

Lateralizing Memory Function in Temporal Lobe Epilepsy:
An Investigation of the Meaning and Utility of the
Wechsler Memory Scale – Third Edition

by

Nancy Jean Wilde
B.Sc., McGill University, 1996
M.Sc., University of Victoria, 1999

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Psychology

© Nancy Jean Wilde, 2005
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by
photocopy or other means, without the permission of the author.

Supervisor: Dr. Esther H. Strauss

ABSTRACT

The Wechsler Memory Scale (WMS) is the most extensively used battery for memory assessment of adults. The third edition of the WMS (WMS-III) represents a substantial revision of previous versions. Accordingly, issues of validity of the revised instrument need to be addressed. The purpose of these studies was to contribute to the validation of the scale in the assessment of patients with temporal lobe epilepsy (TLE). An important role of the neuropsychological evaluation in TLE is to aid in the localization and lateralization of dysfunction. This is based on the premise that the temporal lobes are specialized for the acquisition of material-specific information, with dysfunction in the left and right mesial temporal regions being associated with verbal and nonverbal memory impairment, respectively. Since the WMS is utilized by the vast majority of epilepsy centres, evaluation of its meaning and utility in this population is essential.

In Study 1, the utility of the WMS-III in detecting lateralized impairment was examined in a sample of patients with left ($n = 55$) or right ($n = 47$) TLE. Methods of analysis included evaluation of group means on the various indexes and subtest scores, the use of ROC curves, and an examination of Auditory-Visual Index discrepancy scores. The Auditory-Visual Delayed Index difference score appeared most sensitive to side of temporal dysfunction, although patient classification rates were not within an acceptable range to have clinical utility. The ability to predict laterality based on statistically significant index score differences was particularly weak for those with left temporal dysfunction. The use of unusually large discrepancies led to improved prediction; however, the rarity of such scores limits their usefulness.

In Study 2, five competing models specifying the factor structure underlying the WMS-III primary subtest scores were evaluated in a large sample of patients with TLE (N = 254). Models specifying separate immediate and delayed constructs resulted in inadmissible parameter estimates and model specification error. There were negligible goodness-of-fit differences between a 3-factor model of working memory, auditory memory, and visual memory, and a nested—more parsimonious—2-factor model of working memory and general memory. The results suggested that specifying a separate visual memory factor provided little advantage for this sample— an unexpected finding in a population with lateralized dysfunction, for which one might have predicted separate auditory and visual memory dimensions.

These findings add to a growing literature which suggests that the WMS-III has little utility in detecting lateralized dysfunction in TLE. This has important implications for the preoperative assessment of epilepsy patients.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iv
List of Tables	vi
List of Figures.....	vii
CHAPTER 1	
General Introduction	1
The Wechsler Memory Scales	1
Test Validation	5
Criterion Validation & Diagnostic Utility of the WMS-III in Determining Laterality of Dysfunction (Study 1).....	6
Construct Validation of the WMS-III in Temporal Lobe Epilepsy Using Factorial Methods (Study 2)	8
CHAPTER 2	
Study 1: The Utility of the WMS-III in Differentiating Lateralized Temporal Epileptogenic Dysfunction	11
Method	16
Participants	16
Procedures	18
Statistical Analysis.....	18
Results.....	19
Sample Characteristics	19
Group Differences Among the Primary Indexes and Subtest Scores.....	20
Auditory - Visual Index Discrepancy Comparison.....	20
Comparison of Immediate and Delayed Index Scores	22
Receiver Operating Characteristic Curves	23
Patient Classification Using Significant and Infrequent Index Discrepancies	26
Discussion.....	29

CHAPTER 3

Study 2: Confirmatory Factor Analysis of the WMS-III in Patients with Temporal Lobe Epilepsy.....	36
Method	40
Participants	40
Procedures	42
Statistical Analysis.....	43
Results.....	46
Discussion.....	50

CHAPTER 4

General Discussion	56
Additional Studies on the WMS-III in TLE	58
Visual Memory and the Right Temporal Lobe.....	61
The Utility of Separate Immediate and Delayed Indexes	67
Factor Structure of the WMS-III	68
Neuropsychological Testing, Lateralization, and Material-Specific Memory Outcome.....	70
References	73

LIST OF TABLES

Table 1.1. WMS-III Subtest, Index, and Composite Configuration.....	4
Table 2.1. Sample Characteristics for Each Center and the Total Sample	17
Table 2.2. Mean WMS-III Scores for the Right and Left Temporal Lobe Epilepsy Groups.	21
Table 2.3. Immediate-Delayed Index Score Differences	23
Table 2.4. ROC Curve Statistics for WMS-III Index and Subtest Scores	25
Table 2.5. Auditory - Visual Index Difference Scores and Classifications	28
Table 3.1. Demographic and Clinical Epilepsy Characteristics for the Total Sample and Individual Centers	42
Table 3.2. Mean WMS-III Age-Corrected Subtest Scores	44
Table 3.3. Intercorrelations Between WMS-III Age-Corrected Subtest Scores	44
Table 3.4. Goodness-of-Fit Statistics	47
Table 3.5. Standardized Parameter Estimates for Model 2.....	48

LIST OF FIGURES

- Figure 2.1. Mean z scores for the RTLE (right temporal lobe epilepsy) and LTLE (left temporal lobe epilepsy) groups on the WMS-III indexes and individual subtests. Note that better performance is represented by z values closer to the normative mean of 0. Square markers denote index scores and circle markers denote subtest scores.22
- Figure 2.2. Receiver operating characteristic (ROC) curve for the WMS-III Auditory-Visual Delayed Index difference score. AUC = Area under the curve.....24

Chapter 1

GENERAL INTRODUCTION

The purpose of these studies was to contribute to the validation of the newest version of the Wechsler Memory Scale, one of the most widely used measures of memory in clinical practice. Two studies were conducted. The first examined the utility of the scale in differentiating groups with lateralized brain dysfunction; specifically – temporal lobe epilepsy. The second study investigated issues of construct validity by using confirmatory factor analysis to evaluate the latent variable structure of the test in this same population.

This Introduction is divided into two sections. The first provides a brief overview of the development and structure of the Wechsler Memory Scales. The second section considers general issues in test validation, and outlines the ways in which each study contributes to the validation of the scale. Subsequent chapters describe the individual studies in more detail. Note that both studies have been published in peer-refereed journals and are reproduced in this document in their original format, with minor alteration in wording to reduce redundancy and improve readability. Relevant research conducted following the publication of these studies is commented on in the General Discussion.

The Wechsler Memory Scales

The Wechsler Memory Scale (WMS; Wechsler, 1945) was one of the first standardized memory tests (Franzen, 1989), and along with its revisions, is the most frequently used clinical measure of memory (Erickson & Scott, 1977; Lees-Haley, Smith, Williams, & Dunn, 1996; Rabin, Barr, & Burton, 2005). The original WMS consisted of seven subtests, designed to measure different aspects of memory. The various subtests assessed basic orientation to

person, place, and time, passage narrative recall, verbal paired associate learning with repetition, the reproduction of visual figures after a brief exposure, forward and backward digit recall, and the ability to perform various mental tasks under time pressure (e.g., recite the alphabet, count backwards). The WMS was weighted substantially toward the assessment of verbal memory (there was only one visual subtest). In addition, no tests of memory with interference or delayed memory were included (Franzen, 1989; Spreen & Strauss, 1991).

Extensive criticism (e.g., Prigatano, 1977; 1978) of the WMS was directed at the inadequacy of the standardization sample (which consisted of 200 normal subjects who ranged in age from 25 to 50), the fact that scores were combined into a single summary score, the over reliance on immediate recall and verbal tasks, and the inclusion of tasks which were not felt to be genuine measures of memory (Franzen, 1989).

The first revision of the Wechsler Memory Scale was published in 1987 (WMS-R; Wechsler, 1987) and addressed a number of the limitations of the WMS. Norms stratified at nine age levels were provided, five composite scores replaced the single global summary score of the WMS, tests aimed at evaluating visual memory and delayed recall were included, and the scoring procedures were revised to improve scoring accuracy (Wechsler, 1987). While the WMS-R represented an improvement over the original version, a number of problems with the revised test were identified (Bornstein & Chelune, 1988; Elwood, 1991; Loring, 1989). As with the WMS, criticisms were aimed at the normative data (i.e., norms for certain age groups were interpolated, no normative data for subjects older than 74 years were provided) and the weighting was biased towards verbal memory measures. Other limitations were directed at the inadequacy of the “nonverbal” memory subtests, the lack of recognition measures, the high floor for the index scores, and the low reliabilities of the subtests and indexes (Franzen, 2000; Spreen & Strauss, 1998).

The most recent version of the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) states that “a test should be amended or revised when new research data, significant changes in the domain represented, or newly recommended conditions of test use may lower the validity of test score interpretations” (p. 48). The third edition of the Wechsler Memory Scale (WMS-III; Wechsler, 1997b) is “meant to be more than a revision of the WMS-R. It is meant to reflect a contemporary conceptualization of memory and its disorders” (Franzen, 2000, p. 225). In the development of the WMS-III, there were substantial revisions of the WMS-R subtests, administration and scoring procedures, and index configurations (The Psychological Corporation, 1997; Tulskey & Ledbetter, 2000; Wechsler, 1997b). The standardization sample was increased to 1250 adults between the ages of 16 and 89 years. Subtests were individually normed and are presented as scaled scores, and there are more specific index scores. The scale was co-normed with the Wechsler Adult Intelligence Scale - Third Edition (WAIS-III; Wechsler, 1997a) to allow better comparison of intellectual functioning and memory performance. In response to criticism of the WMS-R visual memory subtests (Chelune & Bornstein, 1988; Heilbronner, 1992; Loring, 1989), major changes have occurred to the visual memory stimuli. Two out of three WMS-R visual memory subtests were deleted from the scale, and the third was turned into an optional subtest. Two new visual subtests were introduced and are now the core visual subtests for interpretation.

The Primary Indexes, meant to be the main interpretive focus on the WMS-III, have increased from five in the WMS-R to eight in the WMS-III. A Working Memory index, composed of two new subtests, was introduced. Delayed recognition tasks were included for comparison with performance on the delayed recall tasks. The method of calculating several WMS-III index scores also differs significantly from that used in the WMS-R in that scaled

scores, rather than raw scores, are summed to ensure equal weighting of the component scores. A summary of the organization of the WMS-III is shown in Table 1.1.

Table 1.1
WMS-III Subtest, Index, and Composite Configuration

Primary Subtests and Indexes			
<i>Indexes</i>		<i>Subtests</i>	<i>Changes from WMS-R</i>
Immediate Memory	Auditory Immediate	Logical Memory I Verbal Paired Associates I	Revised from WMS-R
	Visual Immediate	Faces I Family Pictures I	New to WMS-III
General Memory (Delayed)	Auditory Delayed	Logical Memory II Verbal Paired Associates II	Revised from WMS-R
	Visual Delayed	Faces II Family Pictures II	New to WMS-III
	Auditory Recognition Delayed	Logical Memory II Verbal Paired Associates II	New to WMS-III
Working Memory		Letter-Number Sequencing Spatial Span	New to WMS-III
Optional Composites			
Auditory Process Composites	Single-Trial Learning		
	Learning Slope	Logical Memory I & II	Composites new to WMS-III
	Retention Retrieval	Verbal Paired Associates I & II	
Optional Subtests			
		Information & Orientation	Same as WMS-R
		Word Lists I & II	New to WMS-III
		Visual Reproduction I & II	Revised from WMS-R
		Mental Control	Revised from WMS-R
		Digit Span	Revised from WMS-R

In sum, there have been substantial revisions to both the content and the index structure of the WMS-III. Changes are particularly dramatic for the visual memory stimuli: the correlations between the WMS-R and the WMS-III visual memory subtests are low, perhaps understandably, given the significant change in subtest content (Tulsky & Ledbetter,

2000). “The practical effect for the clinician is that WMS–III Visual Memory Index may perform dramatically different as compared with the WMS–R” (Tulsky & Ledbetter, 2000, p. 259). The Working Memory Index and its composite subtests are also new to this edition and thus relatively unstudied. Even the auditory subtests of the WMS–R, which have been retained in the WMS–III, have undergone content revision and changes in administration and scoring procedures.

When a test is revised, whether to update the norms, to encompass new concepts and research developments, because the psychometric properties are deemed unsatisfactory, or due to inadequate construct representation (Reise, Waller, & Comrey, 2000; Strauss, Spreen, & Hunter, 2000), issues of validity of the revised instrument need to be addressed. Given the extent of changes between the WMS–R and the WMS–III, a thorough examination of the new scale is essential to establish its validity in various contexts, and to facilitate its clinical interpretation.

Test Validation

The *Standards for Educational and Psychological Testing* (AERA et al., 1999) state:

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretation of test scores required by proposed uses that are evaluated, not the test itself. When test scores are used or interpreted in more than one way, each intended interpretation must be validated. (p. 9)

This excerpt highlights several important points on test validation. First, validity considerations are of utmost importance in evaluating the utility of a test. Second, tests themselves are not valid or invalid, but rather it is the inferences drawn from test results that can be described as either valid or invalid (Franzen, 2000). Finally, the process of validation

is an empirical one, which may be described operationally as a statistical relationship between the results of a particular procedure and other independently observed events (Anastasi & Urbina, 1997; Franzen, 2000; Nunnally, 1978). These relationships may be defined in terms of the content of a test, related criteria, and underlying constructs.

Historically, validity has been composed of three main entities: content validity, criterion-related (composing both predictive and concurrent) validity, and construct validity (Cronbach & Meehl, 1955; Nunnally, 1978). Throughout the years, many additional types of validity have been proposed, such as diagnostic, ecological, factorial, or social validity. Current thought tends to view construct validity as the fundamental and all-inclusive validity concept, insofar as it specifies what the test measures and provides the evidential basis for score interpretation (Anastasi & Urbina, 1997; Messick, 1995). Content analysis, predictive relationships, and factorial validation procedures are among the many sources of information that contribute to the definition and understanding of the constructs of the test (Anastasi & Urbina, 1997; Franzen, 2000). Thus, rather than referring to distinct kinds of validity, the 1999 *Standards* refers to types of validity evidence to reinforce the notion that validity is a unitary concept.

The first study made use of test-criterion methods to evaluate the validity of the scale in reference to the criterion of lateralized brain dysfunction. The second study utilized factorial validation procedures to investigate the validity of the scale's proposed underlying structure in this population.

Criterion Validation & Diagnostic Utility of the WMS-III in Determining Laterality of Dysfunction
(Study 1)

Criterion-validating procedures demonstrate a test's effectiveness in a given context (Anastasi & Urbina, 1997). Correlations between test scores and criterion measures

contribute to the joint construct validity of both predictor and criterion. Empirical relationships between predictor scores and criterion measures should make theoretical sense in terms of what the predictor test is interpreted to measure and what the criterion is presumed to embody (Gulliksen, 1950; Messick, 1995). The method of contrasted groups (Anastasi & Urbina, 1997) is frequently employed in this context, in which the researcher evaluates the results of testing two groups that are assumed to differ on the criterion of interest (Franzen, 2000). The scores are then compared by some statistical method, traditionally by testing the significance of the difference between the mean scores for the two groups. In clinical neuropsychological research, concurrent neurophysiologic measures (i.e., CT, EEG, MRI, rCBF) provide major sources of criterion validation (Franzen, 2000). In this way, the diagnostic or localization accuracy of predictor scores can be evaluated.

Temporal lobe epilepsy (TLE) is the largest single type of seizure disorder – of the approximately 2% of the general population with epilepsy, 40% to 60% have seizures of temporal lobe origin (Snyder, 1998). Due to neuronal dysfunction in the medial temporal lobe region, explicit memory disturbances are endemic to this population (cf. Jones-Gotman, 1991). Hence, one requires a test that will provide a valid characterization of memory impairment in these individuals. Furthermore, the temporal lobes are specialized for the acquisition of material-specific information. Thus, left-sided dysfunction tends to result in significant disturbances of verbal memory (Milner, 1968a; 1970; 1971), while right-sided temporal dysfunction may produce deficits in nonverbal or visuospatial memory (Kimura, 1963; Milner, 1965; 1968b).

In Study 1, the ability of the WMS-III to differentiate those with lateralized (i.e., left/right) dysfunction was examined via the method of contrasted groups. The purpose of this study was to determine whether the performance characteristics of the scale followed

hypothesized brain-behavior criterion relationships – specifically, if the scale distinguished left and right temporal dysfunction associated with a unilateral seizure onset.

A common use of neuropsychological assessment is to provide a diagnostic impression. While predictor-criterion relationships at the mean level may provide useful information for test validation, their utility for interpreting results in the assessment of a single individual may be limited. Diagnostic validity is concerned with whether a single test, or a combination of tests, can accurately identify or classify individuals with a given diagnosis. In examining this, one goes beyond the method of contrasted groups (Anastasi & Urbina, 1997) and attempts to apply that information to a single case. As stated by Franzen (2000):

The ultimate evaluative-validators demonstration of a test is its clinical utility. Tests survive to the extent that they provide information that is useful in the individual case as well as on the average. The various forms of validation are markers that indicate to a clinician the relative importance of a pattern of performance and the likelihood that a given pattern of performance is diagnostically or predictively contributory to the overall understanding of the patient. (p. 54)

In individuals with temporal lobe epilepsy, the information derived from memory tests is assumed to assist in the localization of neurological dysfunction (Jones-Gotman, Smith, & Zatorre, 1993). Thus, the second purpose of Study 1 was to evaluate the diagnostic validity or classification accuracy of the WMS–III in differentiating those with unilateral temporal seizure disorder.

Construct Validation of the WMS–III in Temporal Lobe Epilepsy Using Factorial Methods (Study 2)

The first detailed exposition on construct validity appeared in 1955 in an article by Cronbach and Meehl: “Construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not ‘operationally defined’. The problem faced by the investigator is, ‘What constructs account for variance in test performance?’” (p. 282).

Of several common ways to investigate the construct validity of a test, the *multitrait-multimethod matrix* (Campbell & Fiske, 1959), which involves systematic experimental design for the dual approach of convergent and divergent validation, has been most frequently discussed. However, this method has been criticized because there are no guidelines for interpreting the size of resulting zero-order correlations, nor the overall structure of the relationships. Cole (1987) instead suggested the use of confirmatory factor analysis (CFA) to analyze the data and investigate discriminant and convergent validity. CFA procedures have been described as representing “the most significant advance in construct validation research since Cronbach and Meehl (1955)” (Fletcher et al., 1996, p. 23). CFA allows for more precise formulation and a statistical test of measurement models, rather than only a description of the relationships among variables. CFA procedures can also control for the effects of correlated errors, and can be used to determine if measurement models are invariant across populations of interest.

In the use of factor analysis for evaluating a proposed model of covariability across independent samples, two basic situations often arise (Reise et al., 2000). In the first, data are collected on a given sample and a researcher wishes to evaluate whether the sample factor structure is consistent with a hypothesized structure – this design is often seen in replication studies, in which samples are drawn from the same population. However, in the same way that the reliability of test scores often depends dramatically on sample variability, an instrument’s factor structure can change depending on the peculiarities of a particular sample (Reise et al., 2000). Thus, in the second situation, data are collected in samples from different populations, and the researcher wants to evaluate whether the factor structures are similar or equivalent between groups (Reise, Widaman, & Pugh, 1993; Reise et al., 2000). This latter situation is frequently referred to as a measurement invariance study in which a researcher

wishes to test whether an instrument is measuring the same traits in the same way for two or more groups. If the factor structure fails to show invariance across groups, then generalizability is compromised and meaningful comparisons across groups on factor scores is precluded (Floyd & Widaman, 1995). These factor replicability and generalizability questions are increasingly being addressed with CFA procedures (Reise et al., 2000).

There are a number of procedures for assessing measurement invariance. In the least restrictive case, the investigator examines whether the variables show the same pattern of significant factor loadings by testing the goodness of fit of a model with similar patterns of fixed and free factor loadings in each group (Floyd & Widaman, 1995; Jöreskog & Sörbom, 1989). In the more stringent case, all common factor loadings (and perhaps the unique variances) are constrained to be invariant across the groups (Floyd & Widaman, 1995; Reise et al., 1993).

The latent variable structure of the WMS-III in the normative standardization sample has been investigated using confirmatory factor analysis (Millis, Malina, Bowers, & Ricker, 1999; Price, Tulsky, Millis, & Weiss, 2002), in which five competing models were assessed. The purpose of Study 2 was to investigate the generalizability of the factor structure in a clinical group of patients with temporal lobe epilepsy. At the time Study 2 was conducted, it was the first such analysis of a clinical sample using the WMS-III. As such, a less restrictive, hypothesis testing approach was taken in which all five models assessed in the standardization sample were fit to the sample and the goodness of fit of each was evaluated.

*Chapter 2*STUDY 1: THE UTILITY OF THE WMS-III IN DIFFERENTIATING LATERALIZED TEMPORAL EPILEPTOGENIC DYSFUNCTION¹

Memory difficulty in individuals with temporal lobe epilepsy (TLE) is a phenomenon that has long been recognized and documented (Gowers, 1881; Reynolds, 1861). Patients who have undergone temporal lobectomy tend to display material-specific deficits in the ability to learn new material. Early neuropsychological studies indicated that resection of the left temporal lobe may impair the ability to learn verbal material while right temporal resection can produce a deficit in the ability to learn new nonverbal and visuospatial information (Kimura, 1963; Meyer & Yates, 1955; Milner, 1958; 1968b; Taylor, 1969; Weingartner, 1968). Although less pronounced, nonsurgical patients with unilateral temporal lobe seizures exhibit similar impairments (Delaney, Rosen, Mattson, & Novelly, 1980; Hermann, Wyler, Richey, & Rea, 1987; Loring, Lee, Martin, & Meador, 1988; Milner, 1975). Despite these reports, other investigators have failed to detect differential impairment on verbal and visuospatial tasks as a function of seizure laterality (Barr et al., 1997; Glowinski, 1973; Loiseau et al., 1983; Mayeux, Brandt, Rosen, & Benson, 1980; Naugle, Chelune, Cheek, Luders, & Awad, 1993; Naugle, Chelune, Schuster, Luders, & Comair, 1994; Schwartz & Dennerll, 1969). Since many patients undergo surgical resection for intractable temporal lobe epilepsy, it is important for neuropsychologists to develop reliable and valid methods for identifying impairment and for identifying individuals who may be at increased

¹ This work has been published: Wilde, N., Strauss, E., Chelune, G.J., Loring, D.W., Martin, R.C., Hermann, B.P., et al. (2001). WMS-III performance in patients with temporal lobe epilepsy: Group differences and individual classification. *Journal of the International Neuropsychological Society*, 7, 881-891.

risk for cognitive impairment after surgery (Dodrill, Hermann, Rausch, Chelune, & Oxbury, 1993; Jones-Gotman et al., 1993).

The most common tests used to evaluate learning and memory in individuals with epilepsy have been the Wechsler Memory Scale (WMS; Wechsler, 1945) and its first revision, the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987). An international survey of 82 epilepsy surgery centers found that 84% of centers routinely administer all or part of the WMS or the WMS-R in their pre-operative evaluations of epilepsy patients (Jones-Gotman et al., 1993). Despite its wide usage, a number of conflicting findings have been reported in studies comparing WMS or WMS-R performance levels in non-operated left and right temporal lobe epilepsy samples. Some studies have found significant group differences on selected scores (Bornstein, Drake, & Pakalnis, 1988; Delaney et al., 1980; Ivnik, Sharbrough, & Laws, Jr., 1987; Jones-Gotman, 1991; Moore & Baker, 1996), particularly when the differences between verbal and visual performance are compared (e.g., Barr, 1997b). Nonsignificant group differences between patients with left or right temporal lobe onset have also been reported (Barr et al., 1997; Chelune, Naugle, Luders, Sedlak, & Awad, 1993; Delaney et al., 1980; Glowinski, 1973; Ivnik et al., 1987; Loiseau et al., 1983; Mayeux et al., 1980; Naugle et al., 1993). When group differences occurred, they tended to be predominantly on verbal measures, leading researchers to suggest that the WMS and the WMS-R were sensitive to left but not right temporal lobe dysfunction (Chelune & Bornstein, 1988; Loring, 1989). In an analysis of over 1000 individuals with medically refractory seizures, WMS-R verbal memory deficits tended to occur in the context of left-sided dysfunction, whereas visual memory was not related to laterality (Strauss et al., 1995).

It has been suggested that within-subject comparisons may provide a better test of the ability of the WMS-R to detect material specific deficits (Naugle et al., 1993). By subtracting

the visual memory measures from their verbal counterparts, Chelune and Bornstein (1988) found that, in a mixed group of patients, those with left hemisphere dysfunction were less adept at verbal memory and learning tasks, whereas patients with right hemisphere disturbance showed the opposite pattern. Naugle et al. (1993) however, found no significant differences in pre-operative verbal-visual discrepancy scores between left temporal lobe (LTLE) and right temporal lobe (RTLE) epilepsy patients. Nonetheless, the clinical utility of this comparison may be important to the practitioner, who typically looks for intraindividual patterns and discrepancies when attempting to infer lateralization effects on a case-by case basis (Chelune & Bornstein, 1988).

Investigators have also used various magnitudes of discrepancy between verbal and visual indexes to examine the ability of these scores to predict side of temporal dysfunction. Moore and Baker (1996) found that a WMS-R Verbal-Visual Index difference at the .05 level of significance correctly predicted laterality for those people with a left temporal focus but was ineffective for those with right temporal foci, classifying most of them as having a left-sided impairment based on their discrepancy scores. Similar results were obtained in an investigation of patients who had previously undergone temporal lobectomy (Loring, Lee, Martin, & Meador, 1989).

Barr (1997b) used receiver operating characteristic (ROC) curves to determine the diagnostic accuracy of the WMS-R in the classification of epilepsy surgery candidates. Using ROC curves, one can assess the proportion of patients who can be accurately classified into left and right temporal groups based on a given score. Barr concluded that the WMS-R scores provided relatively poor discrimination of patients into left and right temporal groups, yet the highest level of classification accuracy was obtained for a measure of the difference between Verbal and Visual Memory indexes. This supports the contention that within-

subject comparisons of WMS-R scores may be relatively better indicators of lateralized effects among seizure patients than index means.

The Wechsler Memory Scale - Third Edition (WMS-III; Wechsler, 1997b) is the most recent revision of the original WMS and the WMS-R. Although the WMS-III has maintained many aspects of its predecessors, significant changes have been made in response to current research and theory. The content and structure of the WMS-III is considerably different from the WMS-R. Due to research suggesting that the WMS-R visual memory subtests were not adequate measures of a hypothetical “pure” visual memory system and were not differentially sensitive to unilateral lesions (Chelune & Bornstein, 1988; Heilbronner, 1992; Loring, 1989; Naugle et al., 1993), new visual memory subtests were developed, and include both immediate and delayed trials.

The WMS-III nomenclature of the index scores has also changed such that now a distinction is made between “auditory” and “visual” memory to reflect the modality of presentation of the subtests, rather than purporting to tap exclusively a hypothetical verbal or visual memory system as the WMS-R labels suggested (The Psychological Corporation, 1997).

The index structure of the WMS-III also differs considerably from those of its predecessors and is formed by summing the scaled scores of the subtests to ensure equal weighting of the components. In addition to an Auditory Immediate Index and a Visual Immediate Index, three modality-specific delayed indexes are calculated: the Auditory Delayed Index, the Visual Delayed Index, and the Auditory Recognition Delayed Index. It has been suggested that performance differences on the immediate and the delayed tasks have some clinical utility (Tulsky et al., 2000), and that the delayed scores are likely more ecologically valid (The Psychological Corporation, 1997). A Working Memory Index is

composed of one auditory and one visual working memory task (see Wechsler, 1997b and The Psychological Corporation, 1997 for additional information about changes in the WMS-III).

There has been little research thus far on the utility of the WMS-III in patients with epilepsy. The *WAIS-III - WMS-III Technical Manual* (The Psychological Corporation, 1997) provides some preliminary data suggesting that the new measures of auditory and visual memory may be useful in determining laterality of dysfunction among patients who have undergone temporal lobectomies (p. 159), although the sample size was quite small (LTLE = 15, RTLE = 12). Data on pre-operative epilepsy patients is not provided.

As with the revision of any widely used instrument, there is an empirical need to establish its utility. This is of particular relevance in the assessment of patients with epilepsy because scores on the WMS-III are assumed to aid in the localization of dysfunction. Accordingly, the purpose of this study was to assess the criterion validity of the WMS-III in differentiating those with left- or right-sided temporal lobe disturbance. To the extent that the test taps distinct auditory and visual memory processes, it should be able to accurately identify those with lateralized temporal impairment both at the mean level as well as at the level of the individual patient. Methods of analysis included evaluation of group means on the various WMS-III indexes and subtest scores, the use of ROC curves to determine the classification accuracy of the WMS-III, and an examination of WMS-III Auditory-Visual Index discrepancy scores to determine if this within-subject comparison could reliably indicate side of temporal dysfunction. In addition, performance on the immediate and delayed indexes in the auditory and visual modalities was compared within each group to determine the utility of this distinction in this population.

Method

Participants

The study sample was selected from a database of patients with temporal lobe epilepsy and medically refractory seizures from three epilepsy surgery centers participating in the Bozeman Epilepsy Consortium²: the Cleveland Clinic Foundation, Cleveland, Ohio, the Medical College of Georgia, Augusta, Georgia, and the University of Alabama at Birmingham Epilepsy Center, Birmingham, Alabama. Patients were considered for inclusion in this study if they met the following criteria: (a) unilateral seizure onset of temporal lobe origin confirmed by EEG / video monitoring; (b) information was available regarding age of onset of recurrent seizures, duration (computed as age at time of examination minus age at seizure onset), sex, hand preference, and Full Scale IQ (FSIQ) as measured by the Wechsler Adult Intelligence Scale – Third Edition (Wechsler, 1997a) and (c) they had received neuropsychological evaluation including the Wechsler Memory Scale - Third Edition (Wechsler, 1997b). The Cleveland Clinic and Medical College of Georgia also routinely included the intracarotid sodium amytal procedure (IAP) for language and memory in their evaluation of patients. From this information only those patients with left hemisphere language representation were selected for inclusion in the study. Since speech representation data were not available from the University of Alabama, only those patients who demonstrated a right-hand preference were selected from this center, in order to maximize the probability of left hemisphere dominance for speech.

A total of 102 patients met criteria for inclusion in the study. The characteristics of the patients classified by side of dysfunction and examination center are provided in Table 2.1.

² Data contributed from all centers were extracted from deidentified patient registries that were reviewed by their respective institutional review boards.

Table 2.1*Sample Characteristics for Each Center and the Total Sample*

Center	Group	n	Age	Education	Full Scale	Age of	Duration	Gender		Hand.	
			(years)	(years)	IQ	onset	of epilepsy	M	F	R	L
			<i>M</i> (<i>SD</i>)								
Cleveland Clinic Foundation	LTLE	25	32.20 (10.38)	12.76 (1.88)	89.08 (13.29)	13.30 (11.60)	19.02 (11.63)	12	13	21	4
	RTLE	29	33.62 (10.78)	12.83 (2.58)	92.62 (12.53)	13.14 (9.19)	20.48 (12.54)	17	12	25	4
Medical College of Georgia	LTLE	9	28.67 (8.26)	12.44 (2.51)	90.67 (13.83)	16.33 (12.74)	12.33 (8.40)	6	3	8	1
	RTLE	6	32.67 (10.37)	13.50 (2.81)	87.00 (16.20)	12.17 (7.28)	20.50 (12.76)	4	2	6	0
Univ. of Alabama Epilepsy Center	LTLE	21	39.05 (10.98)	12.33 (2.78)	80.10 (14.45)	18.84 (16.63)	21.88 (16.47)	7	14	21	0
	RTLE	12	35.50 (12.66)	13.17 (3.13)	82.92 (13.01)	21.58 (11.59)	13.42 (11.16)	2	10	12	0
Total Sample	LTLE	55	34.24 (10.90)	12.55 (2.32)	86.02 (14.32)	15.91 (13.89)	19.02 (13.48)	25	30	50	5
	RTLE	47	33.98 (11.03)	13.00 (2.70)	89.48 (13.45)	15.17 (10.19)	18.68 (12.36)	23	24	43	4

Note. LTLE = left temporal lobe epilepsy; RTLE = right temporal lobe epilepsy; WAIS-III = Wechsler Adult Intelligence Scale - Third Edition; M = male; F = female; Hand. = handedness; R = right hand preference; L = left hand preference.

The presence of pre-existing differences across centers was examined with separate analyses of variance for age, education level, age of onset, duration, and FSIQ. Chi-square analyses were conducted to evaluate differences in sex and handedness. Center differences were found for age [$F(2, 99) = 3.21, p < .05$], age of onset [$F(2, 99) = 3.16, p < .05$], and FSIQ [$F(2, 97) = 5.57, p < .01$]. In each of these cases, the patients from University of Alabama differed from the other two centers, that is, this patient sample was significantly older, had a later age of onset, and a lower FSIQ than the patients from the other two centers. Sex also differed significantly between the sites [$\chi^2(2, N = 102) = .277, p < .05$], with

the University of Alabama sample containing a larger proportion of female patients than the other two centers.

To examine the effect of center on the WMS-III variables, a one-way MANOVA was conducted on the WMS-III Primary Indices. The multivariate effect was non-significant [Wilks' Lambda $F(12, 176) = 1.73, ns$]. Thus, center was not considered when computing statistical analyses.

Procedures

Participants were administered the WMS-III as part of comprehensive neuropsychological evaluations. Analyses of the data were limited to subtasks common to all centers, which included the Primary Index scores and associated subtest scores. Supplementary scores were not included in the analyses since the specific tasks that were administered differed across sites. In addition, age-corrected scaled and standard scores served as the units of analysis, since raw scores were not available from all centers. Tests were administered and scored by trained personnel according to standardized procedures provided in the WMS-III manual (Wechsler, 1997b).

Statistical Analysis

Data were analyzed in a series of steps designed to evaluate WMS-III performance in individuals with right and left temporal seizure foci. First, a descriptive analysis of the characteristics of the sample was performed. Second, differences in group means for the primary memory indexes and the individual subtests were assessed with a series of independent t tests. Third, discrepancy scores, calculated by subtracting the Visual Memory Index from the Auditory Memory Index, were compared between the groups for both immediate and delayed indexes. Fourth, the immediate and delayed indexes in each modality

were compared within each group via paired sample *t* tests. Because of the large number of comparisons, applying the Bonferroni correction method to account for Type I error was considered. However, this approach is highly conservative, lacks power to reject an individual hypothesis, and may mask actual differences across groups (e.g., Olejnik, Li, Supattathum, & Huberty, 1997; Simes, 1986). Thus, in order to protect against excessive Type I error, while maintaining adequate power and minimizing the risk of Type II error, alpha was set at .025 for statistical significance for all group comparisons.

Fifth, ROC curves were calculated for the WMS-III primary indexes, subtests, and discrepancy scores to evaluate the diagnostic efficiency of the WMS-III. The area under the curve (AUC), the maximal cut-off score, and a suggested cut-off score based on an a priori determination of specificity values greater than 70% (with the highest accompanying level of sensitivity), were calculated using non-parametric analyses (Barr, 1997b).

Finally, Auditory-Visual Immediate and Delayed Index difference scores were further evaluated to examine the utility of different magnitudes of discrepancy for patient classification. Discrepancy criteria were obtained from the WMS-III manual and included (a) the .05 level of statistical significance determined from measurement error of the Auditory and Visual Indices, and (b) the difference between the indexes corresponding to a frequency of occurrence of less than 5% in the standardization sample.

Results

Sample Characteristics

The mean age of the sample was 35.0 years (SD = 11.1) and the mean educational level was 12.8 years (SD = 2.3). Mean WAIS-III FSIQ was 88.7 (SD = 16.1). Patient characteristics of the RTLE and LTLE groups are presented in Table 2.1. A comparison of

groups according to demographic variables revealed no significant differences in group composition for age, education, age at onset or duration, FSIQ, sex, or hand dominance.

Group Differences Among the Primary Indexes and Subtest Scores

The means and standard deviations for the WMS-III primary indexes and subtest scores are provided in Table 2.2. Univariate t tests of the primary index scores indicated that the RTLE and LTLE group differed significantly from one another only on the Auditory Delayed Index [$t(100) = 2.39, p < .025$], with the LTLE group obtaining lower Auditory Delayed Index scores than the RTLE group. Performance on only one subtest, Verbal Paired Associates II [$t(100) = 2.72, p < .01$], differed significantly between the RTLE and the LTLE groups.

Auditory – Visual Index Discrepancy Comparison

Analyses of the Auditory – Visual Index discrepancy scores revealed differences for the RTLE and the LTLE groups, for both the Immediate [$t(98) = 2.95, p < .01$] and the Delayed scores [$t(97) = 3.82, p < .001$]. Furthermore, the net difference scores were in the positive direction for the RTLE group indicating that Visual Index scores were lower than Auditory Index scores, whereas the opposite was the case for the LTLE group.

Figure 2.1 shows each group's mean performance on the individual indexes and subtests, with the scores converted to z -scores for ease of comparison across scales³. Examination of the figure reveals that, while the performance of the LTLE group was uniformly low, performance of the RTLE patients was depressed on the visual subtests only.

³ The scores shown in Figure 2.1 were calculated for each scale by converting individual scaled scores ($M = 10, SD = 3$) or standard scores ($M = 100, SD = 15$) to z -scores by subtracting the normative mean (i.e., 10 for subtest scores and 100 for index scores) from each participant's score, dividing by the normative standard deviation, and then calculating a mean z -score for each group.

Table 2.2*Mean WMS-III Scores for the Right and Left Temporal Lobe Epilepsy Groups*

WMS-III score	<i>n</i>	LTLE		<i>n</i>	RTLE	
		<i>M</i>	(<i>SD</i>)		<i>M</i>	(<i>SD</i>)
Indexes						
General Memory Index	55	82.68	(17.04)	46	86.11	(16.13)
Auditory Immediate Index	55	84.29	(16.58)	47	89.81	(15.87)
Visual Immediate Index	53	85.91	(15.89)	47	81.87	(14.16)
Immediate Memory Index	53	82.19	(16.67)	47	83.15	(16.56)
Auditory Delayed Index*	55	82.64	(17.53)	47	90.91	(17.42)
Visual Delayed Index	55	85.43	(15.92)	46	81.63	(14.67)
Auditory Recognition Index	55	90.55	(16.49)	47	94.15	(15.30)
Working Memory Index	53	87.36	(14.84)	46	91.67	(16.28)
Auditory Immediate - Visual Immediate Index**	53	-1.11	(16.74)	47	7.94	(13.48)
Auditory Delayed Index - Visual Delayed Index***	55	-2.40	(15.25)	46	9.59	(15.90)
Subtests						
Logical Memory I	55	7.38	(3.29)	47	8.36	(3.38)
Logical Memory II	55	7.09	(3.45)	47	8.34	(3.36)
Faces I	55	7.93	(2.61)	47	7.77	(2.15)
Faces II	55	7.96	(2.23)	47	7.66	(2.30)
Verbal Paired Ass. I	55	7.18	(3.01)	47	8.17	(2.67)
Verbal Paired Ass. II**	55	6.87	(3.44)	47	8.64	(3.05)
Family Pictures I	53	7.60	(3.35)	47	6.55	(2.90)
Family Pictures II	55	7.20	(3.63)	45	6.80	(3.09)
Letter-Number Seq.	53	8.13	(3.16)	46	8.39	(3.21)
Spatial Span	54	7.63	(3.14)	47	8.66	(3.49)

Note. *n*'s differ due to cases of missing data. LTLE = left temporal lobe; RTLE = right temporal lobe; WMS-III = Wechsler Memory Scale - Third Edition.

p* < .025. *p* < .01. ****p* < .001.

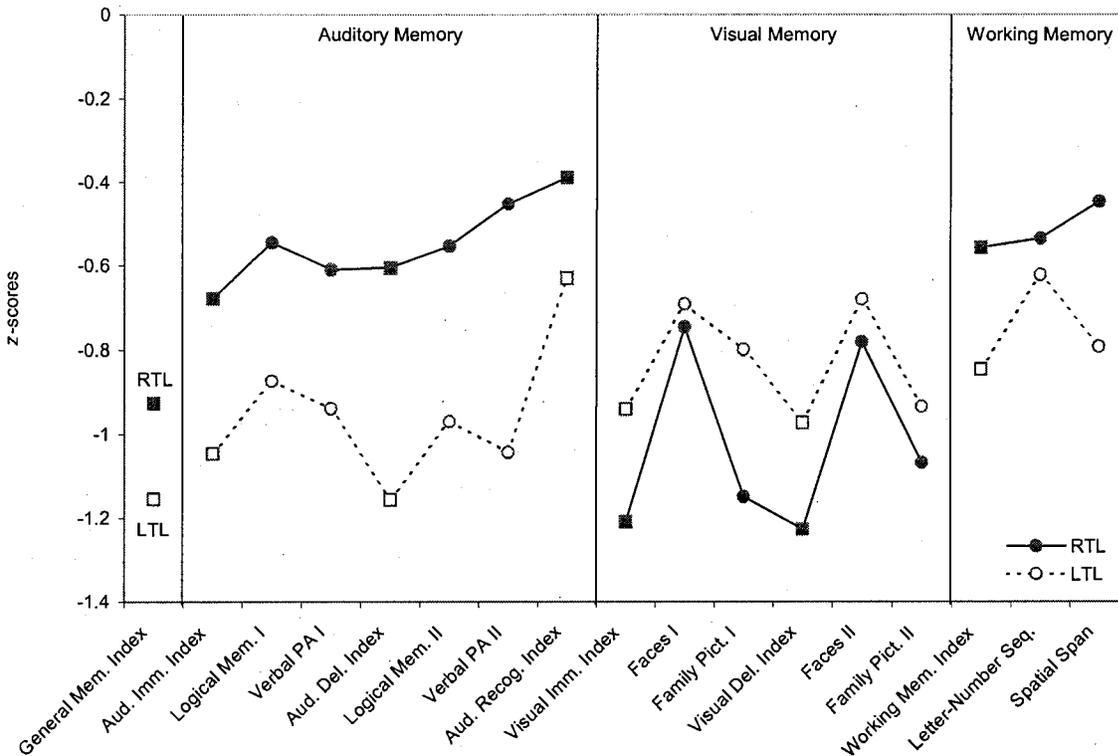


Figure 2.1

Mean z scores for the RTLE (right temporal lobe epilepsy) and LTLE (left temporal lobe epilepsy) groups on the WMS-III indexes and individual subtests. Note that better performance is represented by z values closer to the normative mean of 0. Square markers denote index scores and circle markers denote subtest scores.

Comparison of Immediate and Delayed Index Scores

Performance on the Visual Immediate and Delayed Indexes was compared in each group. This procedure was repeated with the Auditory Index scores. Paired-samples *t* tests revealed that performance differences between immediate and delayed index scores were not statistically significant for either modality, in either group (see Table 2.3).

Table 2.3
Immediate-Delayed Index Score Differences

Index modality	Group	<i>n</i>	Immediate Index score	Delayed Index score	Difference score		<i>t</i> test value	<i>p</i>
			<i>M</i>	<i>M</i>	<i>M</i>	(<i>SD</i>)		
Auditory	LTLE	55	84.29	82.64	1.65	(7.19)	1.71	.09
	RTLE	47	89.81	90.91	-1.11	(8.44)	-.90	.37
Visual	LTLE	53	85.91	85.43	.47	(7.53)	.46	.65
	RTLE	46	82.02 ^a	81.63	.39	(9.01)	.29	.77

Note. Significance test is two-tailed. Difference score = Immediate Index score - Delayed Index score. RTLE = right temporal lobe, LTLE = left temporal lobe. *p* = obtained significance level.
^aIndex mean differs from value listed in Table 2.2 due to missing data for one participant on the Visual Delayed Index score, and thus for calculation of the difference score.

Receiver Operating Characteristic Curves

ROC curve analyses were computed and analyzed in a manner similar to those described by Monsch and colleagues (1992) and Barr (1997b). Each score was treated as a separate cutoff. Measures of sensitivity (*Se*) and specificity (*Sp*) were based on the cumulative number of RTLE and LTLE patients who obtained scores at or below these cutoff values. *Se* and *Sp* (1 - *Sp*) values were plotted graphically to obtain ROC curves for each test.

The results of ROC curve analyses are provided in Table 2.4. The most common index for describing an ROC curve is the area under the curve (AUC; Swets, 1988). Areas close to .50 indicate that the classification is close to chance level, while areas close to 1.0 indicate perfect discrimination. The total areas for individual subtest scores in this study ranged from a low of .524 (Faces I) to a high of .647 (Verbal Paired Associates II). The largest AUC for any score was observed for the difference between the Auditory and Visual Delayed Memory Indexes (AUC = .702). This ROC curve is provided in Figure 2.2.

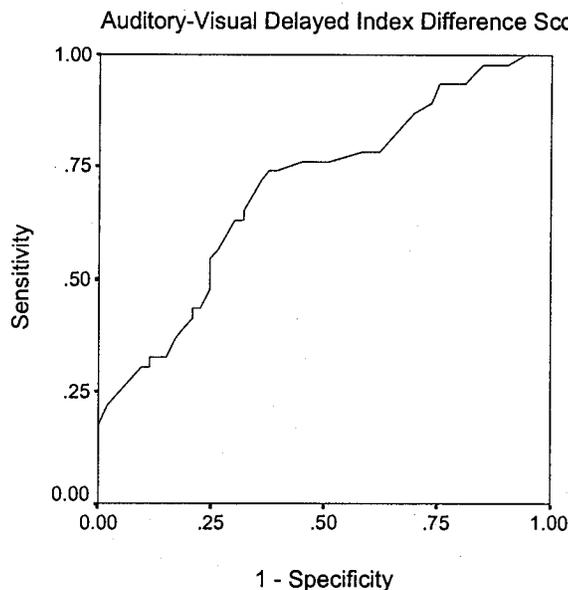


Figure 2.2

Receiver operating characteristic (ROC) curve for the WMS-III Auditory-Visual Delayed Index difference score. AUC = Area under the curve.

Empirically derived cutting scores can be obtained from ROC curves for use in making diagnostic decisions. Two cutting scores were calculated in this study. First, the maximal cutting score defined as those scores where the sum of Se and Sp reaches a maximum value was calculated. These scores provide maximal separation of groups, irrespective of sensitivity and specificity values. Second, it was determined *a priori* based on Barr (1997b) that a suggested cutting score with Sp values exceeding 70% and the highest accompanying level of Se would be most appropriate for making clinical decisions between patients with right and left temporal lobe dysfunction. Maximal and suggested cutting scores and their respective Se and Sp values are included in Table 2.4.

Table 2.4*ROC Curve Statistics for WMS-III Index and Subtest Scores*

WMS-III scale	AUC	Maximal cutting score				Suggested cutting score		
		Score	Se	Sp	Se+Sp	Score	Se	Sp
Indexes								
General Memory Index	0.569	80	0.70	0.49	1.19	90	0.37	0.70
Auditory Imm. Index	0.617	90.5	0.53	0.73	1.26	90.5	0.53	0.73
Visual Immediate Index	0.596	82.5	0.60	0.62	1.22	86	0.49	0.70
Immediate Mem. Index	0.523	88	0.40	0.70	1.10	88	0.40	0.70
Auditory Delayed Index	0.657	93	0.51	0.80	1.31	90.5	0.55	0.73
Visual Delayed Index	0.589	86	0.51	0.70	1.21	86	0.51	0.70
Auditory Recog. Index	0.563	87.5	0.70	0.42	1.12	97.5	0.38	0.71
Working Memory Index	0.586	106.5	0.26	0.92	1.19	97.5	0.43	0.74
Subtests								
Logical Memory I	0.611	8.5	0.66	0.64	1.30	9.5	0.40	0.73
Logical Memory II	0.626	7.5	0.70	0.56	1.27	8.5	0.47	0.71
Verbal Paired Ass. I	0.601	9.5	0.32	0.85	1.17	9.5	0.32	0.85
Verbal Paired Ass. II	0.647	5.5	0.87	0.44	1.31	9.5	0.45	0.76
Faces I	0.524	7.5	0.58	0.51	1.09	8.5	0.38	0.70
Faces II	0.546	8.5	0.42	0.70	1.12	8.5	0.42	0.70
Family Pictures I	0.606	6.5	0.60	0.60	1.20	8.5	0.38	0.77
Family Pictures II	0.539	9.5	0.29	0.84	1.14	8.5	0.36	0.76
Letter-Number Seq.	0.536	8.5	0.57	0.55	1.11	10.5	0.24	0.79
Spatial Span	0.591	10.5	0.40	0.78	1.18	9.5	0.47	0.70
Difference scores								
Auditory - Visual Immediate Index	0.655	-0.5	0.77	0.51	1.28	1.5	0.57	0.70
Auditory - Visual Delayed Index	0.702	0.5	0.62	0.74	1.36	1.5	0.64	0.72

Note. AUC = area under the curve; Se = sensitivity; Sp = specificity; WMS-III = Wechsler Memory Scale - Third Edition.

The most accurate maximal cutting score was obtained from the Auditory-Visual Delayed Memory Index discrepancy. For this measure, a score of 0.5 yielded a sensitivity of .62 and a specificity of .74 (sum = 1.36). This means that 62% of the LTLE patients obtained Auditory-Visual Index difference scores of 0.5 or below, whereas 74% of the RTLE patients obtained scores exceeding that level. The Auditory Delayed Index (score = 93, Se + Sp = 1.31) and the Verbal Paired Associates II subtest (score = 5.5, Se + Sp = 1.31) yielded cutting scores with the next highest levels of maximal separation. The Delayed Index difference score also provided the best separation of groups when utilizing the suggested cutting score, with a Sp value of .72 and a Se value of .64. The Auditory Immediate Index, Auditory Delayed Index and the Auditory-Visual Immediate Memory Index difference score also exhibited modest discrimination, with Se values exceeding 50%.

Patient Classification Using Significant and Infrequent Index Discrepancies

Auditory-Visual Index discrepancy scores were evaluated further to examine the ability of the WMS-III in predicting side of temporal lobe seizure focus. Two discrepancy criteria were used to classify patient performance. The first represented the .05 level of statistical significance determined from the measurement error of the Auditory and Visual Indices, obtained from the WMS-III manual (Wechsler, 1997b). This resulted in discrepancy scores of 15 points for the immediate indexes, and 17 points for the delayed indexes. The second discrepancy criterion was obtained from the rarity of difference scores in the standardization sample. Tables included in the WMS-III manual report the frequency of discrepancies independent of the directionality of the score. That is, the cumulative percentages listed in Table F.2. (p. 206) combine individuals who obtained an Auditory Index score that was higher than their Visual Index score, and people who showed the reverse pattern. Based on suggestions of Tulskey and colleagues for use with the WAIS-III (Tulskey, Rolfhus, & Zhu,

2000), the frequencies reported in Table F.2. should be divided in half to obtain the appropriate base rate when a directional hypothesis is being tested. Thus, to obtain a 95% level of confidence, a 27-point difference was required for the immediate indexes, and a 26-point discrepancy for the delayed indexes (which correspond to cumulative percentages obtained in 10% of the standardization sample as listed in Table F.2.).

Patients were grouped into one of three categories (left, right, inconclusive) on the basis of their Auditory-Visual Memory Index discrepancy scores. If the Auditory Memory Score was significantly below the Visual Memory score, the patient was classified as having probable left temporal dysfunction. Similarly, if the Visual Memory score was significantly lower than the Auditory Memory score, the patient was classified as having probable right temporal dysfunction. Discrepancies not exceeding the criterion were deemed inconclusive for indicating laterality.

Eighteen RTLE patients had immediate index discrepancy scores of 15 or more points. In 16 of 18 patients, the Visual Memory Index was the lower value, which is consistent with right temporal lobe dysfunction. There were 22 LTLE patients meeting the 15-point difference criterion. However, 12 of these patients had significantly lower Visual Memory Indices, suggesting relative impairment of *right* temporal lobe function. These results are shown in Table 2.5.

Similar patterns were evident when statistically significant discrepancies between the delayed indices were examined. The majority of the RTLE patients meeting the 17-point criterion were correctly classified, but a large proportion of LTLE patients exhibited relatively greater impairment on visual memory tasks.

Table 2.5*Auditory – Visual Index Difference Scores and Classifications*

WMS-III Index difference	Size and direction of difference score	LTLE		RTLE	
		<i>n</i>	%	<i>n</i>	%
Auditory – Visual Immediate Index	≤ -15	10	83.3	2	16.7
	≥ 15	12	42.9	16	57.1
	≤ -27	5	83.3	1	16.7
	≥ 27	1	16.7	5	83.3
Auditory – Visual Delayed Index	≤ -17	10	76.9	3	23.1
	≥ 17	9	34.6	17	65.4
	≤ -26	5	83.3	1	16.7
	≥ 26	0	0	8	100

Note. *n* = number of patients in each group with difference scores meeting or exceeding the stated magnitudes; italicized values indicate patients who obtained difference scores in the direction opposite to prediction. % = percent of patients with given difference scores who fall within each group; LTLE = left temporal lobe; RTLE = right temporal lobe; WMS-III = Wechsler Memory Scale – 3rd Edition.

Twelve patients (6 RTLE, 6 LTLE) had discrepancies that exceeded the more conservative 27-point criterion for immediate index differences. In each group, one out of six patients was incorrectly classified. There were 14 patients (9 RTLE, 5 LTLE) who met the 26-point criterion for the delayed discrepancy score. Using this stringent criterion, only one RTLE patient was misclassified.

Alternatively, it is also useful to identify the likelihood of being correct in the classification of laterality when given a certain discrepancy score. As indicated in Table 2.5, the likelihood of correctly classifying a patient with a large negative discrepancy between Auditory and Visual indexes as having left temporal dysfunction was in the range of 75-85% across all given discrepancy criteria. However, positive discrepancies based on difference scores calculated from statistical significance levels led to correct prediction of RTLE

patients in only 55-65% of cases. This is due to the large number of LTLE patients who obtained discrepancy scores in the direction opposite to prediction (i.e., significantly better Auditory than Visual Index scores). With very large and infrequent discrepancy scores, improved prediction of patients was obtained. It must be kept in mind, however, that few individuals (less than 15% of the sample) displayed such large discrepancies.

Discussion

The purpose of the present study was to examine the utility of the WMS-III in predicting laterality of impairment in patients with temporal lobe epilepsy. The results suggest that the new WMS-III does not represent a significant improvement over its predecessors in its ability to distinguish patients with left and right temporal dysfunction associated with a unilateral seizure onset. LTLE patients tended to perform more poorly on the auditory/verbal tasks than the RTLE group, whereas the RTLE patients showed the opposite pattern of performance. However, group performance on the WMS-III indexes and subtests was largely insensitive to laterality. Within subject performance as demonstrated by auditory-visual difference scores appeared most sensitive to side of temporal dysfunction. This is consistent with previous research on the WMS-R, which has suggested that discrepancy scores may be most useful at detecting material-specific memory impairments (Bornstein et al., 1988; Chelune & Bornstein, 1988; see Naugle et al., 1993 for negative findings). It is important to note, however, that when considering the performance within each group, material-specific performance was not observed for the LTLE group. Thus, the LTLE group performed at the same low level on the auditory and the visual subtests. Indeed, the auditory-visual discrepancy scores obtained by the LTLE group did not differ significantly from zero. This calls into question the selectivity of verbal memory deficits in

LTLE patients as measured by the WMS-III. On the other hand, in the RTLE group a more specific pattern of performance was demonstrated such that depressed performance of patients with right temporal lobe seizures was relatively specific to the visual tasks.

At the individual subtest level, some tasks appeared more sensitive to laterality of dysfunction than others. Verbal Paired Associates II was the only subtest to differ significantly between the groups, due to the very low mean performance of LTLE patients. This finding is consistent with much of the literature on memory functioning in epilepsy demonstrating impairment on verbal tasks in patients with left temporal dysfunction (Chelune & Bornstein, 1988; Hermann et al., 1987; Loring et al., 1988; Moore & Baker, 1996). Of the new visual subtests included in the WMS-III, the Family Pictures subtest appeared most sensitive to right temporal lobe dysfunction, despite the fact that patients are required to visually encode as well as verbally recall the content in this task. The fact that dual encoding and processing is required may account at least in part for the nonsignificant difference between LTLE and RTLE patients on this task. Note however that such an explanation cannot account for the failure of the Faces subtest to distinguish between groups.

The area under the ROC curve provides a quantitative index of the diagnostic accuracy of a given score. The area values in this study, while somewhat higher than those reported by Barr (1997b) using WMS-R scores, were still substantially lower than those reported in other studies using ROC curves and neuropsychological test scores with clinical populations (Drebing, Van Gorp, Stuck, Mitrushina, & Beck, 1994; Engelhart, Eisenstein, & Meininger, 1994; Guilmette & Rasile, 1995; Monsch et al., 1992). As would be expected based on the obtained results of group differences, the highest level of classification accuracy in this study was obtained using the auditory-visual discrepancy scores, a finding also observed by Barr

(1997b) in his analysis of the WMS-R (see also Loring, Hermann, Lee, Drane, & Meador, 2000, with regard to the Memory Assessment Scales). The benefit of using ROC curves is that the analysis provides an empirically derived cutting score to aid in diagnostic classification. Using the cutting score of 0.5 obtained from the Auditory-Visual Delayed Index, which had the highest combined level of sensitivity and specificity (maximal cutting score), 38% of the LTLE group and 26% of the RTLE group were incorrectly classified. Thus, although diagnostic accuracy was significantly better than chance, the classification rates obtained from this study were not within an acceptable range to have utility for clinical use.

When discrepancy scores were further examined based on the magnitude of difference, classification rates also provided unsatisfactory results. First, the vast majority of patients did not produce results that met the discrepancy criteria, therefore minimizing the utility of this approach. Second, a large proportion of patients were misclassified. Statistically significant discrepancy scores from the WMS-III were more accurate in predicting laterality for people with right temporal focus than for patients with left temporal dysfunction; many of the LTLE patients would have been classified as having right temporal dysfunction based on their discrepancy scores. Consequently, the ability to correctly predict right-sided laterality given a positive Auditory-Visual Index difference score was especially poor since more than one third of patients who obtained statistically significant positive difference scores were LTLE patients. The particularly poor classification of LTLE patients in this study is contrary to the findings of other researchers examining index score discrepancies using the WMS-R (Loring et al., 1989; Moore & Baker, 1996) who generally found that LTLE patients were correctly classified while RTLE patients were incorrectly classified. The reason for this difference is unclear. At a group level, index scores are within expectations.

Inspection of the individual patients who were misclassified does not suggest any differences in terms of demographic factors, and the WMS-III index scores obtained from these patients spans the range from severely impaired to superior. This finding awaits replication from other researchers, but these preliminary results suggest that the WMS-III may have somewhat different characteristics with respect to laterality than did the WMS-R. Using the more conservative discrepancy criteria of unusually large index differences resulted in improved prediction of laterality; very few individuals were misclassified. However, the rarity of such large discrepancies in this population limits the usefulness of this approach. Specifically, the utility of the WMS-III in characterizing, detecting, and classifying individuals with lateralized temporal dysfunction is put into question by the results of this study.

Another goal of this study was to examine the utility of the immediate versus the delayed memory indexes, since it has been suggested that the delayed measures may be more clinically relevant and ecologically valid than the immediate scores. Factor analytic support of the distinction between immediate and delayed memory dimensions is provided in the WAIS-III – WMS-III Technical Manual (The Psychological Corporation, 1997, p. 115). In this study, there were no differences between immediate and delayed index performance in either the auditory or the visual modality, in either the LTLE or the RTLE group. This sheds some doubt on the particular significance of the delayed memory scores and suggests that the immediate and the delayed subtests may be assessing similar functions. Each subtest on the WMS-III requires the retention of material for the immediate task beyond that which would be possible based on models of working memory, and thus it seems likely that to perform adequately on the immediate memory measures, multiple memory components including encoding, storage, and retrieval would be required. The lack of distinction between immediate and delayed measures may be population specific and awaits further

research with other neurological patient groups in which a distinction might be expected, such as in patients with Alzheimer's disease or Wernicke – Korsakoff's syndrome.

A number of limitations of the present study must be acknowledged. First, analysis of raw scores may have resulted in additional findings, since scaled scores coincide with ranges of raw scores and thus potentially reduce the variance of results, especially at the extremes. In addition, percent retention scores, which have been shown to differentiate LTLE and RTLE groups (e.g., Delaney et al., 1980; Jones-Gotman, 1991), were unavailable in this study. While this study illustrated the relative merit of utilizing discrepancy scores over group means, it will be useful in the future to examine discrepancies between the auditory and visual subtests, with respect to absolute differences and percent retention. For example, studies utilizing the WMS-R have often compared performance on Logical Memory to Visual Reproduction (e.g., Chelune & Bornstein, 1988; Naugle et al., 1993).

In addition, it may be that parsing or combining tasks in a different manner results in an increased sensitivity of the WMS-III to laterality effects. For example, Holley and colleagues (Holley, Lineweaver, & Chelune, 2000) divided the Family Pictures subtest into Character, Location, and Action components, and examined performance in patients who had undergone temporal lobectomies. After statistically removing verbal memory scores, they found that the Location score was sensitive to right temporal lobe dysfunction. Further studies that investigate alternative methods of looking at WMS-III performance will be useful in determining its ability to detect laterality differences.

In addition, this study did not address issues such as the degree of mesial temporal sclerosis and the integrity of the contralateral hippocampus. Memory functioning has been shown to vary according to such indicators (see Bell & Davies, 1998, for review). Thus,

future research in patients with and without hippocampal pathology presents an important avenue for further study.

While the purposes of this study were to examine the ability of the WMS-III to detect laterality in a presurgical sample, it is expected that the magnitude of modality-specific differences would be enhanced following temporal lobectomy. The Technical Manual provides some preliminary data suggesting that the WMS-III scores are sensitive to laterality in postsurgical epilepsy patients. However, the sample size was quite small and analysis of the data was rather limited. A more comprehensive study of WMS-III in epilepsy patients after temporal lobectomy is needed.

Obviously, considerable research is needed before the utility of the WMS-III in patients with epilepsy is known. The present results indicate that the WMS-III alone is limited in the prediction of laterality in epilepsy patients. In particular, selective verbal memory deficits were not demonstrated for those patients with left temporal foci, as indicated by their poor overall performance and high misclassification rate. Future research and clinical experience with the WMS-III will demonstrate whether the findings from this study are replicable and will assist in establishing its usefulness as a neuropsychological measure in the epilepsy population.

This study also emphasizes the fact that scores from the WMS-III should not be used in isolation. It may be that the combination of the WMS-III with other neuropsychological or diagnostic measures provides an improved rate of classification of epilepsy patients. Furthermore, the limitations of the WMS-III in classifying patients with temporal lobe seizures should not be extended to making predictions regarding the use of the WMS-III in other clinical populations. In addition, it should be recognized that neuropsychological testing of epilepsy surgery candidates serves a number of useful purposes aside from

identifying laterality of seizures, such as providing valuable baseline information for evaluating change after surgery and for identifying those who may be at risk for subsequent impairment (Dodrill et al., 1993). Finally, it is important to bear in mind that the utility of the WMS-III lies in its ability to measure memory (that is, its ability to provide an internally and externally valid indication of memory functioning), not only in its ability to differentiate patient groups.

*Chapter 3*STUDY 2: CONFIRMATORY FACTOR ANALYSIS OF THE WMS-III IN PATIENTS WITH TEMPORAL LOBE EPILEPSY⁴

Since its inception, the Wechsler Memory Scale (WMS; Wechsler, 1945) and its first revision, the Wechsler Memory Scale— Revised (WMS-R; Wechsler, 1987) have been among the most frequently used tests to evaluate learning and memory in patients with temporal lobe epilepsy (Jones-Gotman et al., 1993). This established scale is intended to provide an assessment of multiple dimensions of attention and memory in clinical settings (Lezak, 1995).

A number of factor-analytic studies have been reported for the WMS-R in an attempt to identify its underlying structure. However, there exist substantial discrepancies in the reported factor structures identified in this research, including models with a single, general memory factor (Elwood, 1991); two factors (general memory and attention/concentration: Bornstein & Chelune, 1988; Roid, Prifitera, & Ledbetter, 1988; Smith et al., 1992; Wechsler, 1987); and a variety of three-factor solutions (attention/concentration, verbal memory, visual memory: Bornstein & Chelune, 1989; Jurden, Franzen, Callahan, & Ledbetter, 1996; attention/concentration, general memory, percentage retention: Smith et al., 1992; attention/concentration, immediate memory, delayed memory: Burton, Mittenberg, & Burton, 1993; Roth, Conboy, Reeder, & Boll, 1990; Woodward, 1993). Some of the variability in the observed factor structures is likely related to the use of different statistical procedures.

⁴ This work has been published: Wilde, N.J., Strauss, E., Chelune, G.J., Hermann, B.P., Hunter, M., Loring, D.W. et al. (2003). Confirmatory factor analysis of the WMS-III in patients with temporal lobe epilepsy. *Psychological Assessment*, 15, 56-63.

Confirmatory factor analysis (CFA) provides a more theoretically driven approach than exploratory methods (e.g., principal-components analysis, common factor analysis) and is more suitable for confirmation, and comparison, of hypothesized factor structures (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Floyd & Widaman, 1995). In addition, CFA procedures are better equipped than exploratory techniques to partition method variance due to common measurement methods from that associated with the conceptual model. This is of particular importance in analysis of the Wechsler Scales when immediate and delayed trials from the same subtests are included (see Larrabee, Kane, Schuck, & Francis, 1985; Roth et al., 1990; Smith, Malec, & Ivnik, 1992, for discussion of this issue). Studies that have used CFA procedures to investigate the latent structure of the WMS-R have tended to support a three-factor model, consisting of attention/concentration, immediate memory, and delayed memory factors (Burton et al., 1993; Roth et al., 1990; Woodward, 1993; although see Jurden et al., 1996, for contradictory findings).

Although the Wechsler Memory Scale – Third Edition (WMS-III; Wechsler, 1997b) has maintained many aspects of its predecessors, significant changes have been made in response to current research and theory (see The Psychological Corporation, 1997; and Tulskey & Ledbetter, 2000, for an elaboration of changes in the WMS-III). For example, a number of new subtests have been developed, and the index structure has changed considerably. Given the extent of these changes, an examination of the factor structure of the new scale is essential to establishing its construct validity and facilitating its clinical interpretation (MacCallum & Austin, 2000).

Generally, the first step in evaluating the factor structure of a scale is to analyze the performance of the normative group. The *WAIS-III—WMS-III Technical Manual* (The Psychological Corporation, 1997) presents the results of CFAs to examine the latent

structure of the WMS-III with the standardization sample divided into three age bands (16-29, 30-64, and 65-89 years). Five competing models were evaluated, and it was concluded that a three factor model (working memory, auditory memory, and visual memory) provided the best fit of the data for the 16-29 age group, whereas a five-factor model, specifying modality-specific immediate and delayed factors, attained higher fit for the other age groups. These findings were presented as substantiation for the construct validity of the scale: "These results support the WMS-III immediate and delayed auditory memory, immediate and delayed visual memory, and working memory dimensions." (The Psychological Corporation, 1997, p. 115)

Millis, Malina, Bowers, and Ricker (1999) attempted to replicate the factor structure of the WMS-III in the entire standardization sample using data derived from the correlation matrices reported in the manual. In contrast to the results provided by the Psychological Corporation (1997), they reported that models specifying separate immediate and delayed memory factors were hampered by inadmissible solutions signaling model specification error. They concluded that a three-factor model, with separate working memory, auditory memory, and visual memory dimensions, provided the best fit of the data for the entire standardization sample. Rigorous re-analyses⁵, with precise replication of the models and procedures outlined in *The WAIS-III—WMS-III Technical Manual*, have confirmed that the three-factor structure described by Millis et al. provides the best fit to the WMS-III standardization data (Price et al., 2002).

⁵ Millis et al. (1999) noted that the degrees of freedom for their models did not correspond to those in the *WAIS-III—WMS-III Technical Manual* and suggested that the test developers may have estimated additional parameters that were not reported. This is thought to be because Millis et al. included the Auditory Recognition subtest in their methods, whereas this was not the case in the analyses reported in the *WAIS-III—WMS-III Technical Manual* (L. R. Price, personal communication, August 2001). These model differences may provide one explanation for the discrepant results. Because of the dependence of Auditory Recognition on the other two auditory subtests and the skewed properties of this task, it was not included in the present analyses (following the methods of Price et al., 2002, and The Psychological Corporation, 1997).

As Millis et al. (1999) pointed out, the elucidation of the test's structure is rendered more complex given that components of memory may vary as a function of cerebral dysfunction and, thus, may change a test's factor structure from one group to another. Therefore, it is important to investigate whether the factor structure found in the standardization sample generalizes to other samples. For example, distinctions between immediate and delayed memory constructs might emerge in particular clinical samples (e.g., patients with dementia), and individuals with lateralized cerebral damage might show differential performance in visual and auditory domains. However, at present there is very limited data available on WMS-III performance in clinical groups.

As memory deficits are endemic to patients with temporal lobe epilepsy, valid assessment of memory functioning in these individuals is of critical importance. For one, distinctions between auditory and visual memory factors may aid in the differentiation of lateralized memory deficits. Patients with temporal lobe seizures often exhibit material-specific memory impairments, with left-temporal dysfunction being associated with verbal memory impairments and right-temporal disturbance linked to visuospatial memory deficits (e.g. Delaney et al., 1980; Hermann et al., 1987; Loring et al., 1988; Milner, 1975). The separation of auditory and visual factor scores on the WMS-R has been referred to in the context of supporting theories about hemispheric specialization and material specificity of memory deficits (Bornstein & Chelune, 1988; Jurden et al., 1996; Moore & Baker, 1997), and the argument is frequently made that separable visual/auditory factor scores may emerge particularly in populations with lateralized impairment. Although distinct auditory and visual factors on the WMS-III were supported in the normative sample (Millis et al., 1999; Price et al., 2002), it is important to show that this is also the case in patients with temporal lobe epilepsy. Only then can clinicians feel confident in interpreting the Auditory and Visual

Index scores on the WMS-III as representing separate dimensions in this population. Such information is often critical in a clinical context, particularly if patients undergo surgical resection for intractable temporal lobe epilepsy. It is important for neuropsychologists to develop reliable and valid methods for detecting impairment and for identifying individuals who may be at increased risk for cognitive impairment after surgery (Jones-Gotman et al., 1993).

The current study constituted an initial attempt to investigate the generalizability of the WMS-III factor structure described by Millis and colleagues (Millis et al., 1999; see also Price et al., 2002) in a well-defined clinical group. Using CFA procedures, the five models investigated in the standardization sample (The Psychological Corporation, 1997) were evaluated:

Model 1 (M ₁ ; one factor):	single, general memory factor
Model 2 (M ₂ ; two factors):	working memory, general memory
Model 3 (M ₃ ; three factors):	working memory, immediate memory, delayed memory
Model 4 (M ₄ ; three factors):	working memory, auditory memory, visual memory
Model 5 (M ₅ ; five factors):	working memory, auditory immediate memory, visual immediate memory, auditory delayed memory, visual delayed memory

Given the nature of the sample in the current study, that is, most with lateralized temporal lobe disturbance, as well as the findings by Millis et al. (1999) and Price et al. (2002), it was expected that Model 4 would provide the best fit of the five models evaluated.

Method

Participants

This study took advantage of a unique collaborative project (the Bozeman Epilepsy Consortium) that has combined neuropsychological data from a number of epilepsy surgery

centers. The subject pool was selected from this database of patients with temporal lobe epilepsy and medically refractory seizures provided by four epilepsy centers in the Bozeman Consortium. Patients were considered for inclusion in this study if they met the following criteria: (a) seizure onset was of temporal lobe origin (the procedures for diagnosis are described later); (b) information was available regarding patient age, sex, and education; and (c) they had received neuropsychological evaluation including the WMS-III.

The following data were also collected to better characterize the sample: Wechsler Adult Intelligence Scale— Revised (WAIS-R; Wechsler, 1981) or Wechsler Adult Intelligence Test— Third Edition (WAIS-III; Wechsler, 1997a) Full Scale IQ, handedness, age of onset of recurrent seizures, and seizure focus (i.e., left/right/bilateral temporal lobe seizure origin); however, to maximize the sample size for the purposes of CFA, patients were not excluded on the basis of missing data for these demographic or core epilepsy characteristics.

The final sample consisted of 254 patients with a diagnosis of temporal lobe epilepsy from the following four centers⁶: Cleveland Clinic Foundation (CCF; $n = 54$), Medical College of Georgia (MCG; $n = 78$), University of Alabama at Birmingham Epilepsy Center (UAB; $n = 44$), and University of Wisconsin— Madison (UWM; $n = 78$). The characteristics of the patients classified by center are provided in Table 3.1.

The presence of preexisting differences across centers was examined with separate analyses of variance for age, education, IQ, and age of onset. Chi-square analyses were conducted to evaluate differences in the distribution of sex, handedness, and seizure focus. Center differences were found for sex distribution ($p < .01$) and IQ ($p < .01$), although, as might be expected in a sample of patients with medically refractory epilepsy (e.g., Strauss et

⁶ Data contributed from all centers were extracted from deidentified patient registries that were reviewed by their respective institutional review boards. Note that a proportion of this patient sample ($n = 102$) is included in Study 1.

al., 1995), the average IQ was shifted downward from the normative mean in all centers. There were no differences between centers for the other demographic or epilepsy variables. The mean age of the combined sample was 35.6 years ($SD = 11.4$), and the average education level was 12.7 years ($SD = 2.3$). There were 104 men and 150 women. For those patients for whom laterality information was available ($n = 176$), seizure origin was localized to the right temporal lobe in 43.8% of cases, whereas 51.7% showed left temporal seizure onset. Eight patients had bilateral disturbance (4.5%). Information regarding laterality was unavailable in the remainder of the sample ($n = 78$).

Table 3.1

Demographic and Clinical Epilepsy Characteristics for the Total Sample and Individual Centers

Center	<i>n</i>	Age (years)		Education (years)		Full Scale IQ ^a		Gender (%male)	Hand. (% Right)	Age of Onset		Seizure Focus (%Left)
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			<i>M</i>	<i>SD</i>	
CCF	54	33.0	10.5	12.8	2.3	91.0	12.9	53.7	85.2	13.2	10.3	46.3
MCG	78	34.9	11.6	12.6	2.3	85.0	15.0	53.8	89.3	17.7	9.5	53.8
UAB	44	37.0	11.1	12.6	2.6	79.9	15.1	25.0	93.2	18.2	14.5	54.5
UWM	78	37.3	11.6	12.9	2.3	93.7	16.6	28.2	--	15.6	11.4	--
Total		35.6	11.4	12.7	2.3	88.2	15.9	40.9	89.0	15.6	11.9	51.7
Valid <i>n</i> ^b		254		254		244		254	173	182		176

Note. Dash indicates data were not available. CCF = Cleveland Clinic Foundation; MCG = Medical College of Georgia; UAB = University of Alabama—Birmingham Epilepsy Center; UWM = University of Wisconsin – Madison; Hand. = hand preference.

^a IQ scores were based almost exclusively on the standardized administration of the WAIS–III, although in some cases IQ scores were derived based on the WAIS–R. ^b *ns* differ due to cases of missing data for some demographic and epilepsy variables.

Procedures

All patients underwent comprehensive evaluations according to the protocols in place at their respective epilepsy centers, which typically included clinical examination, extensive

invasive and noninvasive electrophysiological monitoring, functional and structural neuroimaging, intracarotid amytal procedure, and neuropsychological assessment.

Diagnoses of temporal lobe epilepsy were confirmed by epileptologists following accepted procedures for the diagnosis and localization of seizures as determined by a consensus conference of epilepsy experts (see Engel, Jr., 1993).

Participants were administered the WMS-III as part of comprehensive neuropsychological evaluations. Tests were administered and scored by trained personnel according to standardized procedures provided in the WMS-III manual (Wechsler, 1997b).

Model fit was examined on the basis of performance on the WMS-III Primary Subtests: Logical Memory I and II, Verbal Paired Associates I and II, Faces-I and II, Family Pictures I and II, Letter-Number Sequencing, and Spatial Span⁷. Age-corrected scaled scores served as the unit of analysis. The mean subtest scores for the sample are reported in Table 3.2, and the subtest intercorrelations are shown in Table 3.3. Given the severity of temporal lobe seizure disturbance in this sample, it is not surprising that performance on the WMS-III subtests falls in the low-average range in comparison to normative standards (Wechsler, 1997b).

Statistical Analysis

CFA was conducted with AMOS Version 4.0 (Arbuckle & Wothke, 1999) using maximum likelihood estimates derived from the covariance matrix. Subtests were constrained to load on only a single factor, but correlated errors were permitted between the immediate and delayed measures of the same subtests. Factor correlations were not restricted in any of the models. Several statistics were used in evaluating and comparing the

⁷ Missing data points, which comprised less than 1% of the data and appeared randomly distributed, were replaced with the mean of the sample.

In addition to the conventional chi-square test, which is affected by sample size and violations from multivariate normality (Hu & Bentler, 1995), the following model fit statistics were examined: (a) the goodness-of-fit index (GFI; Jöreskog & Sörbom, 1981), (b) the nonnormed fit index (NNFI; Bentler & Bonett, 1980, also known as the Tucker-Lewis index), (c) the comparative fit index (CFI; Bentler, 1990), (d) the root-mean-square error of approximation (RMSEA; Steiger, 1990), (e) the Bayes information criterion (BIC), and (f) the expected value of the cross validation index (ECVI; Browne & Cudeck, 1989; Browne & Cudeck, 1993).

For GFI, NNFI, and CFI, higher values are desired, with values above .90 indicating good model fit. RMSEA provides a measure of the discrepancy between elements of the model fitted to the sample and the model fitted to the population covariance matrix. Although exact fit to the model would be indicated by a $RMSEA = 0$, values less than .08 indicate reasonable model fit, and values less than .05 indicate a close fit in relation to degrees of freedom (Browne & Cudeck, 1993). BIC and ECVI are used for model comparison, with lower values indicating better fit. When comparing competing nested models that have reasonable fit to the data, BIC tends to favor simpler, more parsimonious models (Arbuckle & Wothke, 1999). A BIC difference of 5 points provides strong evidence whereas a value above about 10 provides near conclusive evidence of model fit in favor of the model with the lower BIC value (Raftery, 1993). The ECVI is useful for comparison of alternative models, especially when the sample size is not large, providing an indication of which model yields a solution with greatest generalizability (MacCallum & Austin, 2000). Statistical comparisons between models were also made with chi-square tests when one model was nested within another.

To evaluate goodness of fit at the level of individual equations within a model, parameter estimates (i.e., covariances, variances of latent variables, error variances, residual moments) were inspected to evaluate their statistical plausibility and to screen for model specification errors (e.g., Heywood cases, standardized correlation estimates greater than unity). The reliability of parameter estimates was also examined, as the global fit of a model may be very good even if one or more individual parameter estimates are not reliably different from zero (Breckler, 1990). Finally, a squared multiple correlation was computed for each structural equation, which indicates the proportion of variance in each latent endogenous variable accounted for by the equation (Breckler, 1990).

Results

As described above, five models were initially evaluated. However, the solutions for those models that specified separate immediate and delayed factors (M_3 and M_5) were hampered by inadmissible estimates and model specification errors. The solution for M_3 (working memory, immediate memory, and delayed memory) contained negative error variances, whereas the implied covariance matrices for variables in M_5 (working memory, auditory immediate memory, visual immediate memory, auditory delayed memory, visual delayed memory) were not positive definite (i.e., contained negative eigenvalues), which raises questions about the validity of the remaining estimates in the model (Hoyle & Panter, 1995). The boundary solution errors are reflective of the high correlation estimates between the immediate and delayed factors specified in these models; these ranged from .98 to 1.00. In addition, examination of the intercorrelations among the WMS-III subtests (see Table 3.3) revealed that the observed relationships among the subtests were inconsistent with the factor-loading patterns specified by Models 3 and 5. The results mirror those found by Millis

et al. (1999) in that correlations among subtests of different factors were larger than the correlations between subtests of the same factor. This is a strong indication of model specification error—that the models were not supported by the data (Wothke, 1993). The results of these analyses do not support the viability of models with separate immediate and delayed dimensions, and thus the fit indexes that resulted from estimation of these models (M_3 and M_5) were not further evaluated. A summary of the fit statistics for the three remaining models based on the covariance matrix of the 10 primary subtests is shown in Table 3.4.

Table 3.4
Goodness-of-Fit Statistics

Model	$\chi^2 (N = 254)$	<i>dfs</i>	GFI	NNFI	CFI	RMSEA (90% CI)	BIC	ECVI
M_1	132.29*	31	.90	.92	.95	.11 (.09 - .13)	320.45	.71
M_2	79.98*	30	.94	.96	.97	.08 (.06 - .10)	275.98	.51
M_4	70.65*	28	.95	.96	.98	.08 (.06 - .10)	282.33	.49

Note. M_1 has a single factor. M_2 has working memory and general memory factors. M_4 has working memory, auditory memory, and visual memory factors. The solutions for M_3 and M_5 produced inadmissible model estimates and are not reported. GFI = Goodness-of-fit index; NNFI = Nonnormed fit index; CFI = Comparative fit index; RMSEA = Root mean squared error of approximation; BIC = Bayes information criterion; ECVI = Expected cross-validation index.
* $p < .001$.

The chi-square test indicated significant discrepancies between the observed and implied covariance matrices for all models, suggesting that none of the models provided an exact fit of the data. However, the values for the other omnibus fit indexes clearly indicate that fit is better for a two-factor model (M_2 ; working memory and general memory) than for a single-factor model (M_1), and there was a large chi-square difference between M_2 and M_1 , $\chi^2_{diff}(1, N = 254) = 52.31, p < .01$. However, little difference was obtained in the fit values for M_2 and M_4 (working memory, verbal memory, visual memory), with the GFI, NNFI, and CFI

indexes all exceeding .93 indicating acceptable fit to the data. Although a comparison between the chi-square values for the two models favored M_4 , $\chi^2_{diff}(2, N = 254) = 9.33, p < .01$, other comparison indexes (RMSEA, ECVI) did not provide convincing support for this model. In fact, the BIC value for M_2 was *lower* than for M_4 , providing “strong” evidence of better fit for M_2 according to Raftery’s (1993) guidelines. Thus, although M_4 cannot be discounted, the more parsimonious M_2 seems similarly able to account for the data in this sample.

Table 3.5

Standardized Parameter Estimates for Model 2

WMS-III Subtest	Working Memory	General Memory	R^2
Logical Memory I		0.75	0.55
Faces I		0.42	0.18
Verbal Paired Associates I		0.79	0.63
Family Pictures I		0.67	0.45
Letter-Number Sequencing	0.83		0.69
Spatial Span	0.76		0.57
Logical Memory II		0.77	0.59
Faces II		0.48	0.23
Verbal Paired Associates II		0.75	0.57
Family Pictures II		0.72	0.52

A number of factors were investigated at the level of the individual equation to further evaluate the plausibility of M_2 . The standardized parameter estimates for M_2 are presented in Table 3.5. All regression weights differed significantly from zero ($p < .01$). Standardized residual covariances indicate the difference between the sample covariances and the implied covariances from the model (divided by an estimate of its standard deviation). Thus, if the model is correct, each standardized residual covariance should be less than about two in

absolute value (as the critical value for $\alpha = .05$ is ± 1.96). No standardized residual covariances in M_2 exceeded this level. Finally, the correlation between the working memory and general memory factors of M_2 was .70, indicating strongly related, although partially divergent, constructs.

One additional issue was explored. As reported in Millis et al. (1999), examination of the parameter estimates for the Faces subtests revealed that they were uniformly low for all the models examined, even in those models where a separate visual memory factor was proposed. Inspection of the intercorrelations between the subtests (Table 3.3) also reveals that Family Pictures has larger correlations with subtests other than Faces, and thus, it is not surprising that these two subtests did not combine to produce a coherent visual memory factor. In any given study, a particular latent variable is effectively defined as that which its indicators have in common. The nature of the construct can shift with the choice of indicators, which in turn can influence results and interpretation (MacCallum & Austin, 2000). It was felt that the low R^2 values for Faces and the unexpected pattern of intercorrelations obtained for this subtest may have hampered global model-fit estimates. Thus, the fit of the five models after removal of the Faces subtest was examined to explore this possibility.

The fit indexes that resulted from these analyses led to essentially the same results as reported above. That is, M_3 and M_5 produced inadmissible results, and M_2 and M_4 provided better fit estimates than the single-factor model. As above, the BIC value for M_2 was lower than for M_4 , indicating better fit for the two-factor model. Furthermore, in this case the chi-square comparison between M_2 and M_4 was not significant, $\chi^2_{diff}(1, N = 254) = 0.16, ns$, indicating that the more complex model (M_4) was no better able to reproduce the data matrix. Thus, when the Faces subtests were not included, M_2 provided the most

parsimonious solution and the best fit of the data of the five models evaluated. The low observed correlations between the Family Pictures and Faces subtests do not appear to account for the lack of superiority of M_4 over M_2 .

Discussion

The WMS-III represents the most recent revision of the widely used scales of learning and memory. The scale endeavors to assess immediate and delayed memory of auditory and visually presented material, through free recall and recognition procedures. Millis et al. (1999) and, more recently, Price et al. (2002) have evaluated the factor structure of the WMS-III in the standardization sample and have concluded that a three-factor model (working memory, auditory memory, visual memory) provided the best fit of the data. Thus, the multifactorial nature of the WMS-III was substantiated, although the distinction between immediate and delayed factors reported with the WMS-R was not supported statistically. This may be due to the high correlation between the immediate and delayed conditions of the same memory tasks on the WMS-III, which may not be expected to covary significantly in the normative samples that were used in these studies (Price et al., 2002). As score distinctions, and hence the underlying factor structure of the scale, may vary from one group to another, these researchers described a pressing need for the exploration of the WMS-III factor structure in clinical groups (Millis et al., 1999; Price et al., 2002).

To this end, the purpose of the current study was to investigate the generalizability of the WMS-III factor structure in a clinical population of individuals with temporal lobe epilepsy. Of the five models evaluated, a two-factor model composed of working memory and general memory factors seemed to provide the best fit and the most parsimonious solution. Although these findings are not contradictory to that reported in the standardization group,

a distinction between auditory and visual memory constructs was not clearly supported in this sample. This was somewhat unexpected given the many individuals in this sample with lateralized dysfunction, for which one might have predicted separate auditory and visual memory factors. Indeed, exploratory factor analysis of the WMS-R in patients with intractable epilepsy revealed a three-factor structure with separate attention/concentration, verbal memory, and visual memory factors (Moore & Baker, 1997). The fact that specifying a separate visual memory factor provided little improvement in model fit in the current study may be related to properties of the two “visual” tasks.

New visual subtests were developed for the WMS-III in response to criticisms that the WMS-R visual memory subtests were inadequate measures of a pure visual memory system and were not differentially sensitive to unilateral lesions (Chelune & Bornstein, 1988; Heilbronner, 1992; Loring, 1989). In the standardization sample, the new visual subtests of the WMS-III appear to tap a visual memory construct to a greater extent than those of the WMS-R, at least as suggested by the factor analytic results, yet Millis and colleagues (1999) identified concerns with the visual factor. In their analysis, the Faces subtests had insufficient commonality with Family Pictures, which is especially problematic as these two subtests alone compose the visual memory index scores. This finding was also noted by Price et al. (2002) and replicated in this clinical sample. The WMS-III Primary Indexes are intended to be the principal scores used to evaluate memory functioning and are meant to summarize overall performance in a particular domain (The Psychological Corporation, 1997). The fact that these two subtests do not form a coherent factor renders the Visual Indexes suspect. As Kaufman and Lichtenberger (1999) have indicated with regard to interpretation of the WAIS-III factor scores, when subtest scores that contribute to factor-based indexes are markedly discrepant, they may not provide adequate estimates of the constructs being

measured. It should be noted that a construct consisting of only two indicators is a general problem that has been recognized for some time (e.g., Harman, 1967). Fundamentally, this is because a centroid's location is not uniquely defined and hence not determinable from two points alone (see Little, Lindenberger, & Nesselroade, 1999 for review of indicator selection influences on construct representation). This may negatively impact fit estimates for models such as Model 5, in which four of five factors include two indicators.

An additional reason for the poor specification of a separate visual memory factor may be related to the properties of the Family Pictures subtest. This task is clearly not an exclusively visual memory task in that it requires patients to verbally recall as well as visually encode the picture contents. The fact that dual encoding and processing is required may account at least in part for the correlations between Family Pictures and the various auditory subtests seen in the standardization data (see *The WAIS-III—WMS-III Technical Manual*, Appendix A, for WMS-III subtest intercorrelation tables) and in this clinical sample. Further support comes from a recent investigation of memory performance in patients with epilepsy, which showed that auditory memory ability accounted for the majority of the variance in the Family Pictures task (Dulay, Testa, Fargo, Eschleman, & Schefft, 2001). It is tempting to contemplate whether this relationship might be more characteristic of some clinical populations, which may reveal interesting insight into the cognitive bases underlying subtask performance and their interrelationships. For example, one might speculate that patients with right-sided temporal dysfunction show higher correlations between Family Pictures and the auditory tasks, reflecting an increased reliance on verbal mediation as a compensatory mechanism. Unfortunately, this cannot be addressed with the present data because laterality information is not readily available from all participants in the sample.

The question arises whether either the Faces or the Family Pictures subtests provides reasonable measures of visual memory. As discussed, there exist substantial lines of evidence for an auditory-verbal component to the Family Pictures subtest. One may put forth the argument that the lack of commonality of the two tasks may, in fact, reflect the superiority of Faces as a measure of visual memory. Without some external criterion, no firm conclusions about this task's theoretical or clinical utility can be reached. However, the *WAIS-III—WMS-III Technical Manual* indicates that the correlations between Faces and Visual Reproduction, presently a supplemental subtest of the WMS-III, are quite low (around .20). The Faces task was also not found to contribute to differentiating laterality of disturbance in patients with temporal lobe epilepsy (Wilde et al., 2001). Therefore, in addition to the factor analytic results, evidence is accumulating that raises questions about the meaning of this subtest. Further research on the Faces task both pre- and post-temporal lobectomy is required to determine whether it might provide meaningful information regarding outcome in an epilepsy sample.

At least in a temporal lobe epilepsy sample, the findings of this study suggest caution in interpretation of the WMS-III Auditory and Visual Index scores as representing separable dimensions. Similarly, any clinical interpretations of differences between the scores must remain tenuous—it is difficult to interpret patterns of scores that are unrelated to the test's factor structure, particularly when they do not discriminate between contrasting groups (see Wilde et al., 2001). Of course, even though the data cast doubt on the test's utility to characterize auditory and visual memory in this population, the clinical utility of the WMS-III need not depend on this multifactorial structure. It may still be the case that scores on the WMS-III provide useful inferences regarding patients' daily functioning or information regarding risk for cognitive impairment after surgery.

It is obvious that the factor structures revealed in this clinical sample are intimately linked to the nature of the patients who are included, and studies based on other homogeneous patient samples may yield different results. Also, one might question how representative the sample is of epilepsy populations in general. This large sample, derived from multiple centers in North America that each cater to a broad regional area, is considered to provide a good characterization of patients with medically refractory seizures of temporal lobe origin. However, it is important to bear in mind that these are a highly selective subgroup of patients with temporal lobe epilepsy, and caution should be used in generalizing these findings to patients in whom seizure control is achieved with medication.

A limitation of this study is that analysis was limited to the primary subtests. Studies that specify structural models including the primary and supplementary subtests are needed. Further evaluation of the psychometric properties of the supplemental subtests may lead to the creation of a more coherent and statistically sound visual memory factor. Based in part on the factor analyses with the standardization sample, new visual memory indexes are being developed to combine Family Pictures with the Visual Reproduction subtests, instead of the Faces task (Price et al., 2002). It is yet unclear whether the factor structure using this combination will provide improved parameter estimates for a visual memory factor in this population. Another drawback of this study was that the covariance matrix used for model estimation was based on scaled scores. Future research will determine whether analysis based on raw scores leads to better elucidation of the factor structure of the WMS-III in clinical groups.

One other issue deserves comment. The problem with factor analysis in clinical populations is that it is difficult to disentangle whether results relate to the test itself or to the specific population under investigation. Given the sizeable number of patients with

lateralized disturbance, it is unlikely in this case that the poor evidence for a visual memory index is because patients did not show measurable visual memory deficits. However, in the absence of a gold standard of visual memory, this issue remains an open question.

In summary, five competing structural models specifying the relationships among the WMS-III primary subtests were evaluated in a sample of individuals with temporal lobe epilepsy. Of the five models, a two-factor model with separate working memory and general memory factors provided the best and most parsimonious solution. Thus, although support was found for separate auditory and visual memory factors in the standardization group, convincing evidence in favor of such a model was not found in our clinical sample. The results of this study cast doubt on the interpretation of Auditory and Visual Memory Index scores and on discrepancies between them, at least in an epilepsy population (see also Wilde et al., 2001). First, the Faces subtest appears to have insufficient commonality with Family Pictures— which is problematic as these are the only two subtests to compose the Visual Memory Index scores. Second, there is insufficient evidence from the present study that the visual memory subtests are tapping a construct other than that measured by the auditory subtests. It remains to be seen whether these findings are replicated in other patient groups.

Chapter 4

GENERAL DISCUSSION

The Wechsler Memory Scale (WMS) is the most extensively used battery for memory assessment of adults (Rabin et al., 2005). The third edition of the WMS (WMS-III; Wechsler, 1997b) represents a substantial revision of previous versions. Accordingly, issues of validity of the revised instrument need to be addressed. The purpose of these studies was to contribute to the validation of the scale in the assessment of patients with temporal lobe epilepsy (TLE).

An important role of the neuropsychological evaluation in TLE is to aid in the localization and lateralization of dysfunction (Helmstaedter, 2004). This is based on the premise that the temporal lobes are specialized for the acquisition of material-specific information, with dysfunction in the left and right mesial temporal regions being associated with verbal and nonverbal memory impairment, respectively. This premise has been supported by evidence from a number of investigative methods, including the evaluation of patients with lesions, results from the intracarotid amobarbital procedure (IAP), evaluation of patients with temporal lobe epilepsy both preoperatively and following anterior temporal lobectomy⁸ (ATL), and functional neuroimaging studies. However, it should be noted that the association between left-sided TLE (LTLE) and auditory-verbal memory has been a much stronger and consistent finding than the association between right-sided TLE (RTLE) and visual memory. This issue is discussed in greater detail below. Since the WMS (and its revisions) is utilized by the vast majority of epilepsy centres as part of presurgical

⁸ Although the conventional term *lobectomy* is used throughout this document, it should be noted that many centres now perform more restrictive or selective temporal resections.

neuropsychological evaluations of epilepsy patients (Jones-Gotman et al., 1993), evaluation of its meaning and utility in this population is essential. Two studies were conducted.

The purpose of Study 1 was to utilize criterion validating procedures to evaluate the ability of the WMS-III in detecting left versus right TLE. To the extent that the test taps distinct auditory and visual memory processes, it was predicted to be able to accurately identify those with lateralized temporal impairment both at the group level as well as at the level of the individual patient. Auditory-Visual index discrepancy scores appeared most sensitive to side of temporal dysfunction at the group and individual levels. However, group means on the WMS-III indexes and subtests were largely insensitive to laterality, and patient classification rates based on individual index, subtest, or discrepancy scores revealed poor diagnostic accuracy. The ability to predict laterality based on statistically significant index score differences was particularly weak for those with left temporal dysfunction. Classification accuracy improved when unusually large discrepancies were used; however, discrepancies of such magnitude occurred very rarely limiting the overall sensitivity of such an approach.

Study 2 investigated the latent structure of the WMS-III in a sample of patients with TLE. Five models based on the covariance matrix of the primary subtest scores were evaluated. These models were initially evaluated in the standardization sample and correspond to the hierarchical index structure of the WMS-III. The findings revealed negligible goodness-of-fit differences between a 3-factor model of working memory, auditory memory, and visual memory, and a nested—more parsimonious—2-factor model of working memory and general memory. While the 3-factor model provided the best fit in the standardization sample, the results suggested that specifying a separate visual memory factor provided little advantage for this TLE sample. This was an unexpected finding in this

population with lateralized dysfunction, for which one might have predicted separate auditory and visual memory dimensions. Furthermore, models specifying separate immediate and delayed constructs resulted in inadmissible parameter estimates signaling model specification error.

Taken together, these studies raise concerns about the meaning of the separate Auditory and Visual WMS-III scores in a TLE population and the interpretation of discrepancies between them.

Additional Studies on the WMS-III in TLE

Baker, Austin, and Downes (2003) conducted a very similar study to Study 1 and reported congruent findings. In their sample of 99 preoperative patients, the LTLE and RTLE groups were found to differ only on the Auditory Immediate and Delayed indexes. This was driven by a difference in Logical Memory, while Verbal Paired Associates was marginally significant. The groups did not differ on the visual memory indexes or subtests. Within group comparisons revealed no differences between the Auditory and Visual indexes in the left TLE group, although the right TLE group did show a difference in the expected direction. Not surprisingly, classification based on discrepancy scores between the Auditory and Visual indexes were particularly unfruitful for the left TLE group, with correct classification rates ranging from 6 to 24% of the sample depending on the size of the discrepancy scores. The right TLE group fared slightly better, with correct classification rates between 19 and 50%. Nonetheless, the ability of the WMS III to correctly classify the side of pathology was severely limited.

There are a number of possibilities for the relatively poor global performance on the WMS-III in LTLE patients as demonstrated in Study 1 and by Baker et al. (2003). For one, the possibility that patients classified as LTLE may have also had right temporal

abnormalities that went undetected cannot be excluded. Second, LTLE patients may exhibit more cognitive impairment than RTLE patients, that is not restricted to verbal memory (e.g., Chelune, Naugle, Luders, & Awad, 1991; Strauss et al., 1995). This may be related to cerebral reorganization due to early insult to the left hemisphere (as reflected in higher incidence of atypical language organization and left-handedness in LTLE) and concomitant crowding phenomena (Strauss, Satz, & Wada, 1990). Finally, it may also be that the WMS-III visual memory tests are confounded by auditory-verbal memory with resultant reliance on left temporal structures, and thus also impaired in patients with deficits in such skills.

Support for this latter statement comes from the findings of Bell and colleagues (Bell, Hermann, & Seidenberg, 2004), whose goal was to evaluate the utility of the distinction between the Immediate and Delayed index scores on the WMS-III in patients with TLE (see discussion of this below). These authors did not specifically report results by side of seizure focus so the ability of the WMS-III scores to predict laterality in their sample cannot be determined. However, they evaluated the correlation between the WMS-III scores and hippocampal volumes in light of the established relationship between hippocampal integrity (particularly on the left) and memory performance (Bell & Davies, 1998). They found that the WMS-III Auditory Immediate and Delayed indexes correlated with left but not right hippocampal volume, whereas the Visual indexes correlated significantly with both left and right hippocampal volumes. This is consistent with the findings of Study 1 in which LTLE patients as a group performed poorly on both Auditory and Visual WMS-III indexes whereas the memory deficit in RTLE was more specific to the Visual indexes.

Furthermore, the *WAIS-III – WMS-III technical manual* (The Psychological Corporation, 1997) reported descriptive data for a small sample of patients who underwent surgery for temporal lobe epilepsy. While a relative double dissociation in the anticipated direction was

reported (i.e., lower Auditory than Visual index scores for the left-sided group, and vice versa for the right-sided group), absolute Visual index scores were comparable in both groups.

Doss and colleagues (Doss, Chelune, & Naugle, 2004) have published the first large scale study investigating performance on the WMS-III in patients following temporal lobectomy. As Doss et al. point out, a post-operative investigation may be well suited to studying the validity of instruments purporting to measure modality-specific abilities, such as the WMS-III. Their results indicated that the right temporal lobectomy (RTL) group scored significantly lower on the Visual indexes than the Auditory indexes, and lower than the left temporal lobectomy (LTL) group's performance on the Visual indexes, while the LTL group showed the opposite pattern. Thus, a differential pattern of performance was obtained between and within groups, which was in the expected direction and supported the material-specific distinction of the WMS-III associated with temporal lobectomy.

Material-specific memory impairments have been shown more consistently in patients following temporal lobectomy than in preoperative epilepsy patients. Thus, it is not surprising that the results of Doss were more favourable than those of Study 1. Doss et al. do not present preoperative data on their patients, so it is difficult to know how comparable their preoperative results would have been to those reported in Study 1. However, both patient groups were selected from patients at epilepsy surgery centres with consistent procedures, and the largest proportion of both samples were actually from the same centre (Cleveland Clinic Foundation). Assuming that performance in both samples would be roughly comparable, inspection of the mean scores in the preoperative group (Study 1) and the postoperative group (Doss et al., 2004) revealed some interesting findings. Rather than exhibiting a significant decline in their respective modalities following surgery (i.e., auditory

for LTL and visual for RTL), group differences appeared to be mediated largely by increases in performance in each group on one particular subtest: Verbal Paired Associates for the RTL group and Faces for the LTL group. Not surprisingly, Doss et al. reported that these two subtests best distinguished group membership via discriminant function analysis. Due to the effects of practice, the absence of positive gain may actually reflect a decrement in performance in some cases (e.g., McCaffrey, Ortega, & Haase, 1993). Evaluation of test-retest scores in a nonoperated sample of epilepsy patients found statistically significant practice effects on Faces, but not the other WMS-III memory subtests (Martin et al., 2002). Thus, the better performance in the LTL group may be due to practice effects on the Faces task, with the lack of improvement in the RTL group reflecting a decrement (see also Chiaravalloti & Glosser, 2004; Chiaravalloti, Tulsky, & Glosser, 2004). This does not explain the improved performance by the RTL group on Verbal Paired Associates however. “Release” of functions associated with successful seizure control is also a possibility (Helmstaedter, Kurthen, Lux, Reuber, & Elger, 2003). A longitudinal, systematic evaluation of LTLE and RTLE patients pre and post temporal lobectomy is needed to better evaluate the effects of temporal lobectomy on the WMS-III and the different effects in these two groups.

Visual Memory and the Right Temporal Lobe

Much of the evidence to support the notion of material specificity of the temporal lobes comes from studies of patients undergoing ATL for the treatment of medically refractory epilepsy (Milner, 1968b). The specialized contribution of the left mesial temporal lobe for processing and encoding verbal information has been well documented in the literature and is now widely accepted.

In contrast to verbal memory, findings have been less consistent with regard to the role of the right mesial temporal lobe in nonverbal or visuospatial memory. A number of studies have shown visuospatial memory deficits following ATL in the right temporal lobe, evident in impaired learning of new, unfamiliar faces (Milner, 1968b), abstract designs (Jones-Gotman, 1986; Helmstaedter, Pohl, Hufnagel, & Elger, 1991) and object-spatial relationships (Milner, 1965). However, other studies have not found an association between RATL and visuospatial memory deficits as assessed by figural memory (Barr et al., 1997; Barr, Morrison, Zaroff, & Devinsky, 2004), facial recognition (Hermann, Connell, Barr, & Wyler, 1995; Naugle et al., 1994) or memory for spatial location (Barr, 1997a; Malec et al., 1992). Failure to find consistent deficits in visuospatial memory associated with right mesial temporal lobe dysfunction may be due to a number of factors (see also Heilbronner, 1992).

As Helmstaedter (2004) notes, the inconsistent relationship between visual memory and right temporal lobe pathology may be due in part to the effects of sex differences and atypical language dominance. For example, in a group of patients with LTLE, figural memory impairment was evident in women with typical language dominance (but not men), and in men and women with atypical language dominance (Helmstaedter, Kurthen, & Elger, 1999). The finding of visual memory impairment in those with left-sided TLE is consistent with the results of Study 1, and supports the notion that visual memory impairment is not seen exclusively in those with right-sided dysfunction. Therefore, any study in which the evidence of visual memory deficits in RTLE patients is evaluated based on comparisons to LTLE performance would be affected by such a finding.

There are also confounding effects of verbalization on visual memory. For example, Eadie & Shum (1995) compared retention for abstract visual stimuli in patients with unilateral brain lesions, and found that greater impairments were found in the right

hemisphere lesion group with the material that was rated as less verbalizable. Verbal mediation of figural designs has been shown to facilitate performance in normals, and attenuating verbal mediation via dual task interference impedes such memory for designs that are relatively easy to verbalize (Silverberg & Buchanan, 2005).

The Family Pictures subtest appears highly susceptible to verbalization at encoding, and requires a verbal response for recall. It is likely that this “visual” memory test is confounded by auditory-verbal memory and thus would be expected to be reliant on the integrity of both temporal lobes. This is supported by the findings of Study 1 showing no differences between the RTLE and LTLE group on this subtest, and the correlations of the subtest with left hippocampal and lateral temporal lobe white matter volumes as reported by Bell et al. (2004). This may also contribute to the lack of a coherent visual memory factor as reported in Study 2, in which Family Pictures showed stronger correlations with the two auditory subtests than with the Faces task. Similarly, in a sample of patients with epilepsy, performance on the Family Pictures subtest was best predicted by Logical Memory performance (Dulay et al., 2002). These authors concluded that the test relied heavily on auditory-verbal abilities and suggested the use of these scores would be misleading when using the test as a lateralizing index in patients with epilepsy.

While these confounds may be one reason for the failure to find consistent deficits in visual memory associated with right mesial temporal lobe dysfunction, it has been suggested that difficulties in documenting the existence of material specific impairments in nonverbal memory may be due to the lack of clarity of the entire construct of nonverbal memory (Barr et al., 2004). Studies on the role of the right temporal lobe in visual memory have generally focused on a particular task or paradigm and have not been driven by hypotheses based on a coherent theoretical construct (Barr, 1997a). It has been suggested that drawing on the

visual cognition and neurosciences literatures may provide a more informative framework for evaluating these questions (Barr, 1997a; Chiaravalloti & Glosser, 2004).

Specifically, it is well established that visuospatial processing is not a unitary construct. It subsumes different cognitive procedures, which are localized to different cerebral regions. For example, a large body of research indicates two separable cortical pathways for the processing of visuospatial information: an inferior pathway for the visual identification of objects, and a superior pathway for the processing of spatial location (Mishkin, Ungerleider, & Macko, 1983).

Given the existence of these distinct visuospatial processing systems, it is also possible that only certain visuospatial memory functions will be impaired in TLE and following ATL, depending on the location of the lesions (Barr, 1997a; Chiaravalloti & Glosser, 2004). Given the proximity of the right inferior visual processing region to the right mesial temporal lobe, and the known contribution of this area (particularly the fusiform gyrus) to the specialized processing of faces (Kanwisher, McDermott, & Chun, 1997), it might be expected that memory for unfamiliar faces would be selectively impaired following RATL. Such a deficit in memory for faces in RTLE and following RATL has in fact been a relatively consistent finding in the literature (Barr, 1997a; Glogau, Ellgring, Elger, & Helmstaedter, 2004; Glosser, Salvucci, & Chiaravalloti, 2003; Milner, 1968b; Newcombe & Russell, 1969; although see Hermann et al., 1995; Naugle et al., 1994).

Further evidence comes from a meta-analysis of nonverbal memory outcome following right ATL (Vaz, 2004). Conflicting results were reported, with some studies demonstrating a decrease in nonverbal memory performance postoperatively, and others showing an increase. Of all the tests evaluated, only a facial recognition memory test showed a decrease that was reliably different than zero postoperatively. As noted by the author, this isolated

result may indicate that other areas of the brain – particularly the inferior areas thought to be involved in facial recognition – are negatively affected by RATL.

Chiaravallotti and Glosser (2004) examined different aspects of visual memory before and after temporal lobectomy. They found a relative impairment in facial recognition memory in the RTLE group compared to the LTLE group, which was magnified following ATL. This was due to a significant improvement between pre and postoperative testing in the LTL group, while scores in the RTLE group remained constant. Evidence for a dissociation between aspects of visual memory was provided in that the LTLE and RTLE groups did not differ on a spatial sequential learning task, either before or after ATL. Similarly, Barr (1997a) found that RTLE patients obtained lower scores on a facial recognition test than LTLE patients, yet no differences were evident on a spatial learning test.

The findings of Chiaravallotti and Glosser (2004), Barr (1997), and others showing differences between RTLE and LTLE patients on tests of facial recognition memory, are discrepant with those of Study 1, in which performance on the WMS-III Faces subtest did not differ between the RTLE and the LTLE groups. In part, this may be due to procedural factors. For example, some studies have failed to show facial memory impairments in presurgical samples using a forced-choice recognition paradigm (Hermann et al., 1995; Naugle et al., 1994). The WMS-III Faces task is also a forced-choice recognition measure. In addition, the scoring procedure of the WMS-III subtest has been criticized due to its susceptibility to response bias. In response to this, Holdnack and Delis (2004) developed alternative scoring methods based on signal detection theory to improve discriminability.

Chiaravallotti, Tulsky and Glosser (2004) evaluated the validity of the WMS-III Faces subtest by comparing performance to the Graduate Hospital Facial Memory Test (FMT),

which had been shown to differentiate LTLE and RTLE groups in previous studies. They also compared the traditional WMS-III scoring to the methods suggested by Holdnack and Delis (2004). Correlations between the two tests were significant (ranges between .46 and .67 depending on the condition being compared), but as noted by the authors, not as high as might be expected given that they are purportedly measuring the same construct. In contrast to the results of Study 1, RTLE patients performed more poorly than LTLE patients on the WMS-III Faces Immediate subtest pre and post surgically. When the alternate scoring methods were applied, differences were evident on the immediate and delayed subtests. The authors concluded that, while the FMT was more sensitive to facial memory impairments, the WMS-III Faces subtests had clinical utility in this population, which was increased when the scoring methods developed by Holdnack and Delis (2004) were applied. Whether this holds true in additional TLE samples awaits replication.

While the distinction between facial recognition and spatial processing cannot explain all the incongruities among studies evaluating the role of the right temporal lobe in nonverbal memory, it does provide an illustration of how a theoretically guided framework can inform clinical research. Augmenting the methods traditionally used in clinical research (e.g., group differentiation) with methods and theoretically driven questions derived from the cognitive neuroscience literature will ultimately lead to a more sound understanding of our tests, and the meaning of the constructs underlying them. For example, the lack of commonality between Faces and Family Pictures shown in Study 2 and in the standardization sample is not surprising when thought of from a neurocognitive standpoint, despite the perfunctory similarity of being presented to the patient in the visual modality.

The Utility of Separate Immediate and Delayed Indexes

Factor analyses with the standardization sample have failed to show separable Immediate and Delayed Indexes (Millis et al., 1999; Price et al., 2002). Based on this work, and the results of the joint WAIS-III – WMS-III factor analyses (Tulsky & Price, 2003), composite Auditory and Visual indexes have been developed (Tulsky, Ivnik, Price, & Wilkins, 2003). However, the continued clinical use of separate Immediate and Delayed Auditory and Visual Memory indexes has been recommended (Millis et al., 1999; Price et al., 2002; Tulsky et al., 2003) with the rationale that these distinctions may have clinical utility in certain populations.

A distinction between immediate and delayed indexes might be predicted to be of clinical utility in TLE, based on findings that the integrity of the mesial temporal structures is more closely related to delayed than immediate memory (discussed in Bell et al., 2004; Bell, Fine, Dow, Seidenberg, & Hermann, 2005). However, in Study 1, no differences were found between scores on Immediate and Delayed indexes, in either group, in either modality. Furthermore, in Study 2, models specifying separate Immediate and Delayed factors were not supported.

Bell, Hermann, and Seidenberg (2004) investigated this issue further by evaluating whether the WMS-III Immediate and Delayed indexes provided unique information when considered separately, in a sample of patients with TLE. Their results showed that very few patients performed considerably worse on the Delayed index compared to the Immediate index, in either modality. The authors concluded that the WMS-III Immediate and Delayed indexes typically provided overlapping information in TLE patients, and that combining the relevant immediate and delayed scores to form composite indexes could be considered appropriate for this sample. Tulsky et al. (2003) provide tables that allow computation of these composite indexes from the subtest scaled scores.

As cautioned by Bell et al. (2004) there is a rare possibility that a significant discrepancy between immediate and delayed scores on the WMS-III could have some prognostic significance for patients undergoing ATL and this should be investigated. Interestingly, Doss et al. (2004) found significantly more pronounced distinctions between and among post-surgical RTL and LTL groups on the comparison of immediate indexes than in the delayed condition. This was also shown with the Faces subtest (Chiaravalloti & Glosser, 2004).

Taken together, these results suggest that the immediate indexes alone may be sufficient for the evaluation of epilepsy patients, which would cut the administration time of the scale in half. However, additional research is needed on whether there is additive value of the delayed indexes in particular circumstances.

It may be that consideration of separate immediate and delayed indexes might prove to be more useful in other clinical populations. Future research should investigate this possibility both at the score level and via factor analytic methods (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003). Of preliminary note, Heaton et al. (2003) analyzed the sensitivity of demographically corrected factor scores in various small diagnostic groups (The Psychological Corporation, 1997) and failed to show more severe delayed than immediate impairments, even in groups that would be expected to have shown such deficits (e.g., Alzheimer's disease).

Factor Structure of the WMS-III

Although the factor structure of the WMS-III in the standardization sample has been extensively studied (Millis et al., 1999; Price et al., 2002; Tulskey & Price, 2003) all of these authors describe a need for further investigation in clinical groups. Despite this, only one study (other than Study 2) has been published in which the factor structure of the WMS-III was investigated in a clinical sample.

Burton and colleagues (Burton, Ryan, Axelrod, Schellenberger, & Richards, 2003) conducted a confirmatory factor analysis of the WMS-III in a mixed clinical sample (N = 281). They evaluated several competing models based on the 11 primary subtest scores (Auditory Recognition was included) plus scores from one supplemental subtest (Word Lists, including immediate, delayed, and recognition scores). They concluded that a four factor model, composed of auditory memory, visual memory, working memory, and learning factors provided the best fit of the data in their clinical sample. This is congruent with the results obtained in the standardization sample (note, the learning factor was composed of the three Word Lists scores, which were not included in previous studies).

The use of factor analysis in normal populations or *mixed* clinical samples for the purposes of investigating construct validity has been criticized (Delis et al., 2003) as it may mask differential patterns of shared or unique variance resulting from particular cognitive deficits arising in certain subgroups. While Study 2 represented the first attempt at factor analysis in a homogenous clinical group, additional studies are clearly needed. Thus, it remains to be seen whether the less differentiated auditory and verbal memory factors obtained in Study 2 also apply to other clinical groups or epilepsy samples.

In the use of factor analysis as an investigation of construct validity, precautions against artificial covariance due to shared test stimuli and methodology need to be employed (Delis et al., 2003; Larrabee, 2003). In the factor analytic studies of the WMS-III conducted thus far (including Study 2), the error terms between immediate and delayed subtests were allowed to correlate as an attempt to control for method variance. However, this did not fully circumvent the problem, and models specifying separate immediate and delayed factors yielded boundary solution errors in Study 2 and the standardization sample analysis (Price et al., 2002), due to the very high correlations between immediate and delayed scores. A

possible solution has been suggested for evaluating the distinction between immediate and delayed memory while significantly reducing method variance. This would be to employ multiple memory tests, and include the immediate recall from one half of the tests, and the delayed recall from the other half, in one analysis, and then reverse the selection in a second analysis (Delis et al., 2003; Larrabee, 2003). While this method suffers from practical limitations in that it would require a large number of variables and participants, it presents a promising avenue for future investigation.

Neuropsychological Testing, Lateralization, and Material-Specific Memory Outcome

The accurate prediction of localization/lateralization of dysfunction is of utmost importance in the evaluation of TLE patients. While epilepsy surgery is extremely successful in the control of partial (focal) seizures resistant to drug treatment (e.g., Wiebe, Blume, Girvin, & Eliasziw, 2001), surgical decisions are made in part according to a risk-benefit analysis that assesses a patient's chance for seizure control versus acquiring or magnifying a functional deficit.

While advances in neuroimaging and electrophysiological monitoring allow for better diagnoses of seizure foci and subtle structural lesions, the functional integrity of temporal structures underlying memory and other cognitive functions can only be inferred, most prominently through applications of neuropsychological assessment.

Thus, neuropsychological testing plays an important role in most epilepsy surgery centers. From a clinical point of view, inconclusive or false lateralization has important and potentially serious negative impacts. Neuropsychological test results that are at odds with neuroimaging or EEG findings often lead to more extensive and invasive testing such as an IAP (Jones-Gotman et al., 1993) and/or monitoring of seizures via implantation of subdural

electrodes. This may result in a more restricted resection of the epileptic area, or even denial of surgery (Loring et al., 1997).

Despite some promising group analyses, studies investigating the lateralizing value of memory scores in individual patients with temporal lobe epilepsy have continued to yield disappointing results across a variety of memory tests with generally low sensitivity values being reported (see Kim, Yi, Son, & Kim, 2004). The IAP presents another option for evaluating the integrity of temporal structures and material specific memory. There is some debate as to whether the IAP can reliably predict postoperative memory outcome in the individual patient (Martin & Grote, 2002); however, prediction rates for the IAP alone (e.g., Kneebone, Chelune, & Luders, 1997) or in combination with neuropsychological testing (Akanuma et al., 2003) have been shown to be superior to standard neuropsychological testing alone. As with standard neuropsychological testing, the predictive power of IAP for postoperative verbal memory outcome in LATL patients has been consistently shown. However, the predictive power of IAP for postoperative nonverbal outcome in either left or right ATL patients has been more problematic.

The IAP is an invasive, costly, and potentially dangerous procedure, and the utility of its unique or additive value to the clinical decision making process is equivocal (Martin & Grote, 2002). While it is generally recommended as an adjunctive procedure in complex cases, it is no longer thought of as an essential component of every presurgical workup (Helmstaedter, 2004). Focus is shifting to the development of other methods to investigate lateralization issues. Recently, the availability of sodium amytal in Canada has been severely restricted, which makes this issue all the more relevant.

Recently, efforts have focused on developing functional magnetic resonance imaging (fMRI) paradigms for clinical use in patients with TLE both for the examination of language

lateralization and assessment of memory and integrity of mesial temporal structures (See Vingerhoets et al., 2004 for a review of fMRI findings in TLE). fMRI has the ability to provide greater functional and anatomic specificity than the IAP or neuropsychological testing. Efforts are being made to develop memory paradigms that can activate medial temporal regions and thus offer insight into the functional adequacy and postsurgical amnesic risk in patients with TLE. Promising results have been obtained showing the predictive value of fMRI to post-surgical memory (e.g., Janszky et al., 2005; Rabin et al., 2004; Richardson et al., 2004). Furthermore, the recent development of methodologies to allow EEG data to be recorded during MRI scanning has opened up new opportunities for combining the spatial resolution of imaging with the temporal resolution of electrophysiology in seizure localization (Detre, 2004). Although further methodological refinements and prospective clinical validation are required, fMRI might ultimately replace the IAP and neuropsychological testing for the specific purpose of lateralization/localization of dysfunction in TLE.

Neuropsychology will continue to play an important role in the pre and postsurgical evaluation of the TLE patient, however. For example, in assessing baseline performance, which is an important factor in the prediction of postsurgical decline (Stroup et al., 2003), characterizing the deleterious effects of seizures (Dodrill, 2004) and antiepileptic drugs (Ortinski & Meador, 2004) on cognitive status, documenting change after ATL (Martin et al., 2002), and providing quality and outcome control of surgical treatments (Helmstaedter, 2004).

REFERENCES

- AERA, APA, & NCME (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Akanuma, N., Alarcon, G., Lum, F., Kissani, N., Koutroumanidis, M., Adachi, N. et al. (2003). Lateralising value of neuropsychological protocols for presurgical assessment of temporal lobe epilepsy. *Epilepsia*, *44*, 408-418.
- Anastasi, A. & Urbina, S. (1997). *Psychological testing*. (7th ed.) Upper Saddle River, NJ: Prentice Hall.
- Arbuckle, J. L. & Wothke, W. (1999). *AMOS 4 user's reference guide*. Chicago: Smallwaters.
- Baker, G. A., Austin, N. A., & Downes, J. J. (2003). Validation of the Wechsler Memory Scale-III in a population of people with intractable temporal lobe epilepsy. *Epilepsy Research*, *53*, 201-206.
- Barr, W., Morrison, C., Zaroff, C., & Devinsky, O. (2004). Use of the Brief Visuospatial Memory Test-Revised (BVM-T-R) in neuropsychological evaluation of epilepsy surgery candidates. *Epilepsy & Behavior*, *5*, 175-179.
- Barr, W. B., Chelune, G. J., Hermann, B. P., Loring, D. W., Perrine, K., Strauss, E. et al. (1997). The use of figural reproduction tests as measures of nonverbal memory in epilepsy surgery candidates. *Journal of the International Neuropsychological Society*, *3*, 435-443.
- Barr, W. B. (1997a). Examining the right temporal lobe's role in nonverbal memory. *Brain & Cognition*, *35*, 26-41.
- Barr, W. B. (1997b). Receiver operating characteristic curve analysis of Wechsler Memory Scale--Revised scores in epilepsy surgery candidates. *Psychological Assessment*, *9*, 171-176.
- Bell, B. D., Hermann, B. P., & Seidenberg, M. (2004). Significant discrepancies between immediate and delayed WMS-III indices are rare in temporal lobe epilepsy patients. *Clinical Neuropsychologist*, *18*, 303-311.
- Bell, B. D. & Davies, K. G. (1998). Anterior temporal lobectomy, hippocampal sclerosis, and memory: Recent neuropsychological findings. *Neuropsychology Review*, *8*, 25-41.
- Bell, B. D., Fine, J., Dow, C., Seidenberg, M., & Hermann, B. P. (2005). Temporal Lobe Epilepsy and the Selective Reminding Test: The Conventional 30-Minute Delay Suffices. *Psychological Assessment*, *17*, 103-109.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

- Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bornstein, R. A. & Chelune, G. J. (1988). Factor structure of the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist*, 2, 107-115.
- Bornstein, R. A. & Chelune, G. J. (1989). Factor structure of the Wechsler Memory Scale-Revised in relation to age and educational level. *Archives of Clinical Neuropsychology*, 4, 15-24.
- Bornstein, R. A., Drake, M., & Pakalnis, A. (1988). Verbal and nonverbal memory and learning in patients with complex partial seizures of temporal lobe origin. *Journal of Epilepsy*, 1, 203-208.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107, 260-273.
- Browne, M. W. & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, 24, 445-455.
- Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A.Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Burton, D. B., Mittenberg, W., & Burton, C. A. (1993). Confirmatory factor analysis of the Wechsler Memory Scale-Revised standardization sample. *Archives of Clinical Neuropsychology*, 8, 467-475.
- Burton, D. B., Ryan, J. J., Axelrod, B. N., Schellenberger, T., & Richards, H. M. (2003). A confirmatory factor analysis of the WMS-III in a clinical sample with crossvalidation in the standardization sample. *Archives of Clinical Neuropsychology*, 18, 629-641.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Chelune, G. J. & Bornstein, R. A. (1988). WMS-R patterns among patients with unilateral brain lesions. *The Clinical Neuropsychologist*, 2, 121-132.
- Chelune, G. J., Naugle, R. I., Luders, H., & Awad, I. A. (1991). Prediction of cognitive change as a function of preoperative ability status among temporal lobectomy patients seen at 6-month follow-up. *Neurology*, 41, 399-404.
- Chelune, G. J., Naugle, R. I., Luders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, 7, 41-52.
- Chiaravalloti, N. D. & Glosser, G. (2004). Memory for faces dissociates from memory for location following anterior temporal lobectomy. *Brain and Cognition*, 54, 35-42.

- Chiaravalloti, N. D., Tulskey, D. S., & Glosser, G. (2004). Validation of the WMS-III facial memory subtest with the Graduate Hospital Facial Memory Test in a sample of right and temporal lobectomy left anterior patients. *Journal of Clinical and Experimental Neuropsychology*, *26*, 484-497.
- Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, *55*, 584-594.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.
- Delaney, R. C., Rosen, A. D., Mattson, R. H., & Novelly, R. A. (1980). Memory function in focal epilepsy: A comparison of non-surgical, unilateral temporal lobe and frontal lobe samples. *Cortex*, *16*, 103-117.
- Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: lessons from memory assessment. *Journal of the International Neuropsychological Society*, *9*, 936-946.
- Detre, J. A. (2004). fMRI: Applications in epilepsy. *Epilepsia*, *45*, 26-31.
- Dodrill, C. B. (2004). Neuropsychological effects of seizures. *Epilepsy & Behavior*, *5 Suppl 1*, S21-S24.
- Dodrill, C. B., Hermann, B. P., Rausch, R., Chelune, G. J., & Oxbury, S. (1993). Neuropsychological testing for assessing prognosis following surgery for epilepsy. In J. Engel, Jr. (Ed.), *Surgical treatment of the epilepsies* (2nd ed., pp. 263-271). New York: Raven Press, Ltd.
- Doss, R. C., Chelune, G. J., & Naugle, R. I. (2004). WMS-III performance in epilepsy patients following temporal lobectomy. *Journal of the International Neuropsychological Society*, *10*, 173-179.
- Drebing, C. E., Van Gorp, W. G., Stuck, A. E., Mitrushina, M., & Beck, J. (1994). Early detection of cognitive decline in higher cognitively functioning older adults: Sensitivity and specificity of a neuropsychological screening battery. *Neuropsychology*, *8*, 31-37.
- Dulay, M. F., Testa, S. M., Fargo, J., Eschleman, M., & Schefft, B. K. (2001). The WMS-III family pictures subtest and temporal lobe epilepsy. *Clinical Neuropsychologist*, *15*, 271.
- Dulay, M. F., Schefft, B. K., Testa, S. M., Fargo, J. D., Privitera, M., & Yeh, H. S. (2002). What does the Family Pictures subtest of the Wechsler Memory Scale-III measure? Insight gained from patients evaluated for epilepsy surgery. *Clinical Neuropsychologist*, *16*, 452-462.

- Eadie, K. & Shum, D. (1995). Assessment of visual memory: A comparison of Chinese characters and geometric figures as stimulus materials. *Journal of Clinical & Experimental Neuropsychology*, 17, 731-739.
- Elwood, R. W. (1991). Factor structure of the Wechsler Memory Scale-Revised (WMS-R) in a clinical sample: A methodological reappraisal. *The Clinical Neuropsychologist*, 5, 329-337.
- Engel, J., Jr. (1993). *Surgical treatment of the epilepsies*. (2nd ed.) New York: Raven Press.
- Engelhart, C., Eisenstein, N., & Meininger, J. (1994). Psychometric properties of the Neurobehavioral Cognitive Status Exam. *The Clinical Neuropsychologist*, 8, 404-415.
- Erickson, R. A. & Scott, M. L. (1977). Clinical memory testing: A review. *Psychological Bulletin*, 84, 1130-1149.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.
- Fletcher, J. M., Francis, D. J., Stuebing, K. K., Shaywitz, B. A., Shawitz, S. E., Shankweiler, D. P. et al. (1996). Conceptual and methodological issues in construct definition. In G.R.Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 17-42). Baltimore, MD: Paul H. Brookes.
- Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286-299.
- Franzen, M. D. (1989). *Reliability and validity in neuropsychological assessment*. New York, NY: Plenum Press.
- Franzen, M. D. (2000). *Reliability and validity in neuropsychological assessment*. (2nd ed.) New York: Kluwer Academic/Plenum Publishers.
- Glogau, S., Ellgring, H., Elger, C. E., & Helmstaedter, C. (2004). Face and facial expression memory in temporal lobe epilepsy patients: Preliminary results. *Epilepsy & Behavior*, 5, 106-112.
- Glosser, G., Salvucci, A. E., & Chiaravalloti, N. D. (2003). Naming and recognizing famous faces in temporal lobe epilepsy. *Neurology*, 61, 81-86.
- Glowinski, H. (1973). Cognitive deficits in temporal lobe epilepsy. An investigation of memory functioning. *Journal of Nervous and Mental Disease*, 157, 129-137.
- Gowers, W. R. (1881). *Epilepsy and other chronic convulsive diseases*. London: Churchill.
- Guilmette, T. J. & Rasile, D. (1995). Sensitivity, specificity, and diagnostic accuracy of three verbal memory measures in the assessment of mild brain injury. *Neuropsychology*, 9, 338-344.

- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley & Sons.
- Harman, H. H. (1967). *Modern factor analysis*. Chicago: University of Chicago Press.
- Heaton, R. K., Taylor, M. J., & Manly, J. (2003). Demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III. In D.S. Tulsky & D. H. Saklofske (Eds.), *Clinical interpretation of the WAIS-III and WMS-III* (pp. 181-210). San Diego, CA: Academic Press.
- Heilbronner, R. L. (1992). The search for a "pure" visual memory test: Pursuit of perfection? *The Clinical Neuropsychologist*, 6, 105-112.
- Helmstaedter, C. (2004). Neuropsychological aspects of epilepsy surgery. *Epilepsy & Behavior*, 5, S45-S55.
- Helmstaedter, C., Kurthen, M., & Elger, C. E. (1999). Sex differences in material-specific cognitive functions related to language dominance: an intracarotid amobarbital study in left temporal lobe epilepsy. *Laterality*, 4, 51-63.
- Helmstaedter, C., Kurthen, M., Lux, S., Reuber, M., & Elger, C. E. (2003). Chronic epilepsy and cognition: a longitudinal study in temporal lobe epilepsy. *Annals of Neurology*, 54, 425-432.
- Helmstaedter, C., Pohl, C., Hufnagel, A., & Elger, C. E. (1991). Visual learning deficits in nonresected patients with right temporal lobe epilepsy. *Cortex*, 27, 547-555.
- Hermann, B. P., Connell, B., Barr, W. B., & Wyler, A. R. (1995). The utility of the Warrington Recognition Memory Test for temporal lobe epilepsy: Pre- and postoperative results. *Journal of Epilepsy*, 8, 139-145.
- Hermann, B. P., Wyler, A. R., Richey, E. T., & Rea, J. M. (1987). Memory function and verbal learning ability in patients with complex partial seizures of temporal lobe origin. *Epilepsia*, 28, 547-554.
- Holdnack, J. A. & Delis, D. C. (2004). Parsing the recognition memory components of the WMS-III face memory subtest: Normative data and clinical findings in dementia groups. *Journal of Clinical and Experimental Neuropsychology*, 26, 459-483.
- Holley, F. K., Lineweaver, T. T., & Chelune, G. J. (2000). Performance differences on three components of the family pictures subset among right and left temporal lobectomy patients. In Orlando, FL: 20th Annual Meeting of the National Academy of Neuropsychology.
- Hoyle, R. H. & Panter, A. T. (1995). Writing about structural equation models. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: Sage.

- Hu, L. & Bentler, P. M. (1995). Evaluating model fit. In R.H.Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.
- Ivnik, R. J., Sharbrough, F. W., & Laws, E. R., Jr. (1987). Effects of anterior temporal lobectomy on cognitive function. *Journal of Clinical Psychology, 43*, 128-137.
- Janszky, T., Jokeit, H., Kontopoulou, K., Mertens, M., Ebner, A., Pohlmann-Eden, B. et al. (2005). Functional MRI predicts memory performance after right mesiotemporal epilepsy surgery. *Epilepsia, 46*, 244-250.
- Jones-Gotman, M. (1991). Localization of lesions by neuropsychological testing. *Epilepsia, 32 Suppl 5*, S41-S52.
- Jones-Gotman, M., Smith, M. L., & Zatorre, R. J. (1993). Neuropsychological testing for localizing and lateralizing the epileptogenic region. In J.Engel, Jr. (Ed.), *Surgical treatment of the epilepsies* (pp. 245-261). New York: Raven Press, Ltd.
- Jones-Gotman, M. (1986). Right hippocampal excision impairs learning and recall of a list of abstract designs. *Neuropsychologia, 24*, 659-670.
- Jöreskog, K. G. & Sörbom, D. (1981). *LISREL V: A analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources.
- Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7: User's reference guide*. Mooresville, IN: Scientific Software.
- Jurden, F. H., Franzen, M. D., Callahan, T., & Ledbetter, M. (1996). Factorial equivalence of the Wechsler Memory Scale: Revised across standardization and clinical samples. *Applied Neuropsychology, 3*, 65-74.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302-4311.
- Kaufman, A. S. & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment*. New York: Wiley.
- Kim, H., Yi, S., Son, E. I., & Kim, J. (2004). Lateralization of epileptic foci by neuropsychological testing in mesial temporal lobe epilepsy. *Neuropsychology, 18*, 141-151.
- Kimura, D. (1963). Right temporal lobe damage. *Archives of Neurology, 8*, 264-271.
- Kneebone, A. C., Chelune, G. J., & Luders, H. O. (1997). Individual patient prediction of seizure lateralization in temporal lobe epilepsy: a comparison between neuropsychological memory measures and the Intracarotid Amobarbital Procedure. *Journal of the International Neuropsychological Society, 3*, 159-168.

- Larrabee, G. J. (2003). Lessons on measuring construct validity: a commentary on Delis, Jacobson, Bondi, Hamilton, And Salmon. *Journal of the International Neuropsychological Society*, 9, 947-953.
- Larrabee, G. J., Kane, R. L., Schuck, J. R., & Francis, D. J. (1985). Construct validity of various memory testing procedures. *Journal of Clinical and Experimental Neuropsychology*, 7, 159-168.
- Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1996). Forensic neuropsychological test usage: An empirical survey. *Archives of Clinical Neuropsychology*, 11, 45-51.
- Lezak, M. D. (1995). *Neuropsychological assessment*. (3rd ed.) New York: Oxford University Press.
- Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, 4, 192-211.
- Loiseau, P., Strube, E., Broustet, D., Battellochi, S., Gomeni, C., & Morselli, P. L. (1983). Learning impairment in epileptic patients. *Epilepsia*, 24, 183-192.
- Loring, D. W. (1989). The Wechsler Memory Scale-Revised or the Wechsler Memory Scale-Revisited? *The Clinical Neuropsychologist*, 3, 59-69.
- Loring, D. W., Hermann, B. P., Perrine, K., Plenger, P. M., Lee, G. P., & Meador, K. J. (1997). Effect of Wada memory stimulus type in discriminating lateralized temporal lobe impairment. *Epilepsia*, 38, 219-224.
- Loring, D. W., Lee, G. P., Martin, R. C., & Meador, K. J. (1989). Verbal and Visual Memory Index discrepancies from the Wechsler Memory Scale-Revised: Cautions in interpretation. *Psychological Assessment*, 1, 198-202.
- Loring, D. W., Hermann, B. P., Lee, G. P., Drane, D. L., & Meador, K. J. (2000). The Memory Assessment Scales and lateralized temporal lobe epilepsy. *Journal of Clinical Psychology*, 56, 563-570.
- Loring, D. W., Lee, G. P., Martin, R. C., & Meador, K. J. (1988). Material-specific learning in patients with partial complex seizures of temporal lobe origin: Convergent validation of memory constructs. *Journal of Epilepsy*, 1, 53-59.
- MacCallum, R. C. & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51, 201-226.
- Malec, J. F., Ivnik, R. J., Smith, G. E., Tangalos, E. G., Peterson, R. C., Kokmen, E. et al. (1992). Visual spatial learning test: Normative data and further validation. *Psychological Assessment*, 4, 433-441.

- Martin, R., Sawrie, S., Gilliam, F., Mackey, M., Faught, E., Knowlton, R. et al. (2002). Determining reliable cognitive change after epilepsy surgery: Development of reliable change indices and standardized regression-based change norms for the WMS-III and WAIS-III. *Epilepsia*, 43, 1551-1558.
- Martin, R. C. & Grote, C. L. (2002). Does the Wada test predict memory decline following epilepsy surgery? *Epilepsy & Behavior*, 3, 4-15.
- Mayeux, R., Brandt, J., Rosen, J., & Benson, D. F. (1980). Interictal memory and language impairment in temporal lobe epilepsy. *Neurology*, 30, 120-125.
- McCaffrey, R. J., Ortega, A., & Haase, R. F. (1993). Effects of repeated neuropsychological assessments. *Archives of Clinical Neuropsychology*, 8, 519-524.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Meyer, V. & Yates, A. (1955). Intellectual changes following temporal lobectomy for psychomotor epilepsy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 18, 44-52.
- Millis, S. R., Malina, A. C., Bowers, D. A., & Ricker, J. H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale-III. *Journal of Clinical and Experimental Neuropsychology*, 21, 87-93.
- Milner, B. (1958). Psychological defects produced by temporal lobe excision. *Research Publications for Research in Nervous and Mental Diseases*, 36, 244-257.
- Milner, B. (1965). Visually-guided maze learning in man: effects of bilateral hippocampal, bilateral frontal, and unilateral cerebral lesions. *Neuropsychologia*, 3, 317-338.
- Milner, B. (1968a). Disorders of memory after brain lesions in man. Preface: Material-specific and generalized memory loss. *Neuropsychologia*, 6, 175-179.
- Milner, B. (1968b). Visual recognition and recall after right temporal lobe excision in man. *Neuropsychologia*, 6, 11-26.
- Milner, B. (1970). Memory and the medial temporal regions of the brain. In K. Pribram & Broadbent (Eds.), *Biology of memory* (New York: Academic Press).
- Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, 27, 272-277.
- Milner, B. (1975). Psychological aspects of focal epilepsy and its neurosurgical management. *Advances in Neurology*, 8, 299-321.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, 6, 414-417.

- Monsch, A. U., Bondi, M. W., Butters, N., Salmon, D. P., Katzman, R., & Thal, L. J. (1992). Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Archives of Neurology*, *49*, 1253-1258.
- Moore, P. M. & Baker, G. A. (1996). Validation of the Wechsler Memory Scale-Revised in a sample of people with intractable temporal lobe epilepsy. *Epilepsia*, *37*, 1215-1220.
- Moore, P. M. & Baker, G. A. (1997). Psychometric properties and factor structure of the Wechsler Memory Scale-Revised in a sample of persons with intractable epilepsy. *Journal of Clinical and Experimental Neuropsychology*, *19*, 897-905.
- Naugle, R. I., Chelune, G. J., Cheek, R., Luders, H., & Awad, I. A. (1993). Detection of changes in material-specific memory following temporal lobectomy using the Wechsler Memory Scale--Revised. *Archives of Clinical Neuropsychology*, *8*, 381-395.
- Naugle, R. I., Chelune, G. J., Schuster, J., Luders, H., & Comair, Y. (1994). Recognition memory for words and faces before and after temporal lobectomy. *Assessment*, *1*, 373-381.
- Newcombe, F. & Russell, R. (1969). Dissociated visual perceptual and spatial deficits in focal lesions of the right hemisphere. *Journal of Neurology, Neurosurgery & Psychiatry*, *32*, 73-81.
- Nunnally, J. C. (1978). *Psychometric theory*. (2nd ed.) New York: McGraw-Hill.
- Olejnik, S., Li, J., Supattathum, S., & Huberty, C. J. (1997). Multiple testing and statistical power with modified Bonferroni procedures. *Journal of Educational and Behavioral Statistics*, *22*, 389-406.
- Ortinski, P. & Meador, K. J. (2004). Cognitive side effects of antiepileptic drugs. *Epilepsy & Behavior*, *5 Suppl 1*, S60-S65.
- Price, L. R., Tulsky, D., Millis, S., & Weiss, L. (2002). Redefining the factor structure of the Wechsler Memory Scale- III: Confirmatory factor analysis with cross-validation. *Journal of Clinical and Experimental Neuropsychology*, *24*, 574-585.
- Prigatano, G. F. (1977). Wechsler Memory Scale is a poor screening test for brain dysfunction. *Journal of Clinical Psychology*, *33*, 772-777.
- Prigatano, G. F. (1978). Wechsler Memory Scale: A selective review of the literature (monograph). *Journal of Clinical Psychology*, *34*, 816-832.
- Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*, 33-65.
- Rabin, M. L., Narayan, V. M., Kimberg, D. Y., Casasanto, D. J., Glosser, G., Tracy, J. I. et al. (2004). Functional MRI predicts post-surgical memory following temporal lobectomy. *Brain*, *127*, 2286-2298.

- Raftery, A. E. (1993). Bayesian model selection in structural equation models. In K.A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 163-180). Newbury Park, CA: Sage.
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*, 287-297.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566.
- Reynolds, J. R. (1861). *Epilepsy: its symptoms, treatment and relation to other chronic convulsive diseases*. London: Churchill.
- Richardson, M. P., Strange, B. A., Thompson, P. J., Baxendale, S. A., Duncan, J. S., & Dolan, R. J. (2004). Pre-operative verbal memory fMRI predicts post-operative memory decline after left temporal lobe resection. *Brain, 127*, 2419-2426.
- Roid, G. H., Prifitera, A., & Ledbetter, M. (1988). Confirmatory analysis of the factor structure of the Wechsler Memory Scale - Revised. *The Clinical Neuropsychologist, 2*, 116-120.
- Roth, D. L., Conboy, T. J., Reeder, K. P., & Boll, T. J. (1990). Confirmatory factor analysis of the Wechsler Memory Scale-Revised in a sample of head-injured patients. *Journal of Clinical and Experimental Neuropsychology, 12*, 834-842.
- Schwartz, M. L. & Dennerll, R. D. (1969). Immediate visual memory as a function of epileptic seizure type. *Cortex, 5*, 69-74.
- Silverberg, N. & Buchanan, L. (2005). Verbal mediation and memory for novel figural designs: A dual interference study. *Brain & Cognition, 57*, 198-209.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika, 73*, 751-754.
- Smith, G. E., Ivnik, R. J., Malec, J. F., Kokmen, E., Tangalos, E. G., & Kurland, L. T. (1992). Mayo's older americans normative studies (MOANS): Factor structure of a core battery. *Psychological Assessment, 4*, 382-390.
- Smith, G. E., Malec, J. F., & Ivnik, R. J. (1992). Validity of the construct of nonverbal memory: A factor-analytic study in a normal elderly sample. *Journal of Clinical and Experimental Neuropsychology, 14*, 211-221.
- Snyder, P. J. (1998). Epilepsy. In P.J. Snyder & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment* (Washington, DC: American Psychological Association).
- Spren, O. & Strauss, E. (1991). *A compendium of neuropsychological tests*. New York: Oxford University Press.

- Spreen, O. & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary*. (2nd ed.) New York: Oxford University Press.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Strauss, E., Satz, P., & Wada, J. (1990). An examination of the crowding hypothesis in epileptic patients who have undergone the carotid amytal test. *Neuropsychologia*, 28, 1221-1227.
- Strauss, E., Spreen, O., & Hunter, M. (2000). Implications of test revisions for research. *Psychological Assessment*, 12, 237-244.
- Strauss, E., Loring, D., Chelune, G., Hunter, M., Hermann, B., Perrine, K. et al. (1995). Predicting cognitive impairment in epilepsy: Findings from the Bozeman Epilepsy Consortium. *Journal of Clinical and Experimental Neuropsychology*, 17, 909-917.
- Stroup, E., Langfitt, J., Berg, M., McDermott, M., Pilcher, W., & Como, P. (2003). Predicting verbal memory decline following anterior temporal lobectomy (ATL). *Neurology*, 60, 1266-1273.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- Taylor, L. B. (1969). Localization of cerebral lesions by psychological testing. *Clinical Neurosurgery*, 16, 269-287.
- The Psychological Corporation (1997). *WAIS-III-WMS-III technical manual*. San Antonio, TX: Author.
- Tulsky, D. S. & Ledbetter, M. F. (2000). Updating to the WAIS-III and WMS-III: Considerations for research and clinical practice. *Psychological Assessment*, 12, 253-262.
- Tulsky, D. S., Manly, J., Heaton, R., Lineweaver, T., Chelune, G., Hawkins, K. et al. (2000). Issues in the assessment of cognitive functioning: New research on the WAIS-III and WMS-III. In National Academy of Neuropsychology (Ed.), Orlando, FL.
- Tulsky, D. S., Rolfhus, E. L., & Zhu, J. J. (2000). Two-tailed versus one-tailed base rates of discrepancy scores in the WAIS-III. *Clinical Neuropsychologist*, 14, 451-460.
- Tulsky, D. S., Ivnik, R. J., Price, L. R., & Wilkins, C. (2003). Assessment of cognitive functioning with the WAIS-III and WMS-III: Development of a six-factor model. In D.S. Tulsky & D. H. Saklofske (Eds.), *Clinical interpretation of the WAIS-III and WMS-III* (pp. 147-179). San Diego, CA: Academic Press.
- Tulsky, D. S. & Price, L. R. (2003). The joint WAIS-III and WMS-III factor structure: Development and cross-validation of a six-factor model of cognitive functioning. *Psychological Assessment*, 15, 149-162.

- Vaz, S. A. M. (2004). Nonverbal memory functioning following right anterior temporal lobectomy: a meta-analytic review. *Seizure-European Journal of Epilepsy*, 13, 446-452.
- Vingerhoets, G., Deblaere, K., Backes, W. H., Achten, E., Boon, P., Boon, P. J. et al. (2004). Lessons for neuropsychology from functional MRI in patients with epilepsy. *Epilepsy & Behavior*, 5, S81-S89.
- Wechsler, D. (1945). A standardized memory scale for clinical use. *The Journal of Psychology*, 19, 87-95.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation.
- Wechsler, D. (1987). *Manual for the Wechsler Memory Scale-Revised*. New York: The Psychological Corporation.
- Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale-Third Edition: A administration and scoring manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997b). *Wechsler Memory Scale-Third Edition: A administration and scoring manual*. San Antonio, TX: The Psychological Corporation.
- Weingartner, H. (1968). Verbal learning in patients with temporal lobe lesions. *Journal of Verbal Learning and Verbal Behavior*, 7, 520-526.
- Wiebe, S., Blume, W. T., Girvin, J. P., & Eliasziw, M. (2001). A randomized, controlled trial of surgery for temporal-lobe epilepsy. *New England Journal of Medicine*, 345, 311-318.
- Wilde, N., Strauss, E., Chelune, G. J., Loring, D. W., Martin, R. C., Hermann, B. P. et al. (2001). WMS-III performance in patients with temporal lobe epilepsy: Group differences and individual classification. *Journal of the International Neuropsychological Society*, 7, 881-891.
- Wilde, N. J., Strauss, E., Chelune, G. J., Hermann, B. P., Hunter, M., Loring, D. W. et al. (2003). Confirmatory factor analysis of the WMS-III in patients with temporal lobe epilepsy. *Psychological Assessment*, 15, 56-63.
- Woodward, J. L. (1993). Confirmatory factor analysis of the Wechsler Memory Scale-Revised in a mixed clinical population. *Journal of Clinical and Experimental Neuropsychology*, 15, 968-973.
- Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K.A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256-293). Newbury Park, CA: Sage.