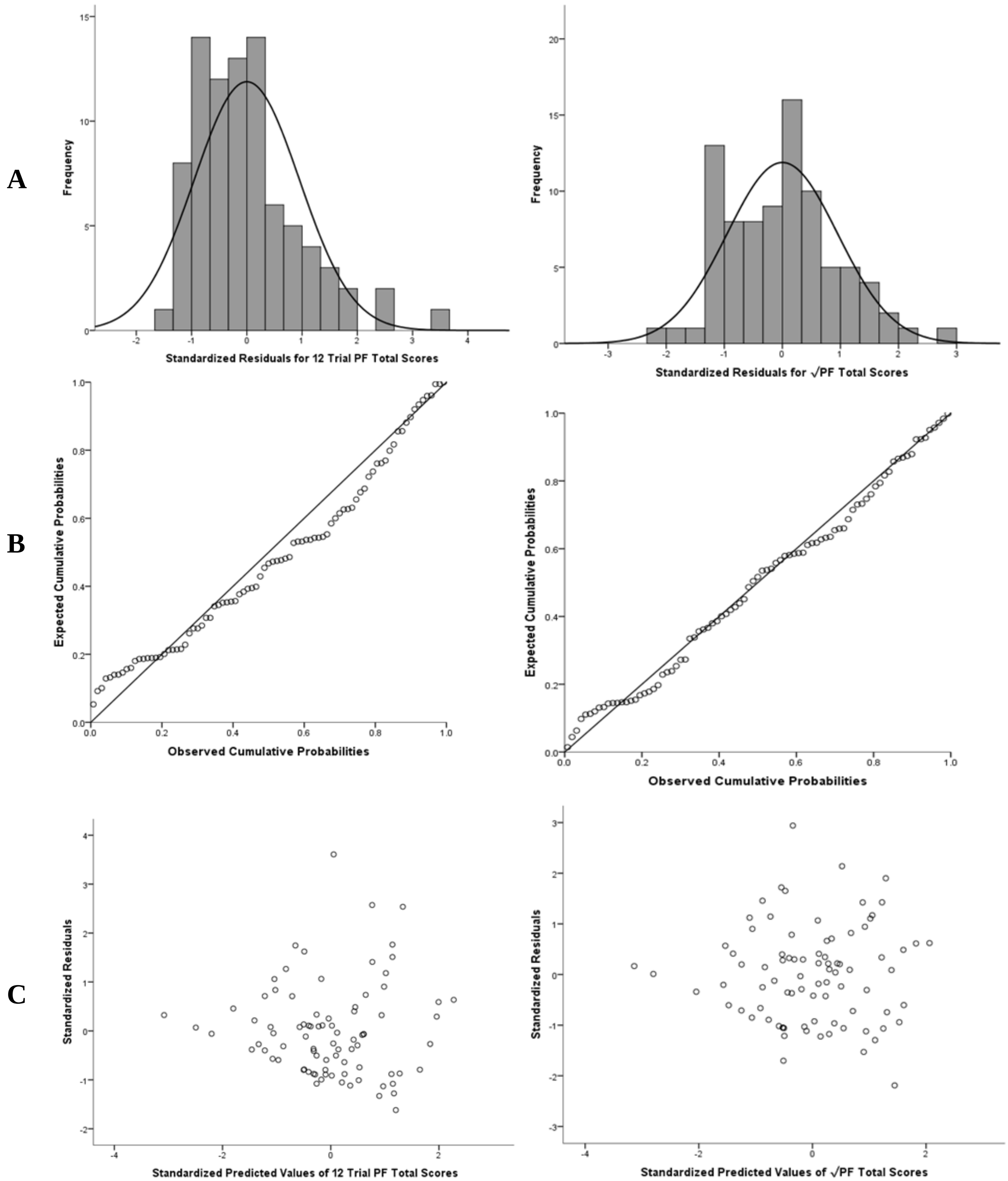


Figure 10. Comparison of residual statistics on initial hierarchical regression of raw 12 trial PF scores to square root transformed scores at six months post injury. A) Distribution of regression standardized residuals from√ PF scores. B) Normal probability plots for standardized residuals. C) Scatterplot of predicted √ PF scores vs. standardized residuals.

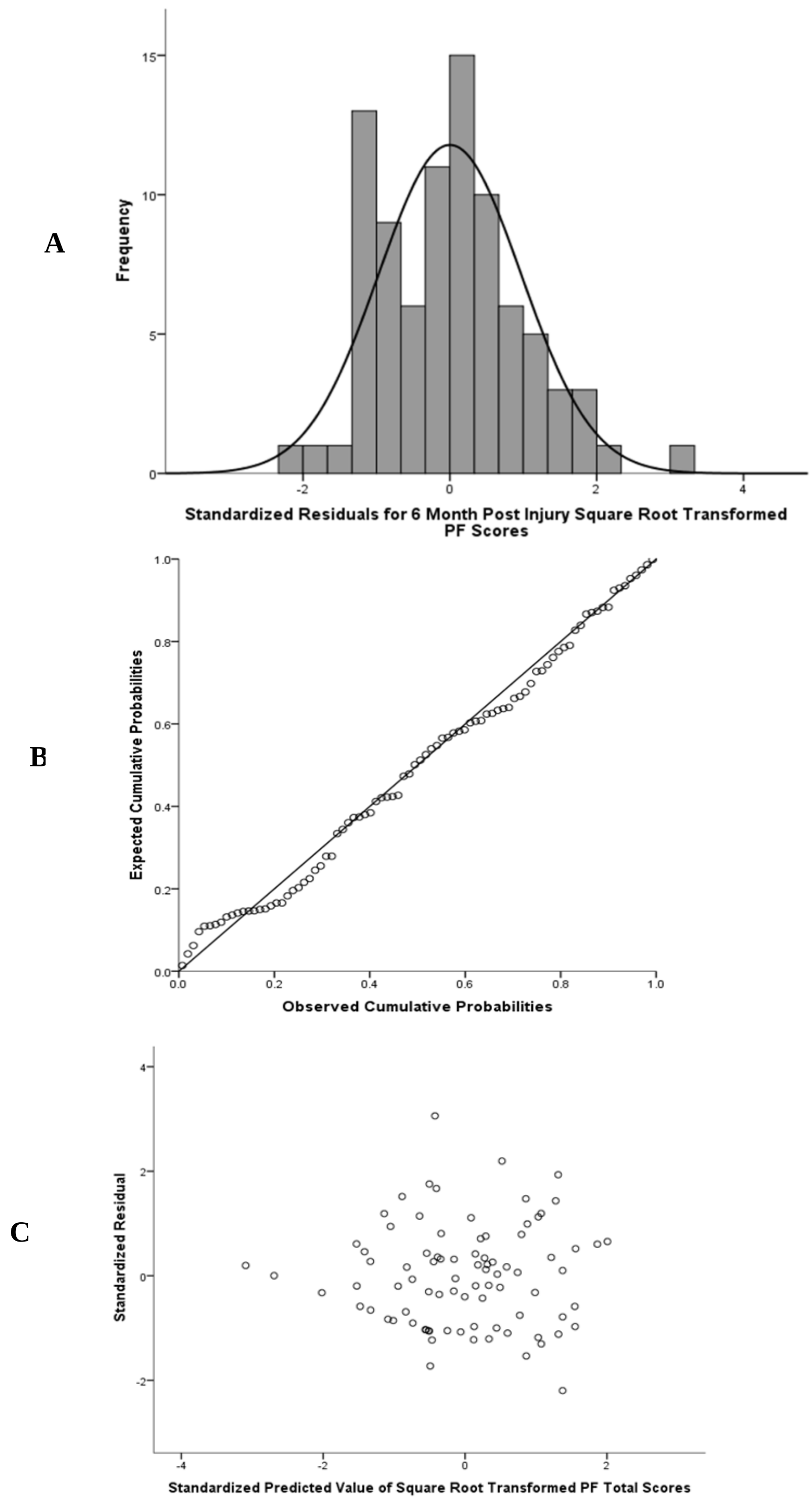


Figure 11. Residuals for final regression model of prediction of √PF total scores at six months post injury. A) Distribution of regression standardized residuals. B) Normal probability plot for observed vs. expected residuals. C) Plot of standardize predicted values of √PF scores vs. standardized residuals.

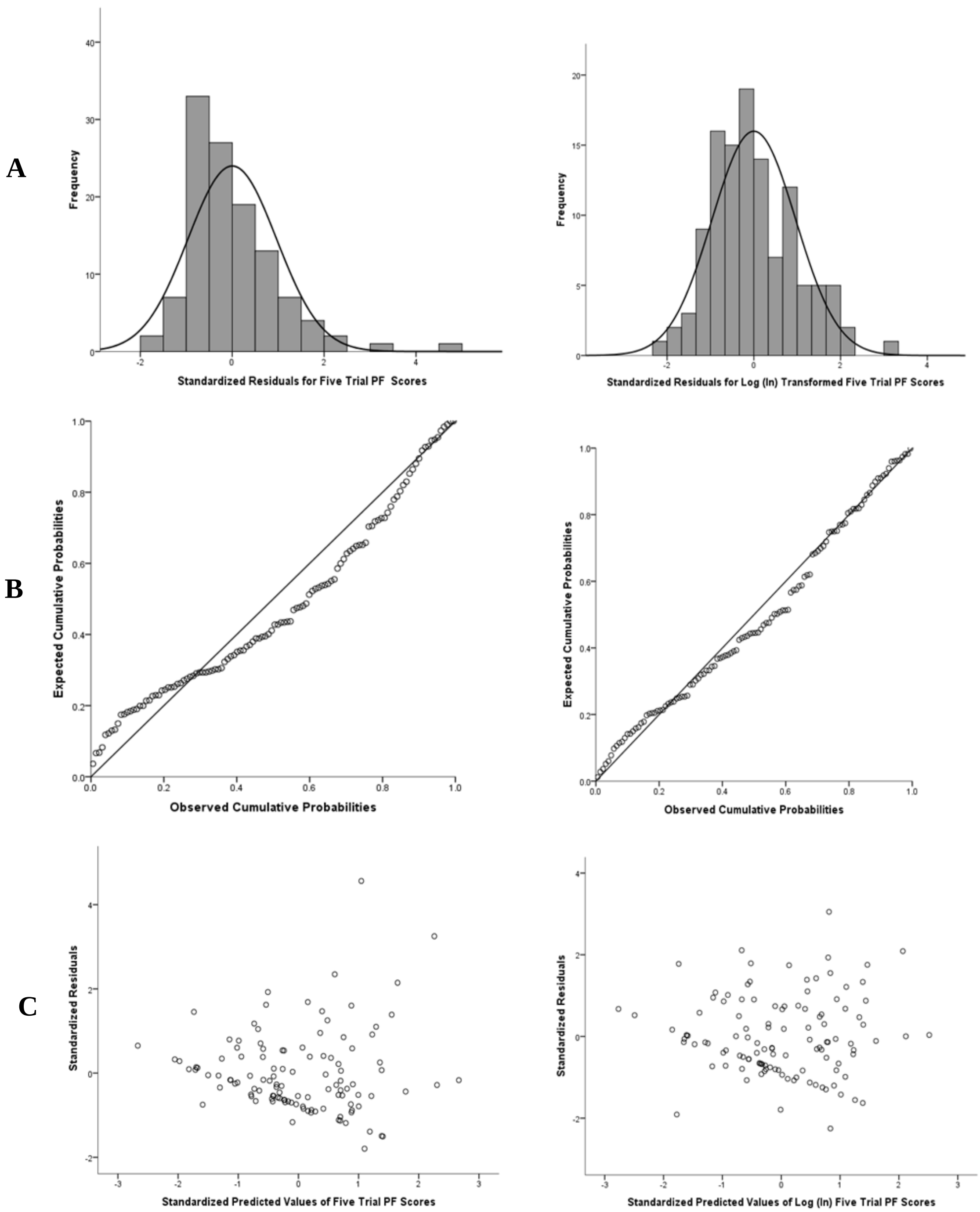


Figure 12. Comparison of residual statistics on initial hierarchical regression of raw 5 trial PF scores and 5 trial lnPF scores at three months post injury. A) Distribution of regression standardized residuals from 5 trial lnPF total scores. B) Normal probability plots for standardized residuals. C) Scatterplot of predicted 5 trial lnPF total scores vs. standardized residuals.

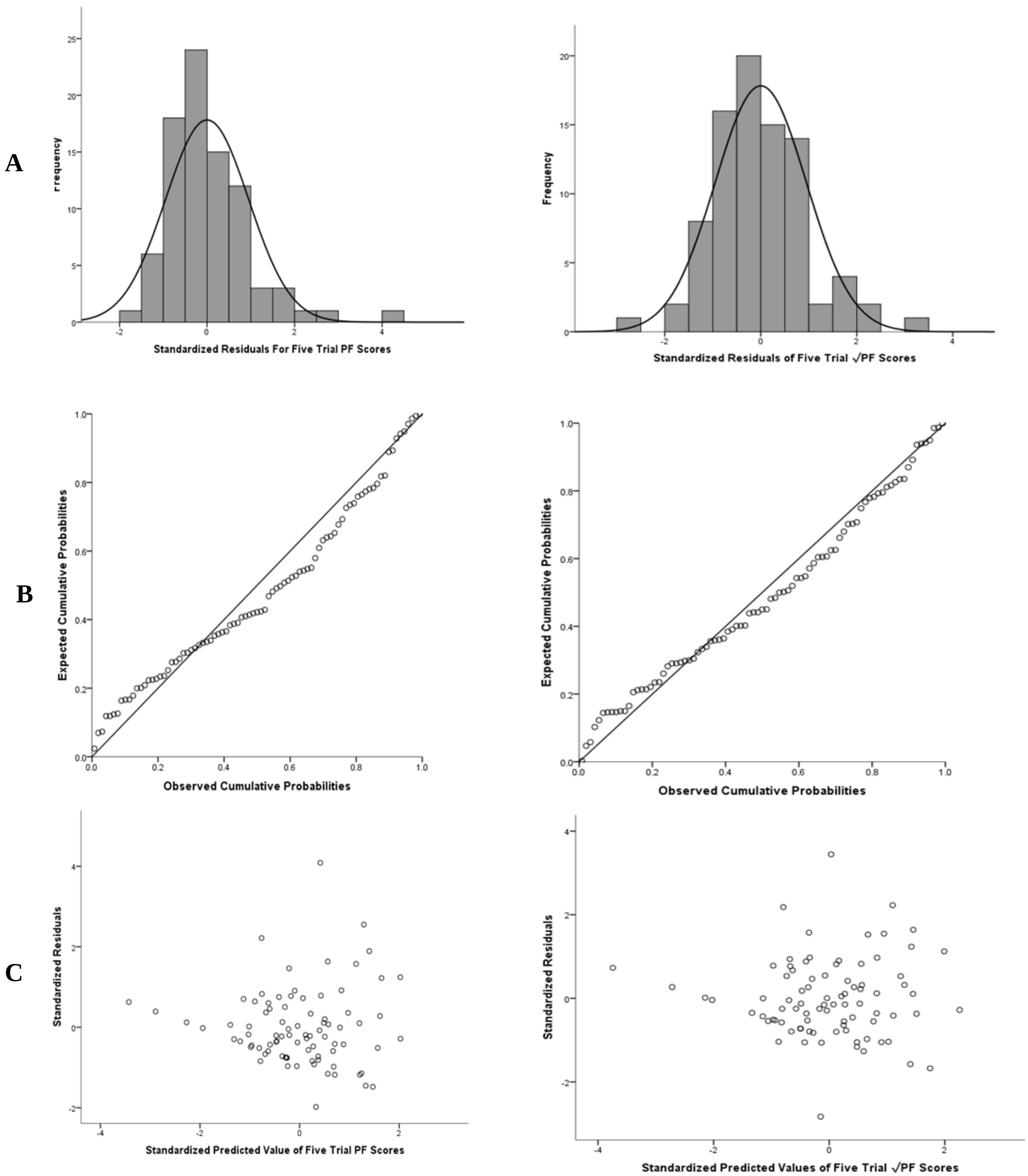


Figure 13. Comparison of residual statistics on initial hierarchical regression of raw 5 trial PF scores and 5 trial √PF scores at six months post injury. A) Distribution of regression standardized residuals from 5 trial √PF total scores. B) Normal probability plots for standardized residuals. C) Scatterplot of predicted 5 trial √PF total scores vs. standardized residuals.

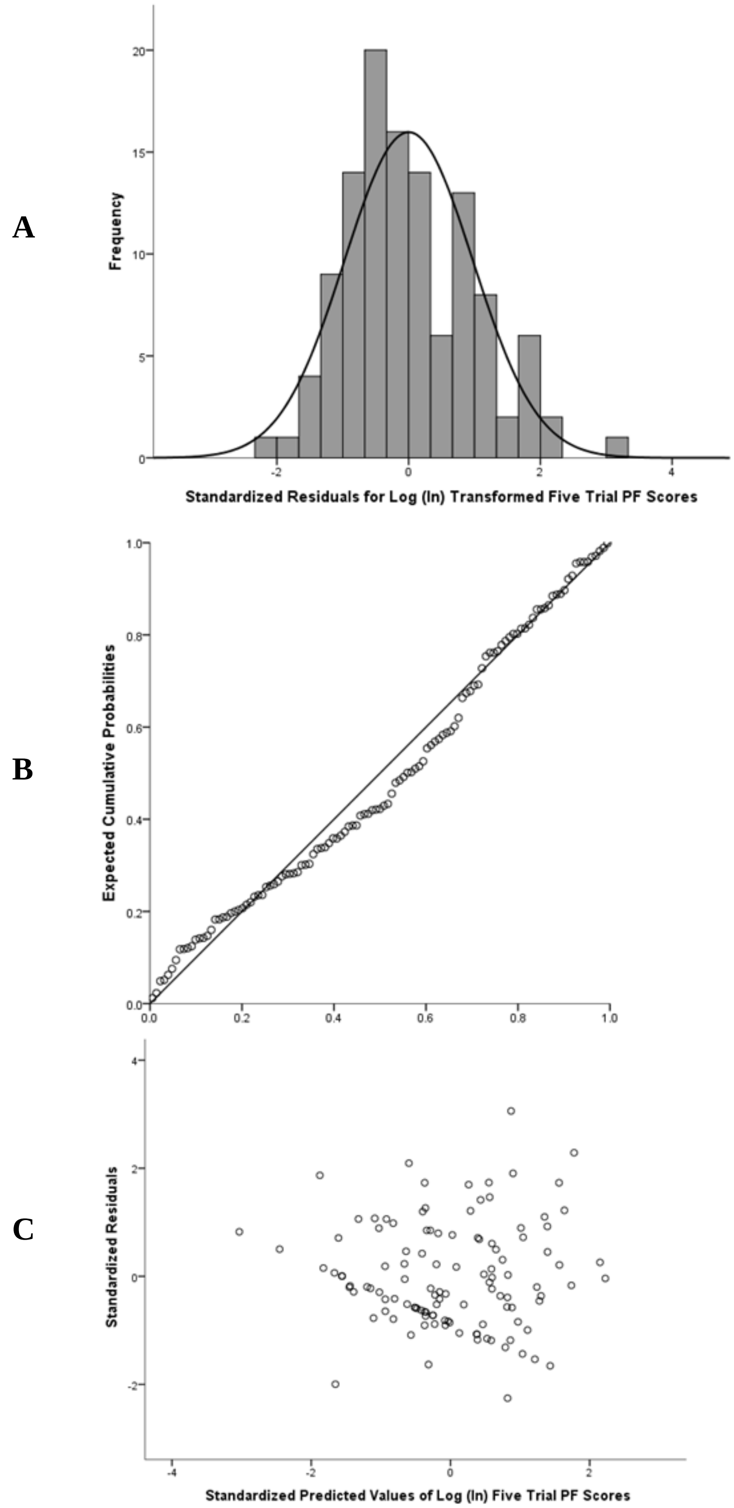


Figure 14. Residuals for final regression model of prediction of five trial lnPF total scores at three months post injury. A) Distribution of regression standardized residuals. B) Normal probability plot for observed vs. expected residuals. C) Plot of standardized predicted values of five trial lnPF scores vs. standardized residuals.

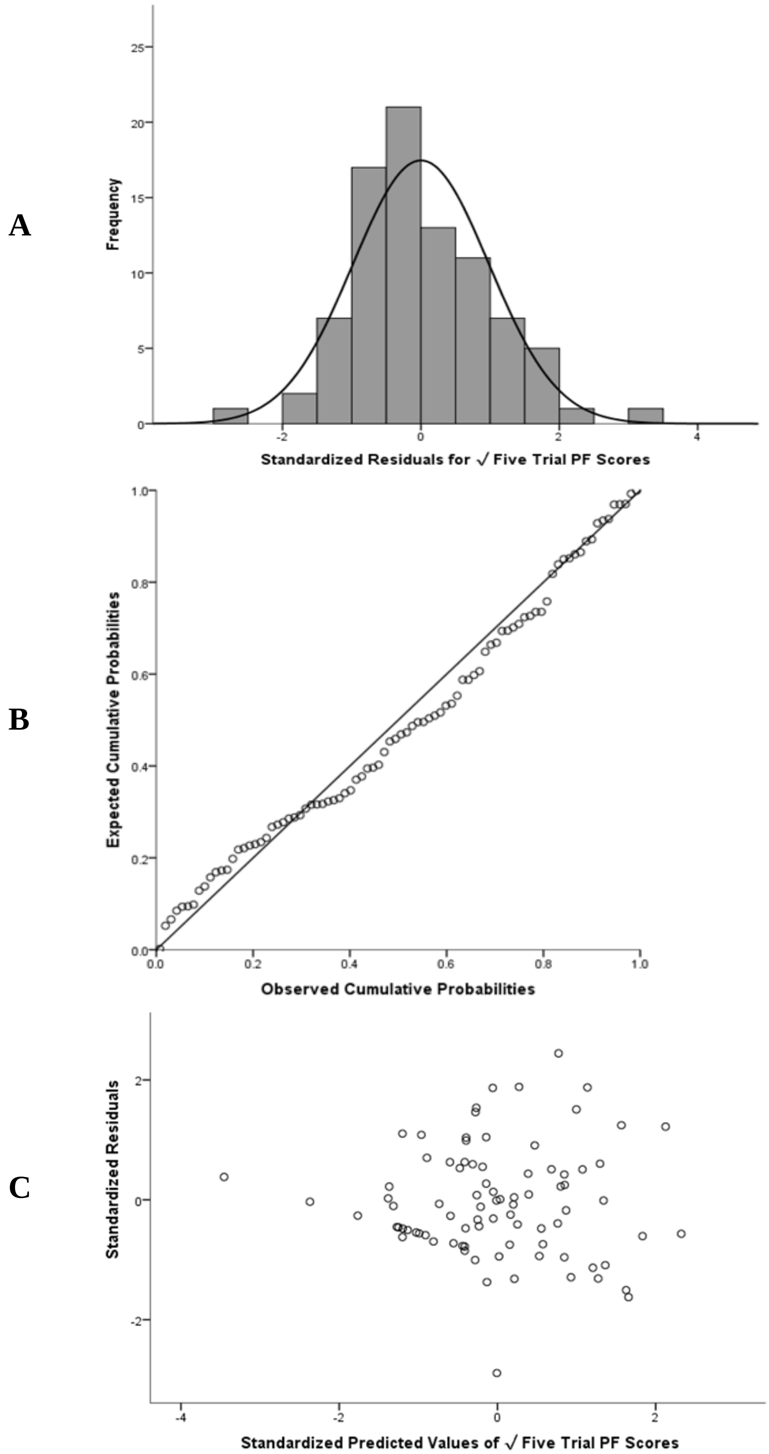


Figure 15. Residuals for final regression model of prediction of five trial √PF total scores at six months post injury. A) Distribution of regression standardized residuals. B) Normal probability plot for observed vs. expected residuals. C) Plot of standardized predicted values of five trial √PF scores vs. standardized residuals.

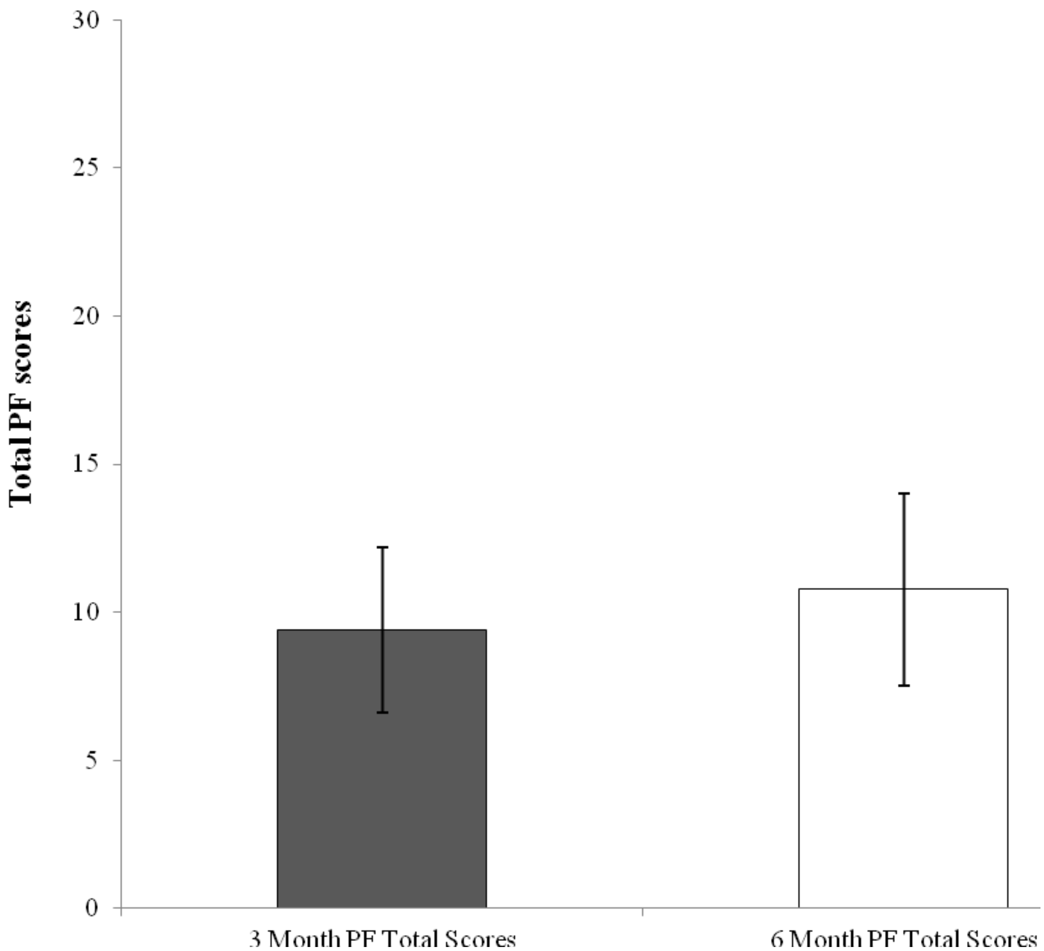


Figure 16. Comparison of average three and six month total pair frequency scores in a subsample of patients tested at both three and six months post injury. Averages are displayed with age and Best Day 1 controlled for as covariates (age evaluated at 31.92 years and Best Day 1 GCS evaluated at 9.35). Error bars represent 95% confidence intervals around the mean.

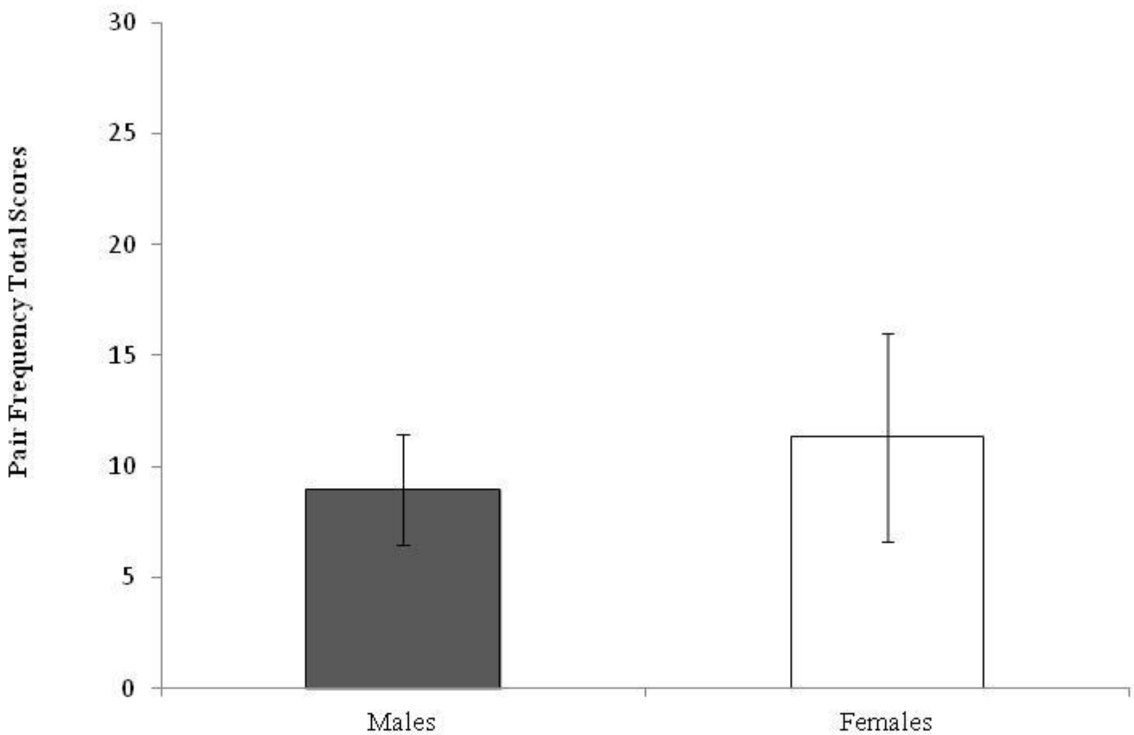


Figure 17. Sex differences in pair frequency total scores in a sample of patients tested at both three and six months post injury. Average scores are displayed with age and Best Day 1 GCS score controlled for as covariates (age evaluated at 31.92 years and Best Day 1 GCS evaluated at 9.35). Error bars represent 95% confidence intervals around the mean.

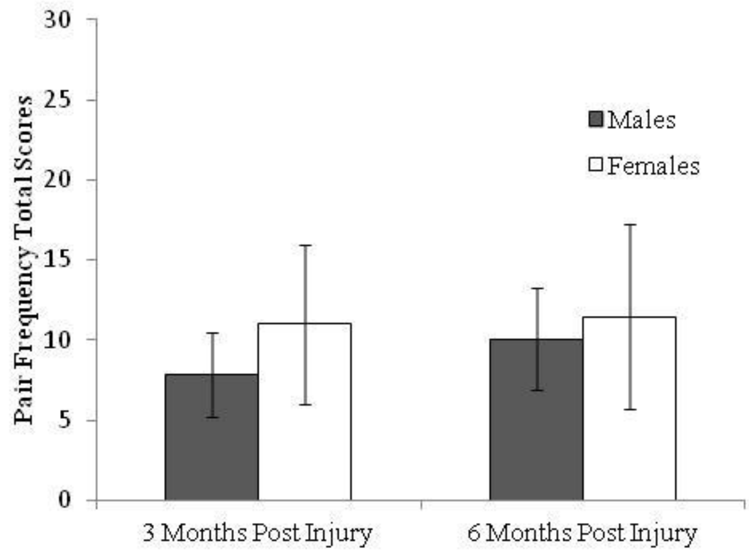


Figure 18. Sex differences in pair frequency total scores at three and six months post injury. Means are evaluated with age and Best Day 1 GCS total score entered as a covariates (age evaluated at 31.92 years and Best Day 1 GCS evaluated at 9.35). Error bars represent 95% confidence intervals around the mean.

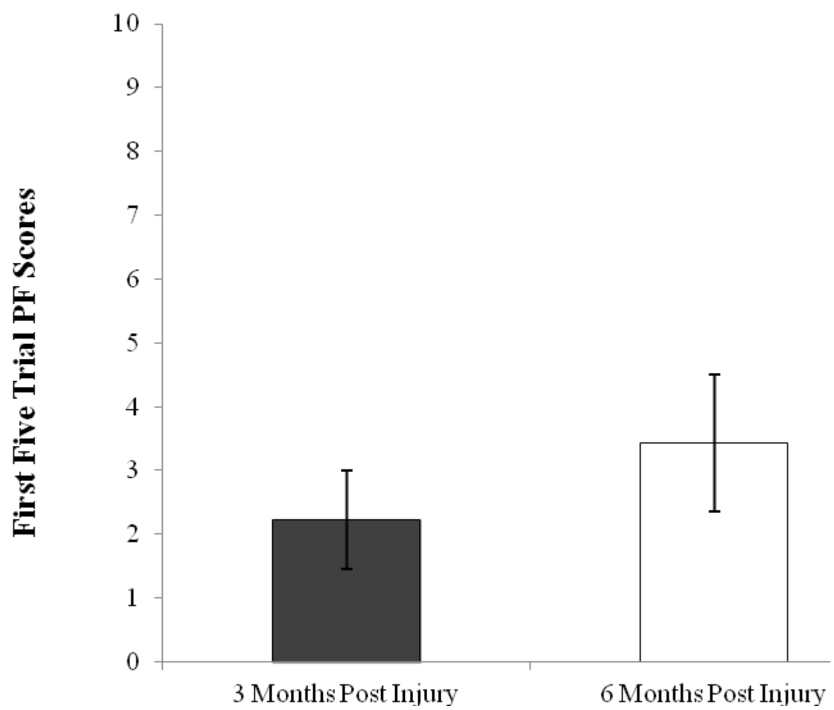


Figure 19. Comparison of average three and six month five trial pair frequency scores in a subsample of patients tested at both three and six months post injury. Averages are displayed with age and Best Day 1 controlled for as covariates (age evaluated at 31.92 years and Best Day 1 GCS evaluated at 9.35). Error bars represent 95% confidence intervals around the mean.

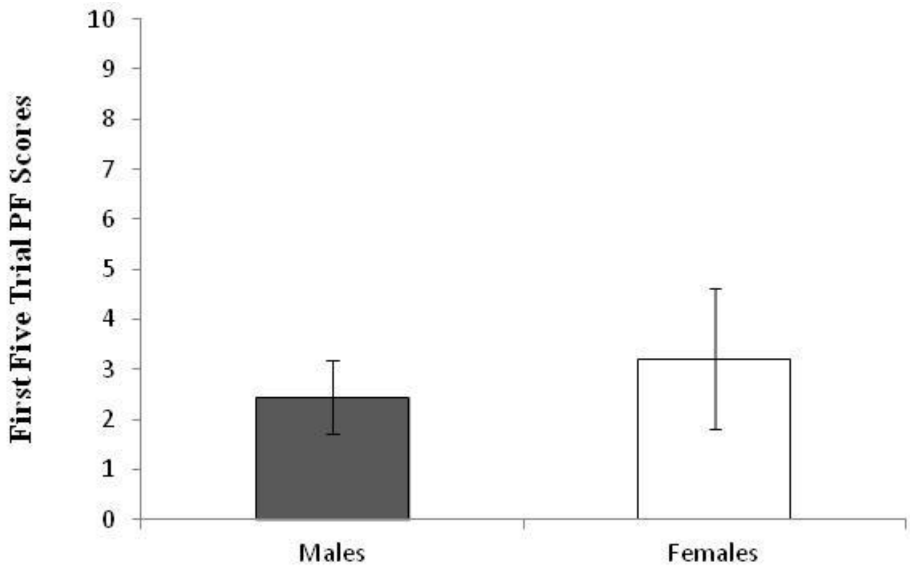


Figure 20. Sex differences in five trial pair frequency scores in a sample of patients tested at both three and six months post injury. Average scores are displayed with age and Best Day 1 controlled for as covariates (age evaluated at 31.92 years and Best Day 1 GCS evaluated at 9.35). Error bars represent 95% confidence intervals around the mean.

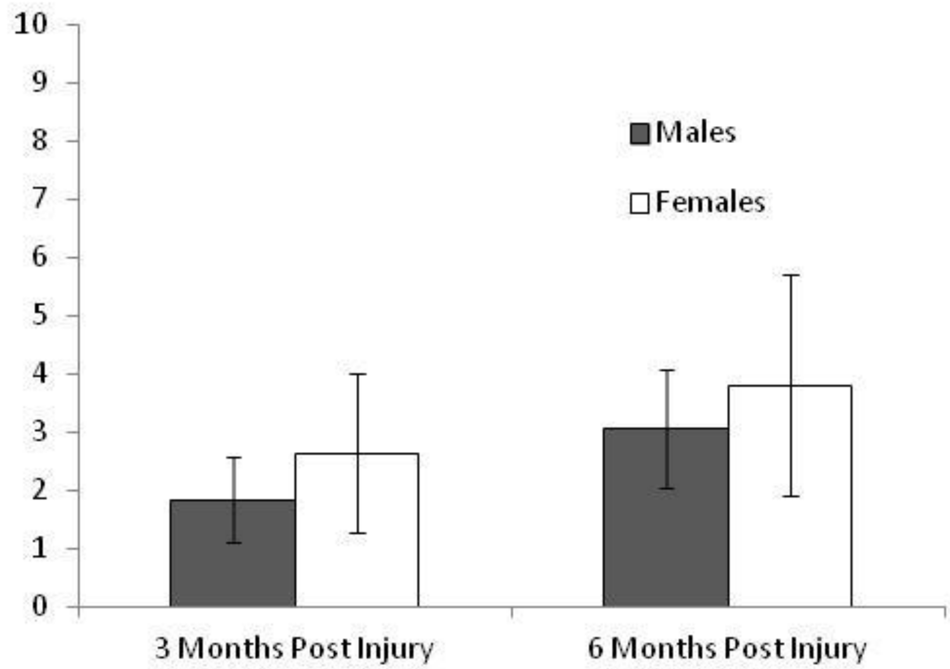


Figure 21. Sex differences in five trial pair frequency scores at three and six months post injury. Means are evaluated with age and Best Day 1 GCS total score entered as a covariates (age evaluated at 31.92 years and Best Day 1 GCS evaluated at 9.35). Error bars represent 95% confidence intervals around the mean.