



This is to certify that the

dissertation entitled

TEST-RETEST PRACTICE EFFECTS, RELIABILITY,
AND STABILITY OF THE WAIS-R IN RECOVERING
TRAUMATICALLY-BRAIN-INJURED SURVIVORS

presented by

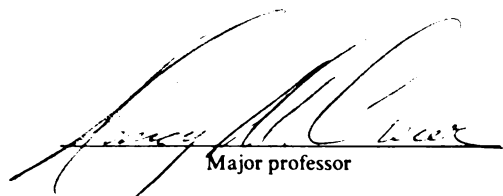
David Brian Rawlings

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling,
Educational Psychology and Special Education

Date

5/31/90



Major professor

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
 TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
AUG 31 1995 35755/VER		

MSU is An Affirmative Action/Equal Opportunity Institution

c:\pic\dtdue.pm3-p.1

TEST-RETEST PRACTICE EFFECTS, RELIABILITY, AND STABILITY
OF THE WAIS-R IN RECOVERING TRAUMATICALLY-
BRAIN-INJURED SURVIVORS

By

David Brian Rawlings

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling,
Educational Psychology and Special Education

1990

655-2869

ABSTRACT

TEST-RETEST PRACTICE EFFECTS, RELIABILITY, AND STABILITY OF THE WAIS-R IN RECOVERING TRAUMATICALLY- BRAIN-INJURED SURVIVORS

By

David Brian Rawlings

Test-retest procedures to investigate practice effects, reliability, and stability of the WAIS-R over time are virtually non-existent. While a limited number of studies have been completed on its predecessor, the WAIS, psychologists have and will continue to practice under a handicap unless a considerable amount of actuarial base-rate information is collected and published on normal as well as patient (i.e., neurologically impaired, developmentally disabled) populations. Without such information, an understanding of the influences of repeated administrations on neuropsychological tests will be difficult, implying that the interpretation of test scores may be erroneous. Having greater specificity and accuracy in determining change over time, particularly in recovering traumatically-brain-injured individuals, is paramount since charting intellectual and cognitive recovery is central to rehabilitation and/or vocational planning, as well as medico-legal determinations.

A group of adult traumatically-brain-injured (TBI) subjects who were tested at approximately 2, 4, 8, and 12 months post-injury with the WAIS-R were compared with a matched control group of similar adult TBI survivors who

were tested only twice with the WAIS-R at approximately 2 and 12 months post-injury. No significant differences were noted between the two groups on numerous pre-injury, injury, or post-injury variables including age, education, severity of injury, length of treatment, CT scan results, and IQ. Both groups demonstrated significant gains in both IQ and subtest scaled scores at one year post-injury, however no differences were noted between the two groups at the one year interval with the exception of the Comprehension and Picture Completion subtests. Significantly greater change scores were demonstrated in the experimental group than in the control group suggesting possible test-retest practice effects in Performance IQ, and the subtests Comprehension, Similarities, Picture Completion, Picture Arrangement, and Object Assembly. Test-retest reliabilities ranged from .48 (Picture Arrangement) to .84 (Vocabulary). Trend analyses detected both linear and quadratic recovery curves with recovery slowing as measured by Verbal subtests by the third evaluation (8 months post-injury). Concerns regarding measures of internal validity, implications of results for clinical practice, and suggestions for future research are discussed.

Copyright by
DAVID BRIAN RAWLINGS
1990

To my wife, Sharon.

ACKNOWLEDGEMENTS

I am deeply indebted to Dr. Nancy Crewe for her willingness to accept, direct, and support this thesis from its inception. I am also extremely grateful for her political acumen when it was needed, and for sharing with me her expertise in rehabilitation and clinical neuropsychology over the years.

I am also indebted to Dr. Michael Harwell for contributing both his time and statistical expertise to this research, and whose interest in and support of my dissertation has been unwavering. I would also like to thank the other members of my committee, Drs. William Hinds and Richard Johnson, for their time and consultation. I also want to thank Dr. Rochelle Habeck for assisting me in the preliminary phase of my proposal, and for bringing this research project full circle.

I would also like to express my appreciation to Drs. Robert Kreitsch and Bill Leer, as well as Mr. Bill Sonday, for their steadfast support and understanding of my work over the years, and for providing the many resources necessary for the completion of this project. I am also thankful for

the opportunity to work with Drs. Maureen Levine, Charles Reeder, and Chet Hoyt whose assistance and encouragement gave this thesis additional meaning. I must also thank Alice Kalush for her assistance in computer programming, and Sandy Savina for working miracles on her word processor.

Last, but certainly not least, I owe my deepest thanks and appreciation to my wife, Sharon, whose love, patience, understanding, and numerous personal and professional sacrifices kept me afloat throughout my graduate school career. This, despite raising three children and being deeply involved in her own career as a practicing attorney. I also owe many thanks to my children Brittany, Jacquelyn, and Andrew, for being wonderful kids during this process and who may have appreciated, but did not always understand, why Dad was still in school.

TABLE OF CONTENTS

I.	RESEARCH PROBLEM	1
	Introduction	1
	Identification of the Problem	2
	Purpose of the Study	10
	Importance	11
	Research Questions	12
	Research Hypotheses	15
II.	LITERATURE REVIEW	18
	Introduction	18
	Neuropsychological Assessment	18
	Reliability	25
	Wechsler Adult Intelligence Scales	30
	Test-Retest Reliability of the Wechsler Scales	36
	Comparison of the WAIS and the WAIS-R	53
	Practice Effects	58
	Practice Effects on the Wechsler Scales	62
	Studies of Gain Scores Not Involving Normals	73
	Factors Which Influence Practice Effects	81
	Intellectual Recovery in TBI Survivors	85
III.	METHODOLOGY	97
	Introduction	97
	Subject Population	100
	Sample Selection	101
	Data Collection	103
	Characteristics of the Sample	107
	Instrumentation	108
	Design	111
	Threats to Validity	113
	Hypotheses	120
	Data Analyses	121
	Limitations of the Analyses	125
IV.	ANALYSIS OF RESULTS	126
	Introduction	126
	Demographic Data	127
	Preliminary Data	136

Tests of Primary Hypothesis 1	142
Hypothesis 2	144
Hypothesis 3	145
Hypothesis 4	148
Hypothesis 5	150
Tests of Secondary Hypothesis 1	160
Hypothesis 2	163
Hypothesis 3	168
 V. DISCUSSION AND CONCLUSIONS	 170
Summary	170
Discussion	173
Implications for Future Research	203
Conclusions	205
 APPENDICES	
A - Approximate Test Times Post-injury	210
B - Data Summary Form	213
C - Michigan State University Committee on Research Involving Human Subjects (UCRIHS) approval letter	215
D - Informed Consent Letter	216
E - Consent Form	217
F - Client Information Release Authorization	219
 BIBLIOGRAPHY	 220

LIST OF TABLES

Table

2.1	Average Reliability Coefficients and Standard Errors of Measurement of the WAIS and WAIS-R	32
2.2	WAIS IQ Test-retest Reliabilities by Subject Population	38
2.3	WAIS Subtest Test-retest Reliabilities by Subject Population	43
2.4	WAIS-R IQ Test-retest Reliabilities by Subject Population	48
2.5	WAIS-R Subtest Test-retest Reliabilities by Subject Population	50
2.6	WAIS/WAIS-R Correlations	54
2.7	WAIS Minus WAIS-R Difference Scores	56
2.8	WAIS IQ Test-retest Gains by Subject Population	65
2.9	WAIS Subtest Test-retest Gains by Subject Population	69
2.10	Test-retest WAIS-R IQ Gains by Subject Population	71
2.11	WAIS-R Subtest Test-retest Gains by Subject Population	72
2.12	Test-retest Gains of Impaired and Control Populations on WAIS IQ Values	77
4.1	Comparison of Experimental and Control Group Subjects on Pre-injury Demographic Variables	128

4.2	Comparison of Experimental and Control Group Subjects on Injury and Post-injury Variables	132
4.3	Comparison of Demographic Variables Between Experimental and Control Groups	137
4.4	Comparison of WAIS-R IQ and Subtest Scaled Scores Between Experimental and Control Groups at Pretest	141
4.5	Experimental and Control Group Changes in WAIS-R IQ Mean Values from Pretest to Posttest	143
4.6	Experimental Group Changes in WAIS-R Subtest Scaled Score Mean Values from Pretest to Posttest	146
4.7	Control Group Changes in WAIS-R Subtest Scaled Score Mean Values from Pretest to Posttest	147
4.8	Comparison of WAIS-R IQ and Subtest Scaled Scores Between Experimental and Control Groups at Posttest	149
4.9	Comparisons of WAIS-R IQ and Subtest Change Score Means from Pretest to Posttest	151
4.10	Comparisons Between the Pretest/Posttest Correlations of the Experimental and Control Groups	162
4.11	Change in WAIS-R IQ and Subtest Scaled Scores Over Four Test Administrations	164
4.12	Trend Analysis Summary of WAIS-R IQs by Time in the Experimental Group	166

LIST OF FIGURES

Figure

3.1	Research Design	112
3.2	Non-equivalent Control Design	112
4.1	Interaction Between Time and Treatment Effects for Performance IQ Means	154
4.2	Interaction Between Time and Treatment Effects for Similarities Scaled Score Means . .	155
4.3	Interaction Between Time and Treatment Effects for Picture Arrangement Scaled Score Means	156
4.4	Interaction Between Time and Treatment Effects for Object Assembly Scaled Score Means	157
4.5	Interaction Between Time and Treatment Effects for Comprehension Scaled Score Means	158
4.6	Interaction Between Time and Treatment Effects for Picture Completion Scaled Score Means	159
4.7	Changes in Verbal, Performance, and Full Scale IQ Means Over Time	167
4.8	Changes in Performance IQ Variance Over Time	186

CHAPTER I

Research Problem

Introduction

The growing emphasis in health and rehabilitation fields on continuity of care, and the importance of prolonged monitoring of patient behavior suggest an important role for information derived from serial testing in a number of different contexts (Seidenberg, O'Leary, Giordani, Berent & Boll, 1981). Nowhere is this more true than in the field of clinical neuropsychology. In particular, there is an increasing need to provide information regarding the developmental progression or course of neurological events (Golden, 1976; Seidenberg et al., 1981), to examine the effects of neurosurgery and psychopharmacological agents on human behavior (Campbell, 1983), to document the nature and extent of recovery from cerebral trauma (Campbell, 1983; Wolfe, 1987), and to assess the efficiency of treatment intervention in traumatically-brain-injured (TBI) survivors through repeat test administrations (Campbell, 1983; Golden, 1976; Lezak, 1983; Wolfe, 1987).

Serial testing is not unusual. In fact, there appears to be an increase in the number of tests readministered to individuals over time in clinical practice (Matarazzo &

Herman, 1984). In part, this seems to stem from the fact that many tests lack equivalent, alternative forms; that is, two tests which are comparable including the same number of items, level of difficulty, content, and test administration procedures. In addition, test-retest procedures (administering the same test on a second occasion) represent the most substantively direct, as well as accurate means of assessing change in individuals' level of functioning (Seidenberg et al., 1981; Tabaddor, Mattis, & Zazula, 1984). Unfortunately, without understanding the influence of repeated administrations on neuropsychological test results and the magnitude of error variance contiguous with an individual's obtained score, a psychologist evaluating the effects of treatment or monitoring a patient's recovery from cerebral trauma will be unable to determine the true nature of changes observed in the patient's successive test performances (Campbell, 1983; Matarazzo & Herman, 1984). Interpretation of observed changes is further confounded by a dearth of standardized data collected upon various populations that would otherwise provide normative information about test-retest scores in clinical practice (Matarazzo & Herman, 1984; Seidenberg et al., 1981).

Identification of the Problem

Since intellectual functions are characteristically and often seriously affected by traumatic-brain-injury (Campbell, 1983), there is a preference among psychologists

to utilize the Wechsler Intelligence Scales as measures of cognitive outcome, particularly in the determination of long-term recovery following head injury (Levin, Benton & Grossman, 1982). In a recent survey, the Wechsler Intelligence Scales were reportedly employed in 97% of all neuropsychological evaluations performed by members of the National Academy of Neuropsychologists (Seretny, Dean, Gray & Hartlage, 1986). The popularity of the Wechsler Scales is further underscored by the vast amount of research that has been conducted with the scales, and their universal acceptance (Lezak, 1983). The ecologic validity of the Wechsler scales has been determined (Levin et al., 1982), and extensive reviews concerning other validity studies involving the scales are noted elsewhere (e.g., Anastasi, 1976; Sattler, 1988; Zimmerman & Woo-Sam, 1973). For these reasons, a further investigation of the validity of the Wechsler Scales is beyond the scope of this study and will not be addressed.

The Wechsler Scales for adults consist of the Wechsler-Bellevue Intelligence Scale (W-B) (Wechsler, 1939), the Wechsler Adult Intelligence Scale (WAIS) (Wechsler, 1955), and the Wechsler Adult Intelligence Scale - Revised (WAIS-R) (Wechsler, 1981). The Wechsler Scales contain eleven subtests grouped into Verbal and Performance sections. The six Verbal Scale subtests are Information, Digit Span, Vocabulary, Arithmetic, Comprehension, and Similarities. The five Performance Scale subtests are

Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Digit Symbol. The Verbal Scale subtests yield a Verbal IQ and the Performance Scale subtests yield a Performance IQ. Together, the two sections (eleven subtests) yield a Full Scale IQ value.

Developed to measure general intellectual aptitude as well as the potential for purposeful and useful behavior (Wechsler, 1981), the Wechsler Scales, when employed in neuropsychological practice with the traumatically-brain-injured, can be used as a guide to the level of ability and the range of general knowledge acquired by a patient prior to a head injury (Newcombe, 1982). In addition, the scales may serve as useful criteria for the evaluation and re-evaluation of selective post-traumatic deficits subsequent to the head injury (Lezak, 1983; Newcombe, 1982). The scales also permit a direct comparison of different abilities and yield a profile of subtest scores; the interpretation of which is augmented by the availability of comprehensive standardized data for a wide distribution of ages (Levin et al., 1982).

When used serially in test-retest procedures with either normal populations or brain-injured patients, information concerning test-retest reliability and stability, standard error of measurement and test-retest practice effects are imperative (Campbell, 1983). Test-retest reliability is defined by Anastasi (1976) as the correlation between the scores obtained by the same persons on two administrations

of the same test. Cronbach (1960) referred to this correlation as a coefficient of stability since it reflected the stability of performance over time. Matarazzo, Carmody, and Jacobs (1980) noted that references to test-retest stability are most pertinent when test-retest intervals are greater than two months and indicated that test-retest reliability and test-retest stability are terms often used interchangeably. The standard error of measurement of a test describes the amount of flexibility that should accompany the use of observed scores to estimate an individual's theoretical "true" score on that test. The true score would be the score that would be obtained on average if the same person could be tested a large number of times on the same test and effects such as practice could be ruled out (Wechsler, 1981). Test-retest practice effects result in improvements in performance which reflect the influences of learning and positive carry-over as a result of having been exposed to the tasks on a previous occasion (Seidenberg et al., 1981).

Without this information noted above, it is impossible to discern whether observed changes upon re-evaluation are due to treatment, spontaneous recovery, test-retest practice effects, or test unreliability (Campbell, 1983). Unfortunately, relatively little research has been published on the test-retest reliability of the Wechsler Adult Intelligence Scale (Wechsler, 1955) and less so on its revision, the Wechsler Adult Intelligence Scale - Revised (Wechsler,

1981), in either normal or clinical populations (Brown & May, 1979; Matarazzo, Carmody & Jacobs, 1980; Matarazzo & Herman, 1984).

Those studies which have been published report that the Wechsler Scales test-retest reliabilities are "statistically and clinically very robust" (Matarazzo et al., 1980, p. 82), suggesting high test-retest stability in both clinical and normal samples with varying test-retest intervals. Matarazzo and Herman (1984), however, were quick to point out that the psychometric definition of test-retest reliability reported in the research is not necessarily synonymous with a clinical interpretation of test-retest reliability. This conclusion stems from the assertion that psychometric stability and/or reliability is reflected by a test-retest correlation in studies conducted to determine the psychometric properties of the test itself, whereas clinical stability and/or reliability involves an investigation of test-retest changes that take place in an individual or group due to specific interventions or other life span changes (Matarazzo et al., 1980). Thus, high psychometric stability or reliability would be reflected by a sizeable test-retest correlation, and high clinical stability or reliability would be demonstrated by the absence of a meaningful score change (Sattler, 1988). In this regard, the Wechsler Scales are less reliable in a clinical sense than might be inferred from the test-retest coefficients alone (Ryan, Georgemiller, Geisser & Randall, 1985) since all

scores may increase on retest due to within subject variability, and only the rank order of the scores are similar (Catron & Thompson, 1979). Thus, in the presence of high reliability coefficients, a psychologist might erroneously conclude that the Wechsler Scales will produce scores on retest nearly identical in value to scores obtained in the initial test, if in fact, only rankings in performance are similar, but substantial gain in test scores occurred due to retest effects (Matarazzo et al., 1980).

Matarazzo et al. (1980) and Matarazzo and Herman (1984) reported gains on both subtest scaled scores and IQ scores when normals were retested with the WAIS and WAIS-R without intervention of any kind. The authors attributed these gains to motivational differences in individuals' test-taking abilities on the two occasions, the less than perfect reliability of the tests themselves, practice or test-retest effects, random error of measurement, or some other as yet undiscovered factor. This led Matarazzo et al. (1980) and Matarazzo and Herman (1984) to suggest rules that "a change of 3 to 5 points in a subtest score and a change of 15 points or more in an IQ score" (p. 103) may be interpreted as potentially clinically meaningful on retest requiring further analysis and clarification. Without behavioral or clinical corroborative data to substantiate a patient's change in Wechsler Scale score(s), such changes in isolation would not be robust proof that a true clinical change had occurred (Matarazzo et al., 1980), and should be considered

a practice or test-retest effect (Matarazzo, 1972). Therefore, the possibility always exists that in a serial investigation of the recovery from TBI, improvements in an individual's score may be a function of a second administration of the test rather than genuine recovery of the patient (Mandleberg & Brooks, 1975).

A second, equally important consideration in evaluating gains in test scores upon retesting is the phenomenon of regression-toward-the-mean. This is particularly true when individuals are selected because they deviate from the mean on some variable (Glass & Hopkins, 1984). Assuming that the time interval between test and retest is such that there is absolutely no practice effect, there is a definite and pronounced tendency for subjects to regress toward the mean to the extent that subjects scores tend to be, on average, closer to the posttest mean than would be expected from their pretest score. The magnitude of the regression-toward-the-mean phenomenon is a function of the size of the pretest posttest correlation coefficient (test-retest reliability) (Glass & Hopkins, 1984). Perfect test-retest reliability (i.e., correlation = ± 1.00) rules out regression-to-the-mean; that is, test-retest correlations less than ± 1.00 ensure that regression-to-the-mean will be present. The latter is almost always the case. Thus, in the absence of treatment or practice effects, posttest scores are likely to be higher, on average, than pretest scores. This is particularly true in measuring recovery

subsequent to head injury. When a traumatic-brain-injury changes IQ scores, it can only lower them due to the trauma imposed on the physiological, and hence, cognitive functions of the brain. Therefore, the mean IQ scores for the group would be reduced initially and any subsequent recovery would result in an elevation of IQ scores relative to the initial scores.

Research by Shatz (1981) and others (Dodrill & Troupin, 1975; Ivnik, 1978; Seidenberg et al., 1981) on patient populations, however, suggested that increases upon post-treatment retesting in individuals with known cerebral dysfunction may be related to intervention strategies and/or improved cerebral functioning, rather than to the test-retest practice effects typically seen in non-neurologically impaired subjects. Mandleberg and Brooks (1975) also reported that repeated exposure to the WAIS did not significantly enhance the IQ scores of recovering patients with head injury at 30 months post-injury when compared to controls tested only once; a finding which is congruent with others (e.g., Levin et al., 1982). Therefore, test-retest practice effects may differ greatly over time when neurologically normal subjects are compared to patient populations with cerebral impairment on dependent measures such as IQ scores.

If this is true, then the application of Matarazzo et al.'s (1980) "rule of thumb" (p. 103) to retest changes in patients with cerebral dysfunction is likely to be

misleading (Shatz, 1981). Therefore, it would be useful to have a data base which includes a range of normative changes observed over time on a particular test, such as the WAIS-R, with a particular population, such as recovering TBI patients (Seidenberg et al., 1981). Only then would the empirical foundation for the evaluation of the WAIS-R and the existence of the possible test-retest practice effects in neuropsychological assessment be established (Matarazzo & Herman, 1984; Shatz, 1981). Changes in test performances over time in head-injured patients could then be more clearly interpreted (Campbell, 1983).

Purpose of the Study

In general, this study provides a data base from which changes observed on the WAIS-R over time in a population of recovering TBI patients will be examined. Within this context, the study: a) provides descriptive information regarding the stability of test scores of individuals recovering from TBI within a one year period of time; b) attempts to determine the existence, extent, and magnitude of test-retest practice effects resulting from repeat administrations of the WAIS-R over time; and c) provides information concerning the test-retest reliability of the WAIS-R through serial testing.

Importance

This study is important for several reasons. First, information obtained from this study should contribute to existing knowledge about the extent of intellectual recovery that might be expected in the TBI survivor over a protracted period of time. Second, is the theoretical and clinical need to distinguish between the processes of genuine recovery in surviving brain-injured individuals and test-retest practice effects due to serial testing (Brooks, Deelman, Van Zomeran, Van Dongen, Van Harskamp & Aughton, 1984). The determination of the existence and extent of test-retest practice effects in repeat testing will, in part, help delineate those factors which influence the magnitude of change and adequacy of an individual's functioning at a specified time post-injury (Dikmen, Reitan, & Temkin, 1983). Third, without being cognizant of the magnitude of test-retest practice effects evident in repeat testing in recovering TBI patients, it would be difficult to discern whether changes are due to the effects of practice, spontaneous recovery, or intervention (Campbell, 1983). Fourth, this study will help determine the test-retest reliability of the WAIS-R and describe patterns of subtest scores for clinical practice. Such information can be crucial when decisions must be made with respect to suitability for therapy, and evaluation of change subsequent to specific treatment procedures (Warner, 1983). Fifth, since information concerning test-retest reliability and practice

effects are important to the clinician employing the WAIS-R, these same data can be utilized to calculate standard errors of measurement. Standard errors of measurement could then be used to assess individual changes over time with confidence and accuracy (Campbell, 1983). Finally, this study will contribute additional information regarding head injury recovery called for by others in the field of neuropsychology and head injury rehabilitation (Campbell, 1983; Catron & Thompson, 1979; Matarazzo & Herman, 1984; Seidenberg et al., 1981; Shatz, 1981; Tabbaddor et al., 1984).

Research Questions

The thrust of this study is to: a) provide descriptive information regarding the stability of test scores of individuals recovering from traumatic-brain-injuries within a one year period of time; b) attempt to determine the existence, extent, and magnitude of test-retest practice effects resulting from repeat administrations of the WAIS-R over time; and c) provide information concerning the test-retest reliability of the WAIS-R across serial testing.

In essence, psychologists tend to measure individual traits such as intelligence over time to determine the course of recovery, and to make determinations and/or recommendations regarding treatment (Campbell, 1983). In order to do so, test-retest procedures are employed. Since psychological tests may provide individuals with opportunities

to remember certain responses as a result of having taken the test previously, the trait being measured may be influenced by the cumulative effects of practice (Brooks et al., 1984; Catron & Thompson, 1979; Seidenberg et al., 1981).

In measuring the intellectual recovery of TBI survivors, psychologists may wish to measure change in individual traits over time. Following head trauma, intellectual changes may occur as a result of many factors, including spontaneous recovery and treatment intervention, in addition to test-retest practice effects (Campbell, 1983). However, research has suggested that gains in recovering TBI survivors are primarily the result of improved cortical functioning, and that test-retest practice effects are not as evident as in normal populations (Campbell, 1983; Shatz, 1981; Warner, 1983). It is also questionable whether it is appropriate to compare and contrast normal subjects with clinical populations, as has occurred elsewhere in studies investigating recovery and practice effects (Brooks et al., 1984; Shatz, 1981), since the use of a nonimpaired control or comparison group may actually obscure true improvement in a brain-impaired experimental group (Shatz, 1981).

The omnibus research question asks whether there is an observed test-retest practice effect in addition to the course of natural recovery in surviving traumatically-brain-injured individuals. Specifically, the three major research questions are:

- (1) What is the magnitude of change in total intelligence scores over time for TBI survivors who have been tested two and four times, respectively?
- (2) Do the IQs and subtest scaled scores measured by the WAIS-R, differ among the recovering TBI survivors who were administered the same test twice and those administered the same test four times?
- (3) Is there an interaction between test-retest gains on total intelligence measures over time and how often a test is administered?

Secondary research questions are:

- (1) What are the test-retest reliabilities (correlation coefficients) of the WAIS-R for this clinical population and do they differ among the TBI survivors who have been tested two and four times, respectively?
- (2) If test-retest gains do exist in the group tested more often, are the cumulative effects of these gains on intellectual measures over time linear and/or curvilinear?

- (3) What is the internal consistency of results obtained on the WAIS-R on TBI survivors, and does this consistency change over time in the group tested more often?

Research Hypotheses

To provide data that will permit an empirical test of the research questions, an experimental group consisting of adult TBI survivors tested four times with the WAIS-R at 2, 4, 8, and 12 months post-injury will be compared with a control group of similar adult TBI survivors. The control group will have been tested two times with the WAIS-R at 2 and 12 months post-injury.

Since the study will use retrospective data the latter group is not a true control group but rather a matched group of subjects incurring similar treatment effects. For simplicity, however, the words "experimental" and "control" will be used throughout this text to differentiate between the two groups.

From the research questions for the study, the following (univariate) directional hypotheses are advanced:

- (1) Full Scale, Verbal, and Performance IQ means for the experimental group at 12 months post-injury will be greater than their respective means at 2 months post-injury. The same hypothesis is advanced for the control group.

- (2) Subtest scaled score means for the experimental group at 12 months post-injury will be greater than their respective subtest scaled score means at 2 months post-injury. The same hypothesis is advanced for the control group.
- (3) Full Scale, Verbal, and Performance IQ means of the experimental group will be greater than the Full Scale, Verbal, and Performance IQ means of the control group at one year post-injury.
- (4) All subtest scaled score means in the experimental group will be greater than all subtest scaled score means of the control group at one year post-injury.
- (5) Pretest posttest differences in IQ and subtest scaled score means over time in the experimental group will be greater than the corresponding differences in the control group (i.e., there will be an interaction of time and number of testings).

Secondary hypotheses:

- (1) Test-retest reliabilities on the WAIS-R for the experimental group will differ from the test-retest reliabilities of the control group at one year

post-injury for Full Scale, Verbal, and Performance IQs and all subtests.

- (2) Changes in Full Scale, Verbal, and Performance IQ means in the experimental group will be quadratic over a one year period post-injury.
- (3) Indices of internal consistency in the WAIS-R for the experimental group will change over the four test administrations.

CHAPTER II

Literature Review

Introduction

The review of related literature is organized under ten major headings. These are: (1) neuropsychological assessment; (2) reliability; (3) the Wechsler Adult Intelligence Scales; (4) test-retest reliability of the Wechsler Scales; (5) comparison of the WAIS and the WAIS-R; (6) practice effects; (7) practice effects on the Wechsler Scales; (8) studies of gain scores not involving normals; (9) factors which influence practice effects; and (10) intellectual recovery in traumatically-brain-injured survivors. Research findings in these areas provide the basis for the research questions and hypotheses for this study.

Neuropsychological Assessment

Neuropsychological assessment entails the evaluation of brain behavior relationships in neurologically-impaired populations. In addition to facilitating diagnostic formulations about neuropathological conditions (Crockett, Clark & Klonoff, 1981; Fuld, 1984; Furst, 1985; Golden, 1976; Lezak, 1983; Wolfe, 1987), neuropsychological assessment attempts to elucidate the extent to which insult to cortical

structures compromise an individual's functioning (Crockett et al., 1981; Wolfe, 1987). In this regard, the extent to which cognitive deficits arise from the insult are established as well as the magnitude of the insult itself (Crockett et al., 1981). In addition, strengths and weaknesses in the skills of the individual are evaluated as precisely as possible (Wolfe, 1987).

Precise information about an individual's cognitive status is essential for careful management and treatment of several neurological disorders including head injury (Lezak, 1983). Rational treatment planning and care, according to Lezak (1983), depends on an understanding of an individual's capabilities, limitations, and potential for maximizing compensatory strategies in a treatment regimen. Information of this type is also important to rehabilitation staff in order to choose appropriate treatment modalities where necessary (Dikmen et al., 1983), to identify goals for cognitive rehabilitation (Lezak, 1983), and to establish realistic parameters in rehabilitation planning (Tabaddor et al., 1984).

Neuropsychological assessment also has an important role to play for the brain-injured survivor by providing that individual with information regarding his or her performance (Lezak, 1983). In this way, he or she may begin to understand the nature and extent of his or her difficulties and set realistic goals and expectations for the future (Levin et al., 1982). In addition, the family must also be

educated as to the extent and nature of the sequelae secondary to the insult so that they may deal with the patient and the deficits appropriately through the recovery period and beyond (Lezak, 1983). Furthermore, neuropsychological assessment may help the patient and the family decide how to arrange the environment to the patient's best advantage and how to offer support when needed (Wolfe, 1987).

In the past, neuropsychological assessment has been devoted largely to defining the nature of cerebral deficits and has been less applicable to monitoring and charting recovery (Diller & Ben-Yishay, 1983). This is a fact that has been conceded by other prominent authors (Dikmen et al., 1983; Eson, Yen & Bourke, 1978; Miller, 1979; Rutter, Chadwick, Shaffer & Brown, 1980; Seidenberg et al., 1981; Tabaddor et al., 1984). This trend, however, is changing. The growing emphasis in the health and rehabilitation fields on continuity of care and importance of prolonged monitoring of patient behavior suggests an important role for serial or successive testing information in a number of different contexts (Seidenberg et al., 1981).

Serial neuropsychological assessment repeated at regular intervals can provide reliable indications of whether an underlying neurological condition is changing, and if so, in what direction (Lezak, 1983). Repeat assessment is also useful in documenting the effects of neurosurgery and psychopharmacological agents on behavior (Campbell, 1983;

Lezak, 1983). Information concerning intellectual functioning acquired through repeat administrations performed longitudinally also provides a basis for the continued refinement of treatment goals in cognitive remediation, rehabilitation, and vocational planning. Repeat testing in a similar vein may indicate an individual's potential for rehabilitation services assessed at various points in time (Tabaddor et al., 1984). In addition, testing performed on a repeated basis may provide information regarding the effectiveness of treatment such as cognitive retraining (Lezak, 1983), psychological interventions (Campbell, 1983), as well as the nature, course and rate of recovery from cerebral trauma itself (Seidenberg et al., 1981; Campbell, 1983; Tabaddor et al., 1984).

Matarazzo et al. (1980) reported that an increasing number of referrals for repeat neuropsychological assessment is occurring in order to document the extent, if any, of recovery. This, in part, seems to stem from the fact that although many tests lack parallel or equivalent forms, test-retest procedures are the most direct and accurate means of assessing changes in individual functioning (Seidenberg et al., 1981; Tabaddor et al., 1984). Unfortunately, without understanding the influences of repeat administrations on neuropsychological test results and the magnitude of error variance inherent in such testing, psychologists evaluating the effects of treatment or monitoring a patient's recovery will be unable to

determine the true nature of the changes observed in the patient's retest performances (Campbell, 1983; Matarazzo & Herman, 1984). In addition, interpretation of those observed changes are confounded by an absence of standardized data collected upon various populations which would otherwise provide normative information about test-retest scores (Matarazzo & Herman, 1984; Seidenberg et al., 1981).

In general, those studies which have attempted to investigate treatment effects, or course of recovery in neurologically-impaired populations such as TBI, suffer from a failure to account for the contaminating effects of test-retest practice and unreliability inherent in such serial testing (Campbell, 1983; Lawson, Inglis, & Stroud, 1983). More specifically Campbell states:

"Repeatedly administering a neuropsychological battery to brain damaged individuals requires that the neuropsychologist be aware of the reliability, the standard error(s) of measurement, and possible impact of practice, memory, and other sources of error variance on test results. Studies using repeated testing to examine the recovery process and/or efficiency of treatment interventions have largely ignored test reliability and the influence of memory and practice on test scores. Consequently, it is presently impossible to discern if the changes in neuropsychological performance are a result of treatment interventions, spontaneous recovery, practice effects, or the unreliability of the test instrument". (p. 7)

Campbell (1983) suggested that information concerning the internal consistency of a test, as well as the test reliability, standard error of measurement and practice effects are vitally important. Once the reliability of a test is established, the standard error of measurement can be

calculated and used to assess individual changes in neuropsychological assessment performances over time with confidence and accuracy. Lawson et al. (1983) also reported that research must also take into consideration regression effects on retest; a phenomenon that is a consequence of the imperfect correlation between test and retest data which is predictable from stability coefficients.

In addition to those concerns noted above, studies of head injury recovery suffer from the following methodological problems as noted by Levin et al. (1982) and others (Dikmen et al., 1983; Williams, Gomes, Drudge & Kessler, 1984): (1) an absence of an appropriate control group; (2) failure to screen for pre-injury conditions that might potentially compromise cognitive efficiency; (3) lack of information concerning pre-injury cognitive abilities; (4) inadequate documentation of acute closed head injury severity; (5) lack of control of the post-injury test-retest interval; and (6) lack of serial testing to depict time course of recovery. Miller (1979) also asserted that the use of unreliable measures, as well as a failure to ensure an adequate range of values for the appropriate variables may likely attenuate any correlations.

Therefore, studies of test-retest procedures are needed to provide the practicing neuropsychologist with information regarding the performance characteristics of these procedures as well as a data base that would include the range of

normative changes observed on particular tests with particular populations over time (Seidenberg et al., 1981). Implicit here is the need for data on the crucial issue of the course of recovery following head injury (Rutter et al., 1980; Tabaddor et al., 1984) as well as those issues surrounding recovery itself (Diller & Ben-Yishay, 1983). In addition, studies must be conducted in such a way as to minimize several of the methodological concerns noted above so that generalization from any single report will not be unduly compromised (Levin et al., 1982).

Studies directed at increasing our knowledge regarding intellectual recovery over time in traumatically-brain-injured survivors through test-retest procedures, with an eye towards the contributing effects of test-retest reliability and practice effects, are vitally important. These studies may help to delineate factors that influence the magnitude of change and the adequacy of functioning at some specified time following injury, as well as developing recovery curves representing the time scale and the amount of spontaneous improvement that occurs from such deficits (Dikmen et al., 1983). Such systematic studies of the process of recovery may also provide a basis for establishing cognitive remediation programs (Tabaddor et al., 1984) and more effective treatment regimens (Eson et al., 1978).

Reliability

Despite the fact that test-retest techniques may be subject to contaminating practice effects, clinical practice often dictates the readministration of psychological tests (Klonoff, Fibiger & Hutton, 1970). Thus, in order to ascertain pathological trends in the individual patient over a period of time, it is necessary to first have normative data on the reliability of the test(s) (Klonoff et al., 1970).

The fact that two sets of measurements of the same features of the same individuals will never exactly duplicate each other is what is meant by unreliability (Stanley, 1971). Nevertheless, repeated measurements on a series of individuals will tend to demonstrate some consistency over time. This tendency, then, toward consistency from one set of measurements to another is referred to as reliability (Stanley, 1971). More specifically, reliability has been defined as the consistency of scores obtained by the same persons when re-examined with the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions (Anastasi, 1976). However, according to Stanley (1971), the evaluation of the reliability of any measure reduces to a determination of how much of the variation in the set of scores is due to certain systematic differences among the individuals in the group,

and how much is due to other sources of variation that are considered (for practical purposes) errors of measurement.

In psychometric theory, an individual's obtained score (X) is represented as the sum of the theoretical true score of that person (T), and an error of measurement of the test (E) (Anastasi, 1976):

$$X = T + E$$

A true score is the score that would be obtained on average if the same person could be tested a large number of times and all other circumstances remained constant. Errors of measurement is the variability of the values in the frequency distribution of repeated measurements (Stanley, 1971). It can be shown that the variance of the obtained scores of a group (σ_X^2) equals the variance of true scores (σ_T^2) plus the variance arising from errors of measurement (σ_E^2):

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

By definition reliability is defined as:

$$p = \sigma_T^2 / [\sigma_T^2 + \sigma_E^2] = \sigma_T^2 / \sigma_X^2$$

The numerical value of the reliability coefficient of a test corresponds to the proportion of the variance in test scores that is due to true differences within that particular population of individuals on the variable being evaluated by the

test (Stanley, 1971). The higher the reliability, the greater the consistency in measurement over time.

The $X = T + E$ expression represents the simplest possible decomposition of the test score variance. In practice there are many possible sources of variance of scores on a particular test which may affect reliability by influencing σ_X^2 and/or σ_E^2 . Freeman (1962) suggested that many sources might include, but are not restricted to the following: (1) actual differences among individuals in the general traits being measured; (2) differences in specific abilities required in a particular test; (3) "test-wiseness", or the converse; (4) "chance" acquisition of particular pieces of knowledge; (5) normal or expected fluctuations in performance from time to time; (6) personal characteristics of the examinee; (7) physical conditions of the testing environment; (8) guessing; and (9) the effects of practice (previous test taking).

There are a number of ways to estimate reliability, and the choices of one estimate over another is often a function of the test and/or the testing environment. Conceptually, estimates of reliability may be approached from two different viewpoints. Stanley (1971) refers to these as measures of intraindividual and interindividual variability, respectively. In the former, one is concerned with the actual magnitude of the error of measurement expressed in the same units as individual scores (Stanley, 1971), and stated in terms of the standard error of measurement.

For example, were a series of measurements to be taken, estimates of one's true score would be obtained. A true score would be the average score an individual would obtain if one's performance were observed through a long series of samples or trials assuming no test-retest practice effects or fatigue from testing (Cronbach, 1960; Wechsler, 1981). While an individual's true score would presumably remain constant from one measure to another, the obtained scores would be expected to vary from the true score from time to time (Freeman, 1968). The variability of obtained values in the frequency distribution is referred to as the variance error of measurement, and its square root is called the standard error of measurement (Stanley, 1971). The standard error of measurement, which is an estimate of the deviation of a set of obtained scores from "true scores" (Freeman, 1962), could then be used to estimate limits of the range of the true score for an individual with a given obtained score (Anastasi, 1968). Thus, the standard error of measurement can be useful in establishing limits or "confidence intervals" using an individual's observed score(s) and within which that individual's true score would be expected to fall with a specified level of confidence. Therefore, an over-emphasis on obtained scores is avoided (Naglieri, 1982).

The latter concept, that of interindividual variability, concerns the consistency with which an individual maintains his/her position in the total group upon repetition of a measurement procedure (Stanley, 1971). Perhaps the most

conceptually appealing method for finding the reliability of test scores is by administering the same test on a second occasion to the same sample, and correlating the two sets of scores. This yields a correlation coefficient referred to by Cronbach (1960) as the coefficient of stability since it reflects the stability of a particular performance. Unfortunately, test-retest reliability is the least frequently reported method of reliability testing due to the time and expense of locating and retesting a large number of subjects (Freeman, 1962; Warner, 1983). In addition, there is the possibility that the person or the trait being measured on the second test has changed as a result of having been previously tested (Campbell, 1983; Catron & Thompson, 1979; Freeman, 1962). This carryover effect is likely to make the test-retest correlation higher than other reliability indices (Cronbach, 1960; Freeman, 1962; Stanley, 1971).

Despite this, Derner, Aborn, and Canter (1950) stated that the stability of an individual's score over time is an important concern in the area of intelligence testing, so the test-retest technique should be the reliability method of choice. In addition, the subsequent readministration of a test has the advantage of providing completely equivalent test content on all occasions, which is an essential consideration in lieu of the costly and difficult task of developing an equivalent form (Freeman, 1962).

The distinction between interindividual and intraindividual variability is important, then, in differentiating

between psychometric and clinical reliability. As noted earlier, psychometric reliability is based on a test-retest correlation which is derived from a change in relative position from test to retest on a sample of test scores. This is not the same as clinical reliability; that is, the presence or absence of a clinically meaningful change in individual scores from test to retest (Matarazzo & Herman, 1984; Ryan et al., 1985). Without this differentiation, high reliability correlations reported in test-retest studies may mask gains that typically occur on measures such as intelligence tests (Catron & Thompson, 1979). These correlations could mislead by implying similarities between test-retest scores and therefore, one may be unaware of the gains due to the retest effect alone (Catron & Thompson, 1979; Matarazzo et al., 1980).

Wechsler Adult Intelligence Scales

The Wechsler Adult Intelligence Scale (WAIS) was published in 1955 to replace the Wechsler-Bellevue Intelligence Scale (W-B) (Wechsler, 1939). The WAIS, like the W-B, is composed of 11 subtests. As noted earlier, six of the subtests comprise the Verbal Scale; five make up the Performance Scale; and all 11 are combined to make the Full Scale. A number of changes occurred in the development of the WAIS, with many subtests undergoing revision. For example, all items on the Vocabulary subtest were newly written.

The WAIS was standardized based on the reports of the 1950 United States Census. Norms were developed for seven age groups ranging from 16 to 64 with an equal number of men and women included in each age group. Standardization also included a mixture of races commensurate with proportions available from the 1950 census.

Reliability coefficients for the individual tests as well as the Verbal, Performance, and Full Scale IQs were provided for three age groups 18 to 19, 25 to 34, and 45 to 54. Table 2.1 presents average reliability coefficients and standard errors of measurement for both the WAIS and the WAIS-R. Average reliability coefficients for the WAIS Verbal IQ were .96; Performance IQ, .93; and Full Scale IQ .97 for the age groups noted above, respectively. Average subtest reliabilities ranged from .67 in Digit Span and Picture Arrangement to .95 for Vocabulary. Standard errors of measurement for Verbal, Performance, and Full Scale IQs average 3.00, 3.86, and 2.60 IQ points, respectively for the three age groups. These values indicate that the chances are about two out of three that an individual's obtained score lies within one standard error of measurement (plus or minus) of his or her "true" score. Average standard errors of measurement for the individual subtests ranged from a high of 1.70 scaled scored points for Digit Span to a low of .68 for Vocabulary.

Table 2.1

Average Reliability Coefficients and Standard Errors of Measurement of the WAIS and WAIS-R

	WAIS		WAIS-R	
	\underline{r}_{11}^a	\underline{SE}_m^b	\underline{r}_{11}	\underline{SE}_m
Information	.91	0.87	.89	0.93
Digit Span	.67	1.70	.83	1.23
Vocabulary	.95	0.68	.96	0.61
Arithmetic	.82	1.32	.84	1.14
Comprehension	.78	1.43	.84	1.20
Similarities	.85	1.19	.84	1.24
Picture Completion	.83	1.16	.81	1.25
Picture Arrangement	.67	1.61	.74	1.41
Block Design	.84	1.20	.87	0.98
Object Assembly	.68	1.63	.68	1.54
Digit Symbol	.92	0.85	.82	1.27
Verbal IQ	.96	3.00	.97	2.74
Performance IQ	.93	3.86	.93	4.14
Full Scale IQ	.97	2.60	.97	2.53

Note. All reliabilities are based on split-half procedures except Digit Symbol.

^a \underline{r}_{11} = reliability. ^b \underline{SE}_m = standard error of measurement.

The WAIS was succeeded by its revision, the WAIS-R (Wechsler, 1981). Like the WAIS, the WAIS-R is composed of 11 subtests; six verbal and five non-verbal. The individual subtests also result in a Verbal and Performance IQ, yielding a Full Scale IQ value. Wechsler (1981) reported that due to the organizational structure of the test, sections of the WAIS-R can be used alone when subjects are handicapped. For example, the Performance section could be used alone with individuals who are unable to comprehend or express language while the Verbal section could be used alone with those who are visually or motorically disabled. Also important is the use of both the Verbal and Performance sections to examine more comprehensively an individual's capabilities (Wechsler, 1981).

Most of the content of the WAIS was retained by the WAIS-R. Items that were dated were either dropped or revised. In all, 80% of the items on the WAIS-R were obtained from the WAIS (Wechsler, 1981). The sequence of administration was also changed so that the Verbal and Performance subtests are alternated. Previously on the WAIS, Verbal subtests were completed in total and were followed by the Performance subtests. Standardization was representative of the U.S. late adolescent and adult population during the 1970's (Sattler, 1988).

Reliability coefficients were provided by Wechsler (1981) for nine age groups. Average reliability coefficients for WAIS-R Verbal, Performance, and Full Scale

IQ values were .97, .93, .97, respectively (see Table 2.1). Average coefficients for the subtests ranged from .68 for Object Assembly to .96 for Vocabulary. Average standard errors of measurement for the subtests ranged from .61 in Vocabulary to 1.54 scaled scored points for Object Assembly. Standard errors of measurement averaged 2.74, 4.14, and 2.53 IQ points for Verbal, Performance, and Full Scale IQs, respectively.

Correlations between the WAIS and the WAIS-R IQs and subtests were also computed. Correlations among the Verbal, Performance, and Full Scale IQs were .91, .79, and .88, respectively. Subtest correlations between the WAIS and the WAIS-R ranged from .50 on Picture Arrangement to .91 for Vocabulary.

Intellectual determinations in normal populations have frequently utilized the Wechsler Scales; most notably the WAIS, but more recently the WAIS-R. Intellectual functioning is often seriously affected by many types of brain injury, and intelligence tests like the WAIS and the WAIS-R have been used to assess general intellectual and specific cognitive abilities of traumatically-brain-injured survivors (Campbell, 1983). In fact, the Wechsler Scales constitute a substantial portion of the test framework of a neuropsychological evaluation and have been the intellectual ability tests of choice for many neuropsychologists (Lezak, 1983). Several prominent authors have also included the Wechsler Scales in both clinical as well as research test

batteries (e.g., Reitan & Davison, 1974; Smith, 1975; Kaplan, 1988). In addition, a survey of the National Academy of Neuropsychologists indicated that the Wechsler Scales were employed in 97% of all neuropsychological evaluations performed by their members (Seretny et al., 1986).

The popularity of the Wechsler Scales is also demonstrated by the vast research and universal acceptance that have evolved from them (Lezak, 1983). When employed in neuropsychological practice with TBI survivors, the Wechsler Scales can be used as a guide to the level of ability achieved and the range of general knowledge acquired by the patient prior to their head injury (Newcombe, 1982). In addition, the scales may also serve as a useful criterion for the evaluation of selective post-traumatic deficits (Newcombe, 1982; Lezak, 1983), to recommend and later modify treatment goals, to monitor improvement and treatment effectiveness (Lezak, 1983), and make vocational and/or medico-legal determinations.

The scales are used frequently as cognitive outcome measures, particularly for the determination of long-term recovery following TBI (Levin et al., 1982). These authors have suggested that the Wechsler Scales permit a direct comparison of different abilities, and the interpretation of scores is facilitated by the availability of comprehensive standardized data for a wide range of ages. The ecologic validity of the scales with respect to the global quality of

adjustment in the community following head injury has also been reported as notable (Levin et al., 1982).

Test-Retest Reliability of the Wechsler Scales

Since the reliability of neuropsychological tests should be a concern for both the clinician and researcher, it is important to know the reliability of those tests (Parsons & Prigatano, 1978). However, despite the Wechsler Scales' importance to clinical and forensic psychology, and clinical neuropsychology, Matarazzo et al. (1980) reported that relatively little has been published on the specific issue of test-retest reliability of these scales, a view which is supported by others (Brown & May, 1979; Campbell, 1983; Matarazzo & Herman, 1984; Wagner & Caldwell, 1979). The few published studies have been severely compromised by methodological shortcomings; mainly, highly variable test-retest intervals, a lack of regard for differences among various clinical populations, and low numbers of subjects. Furthermore, there have been very few studies involving brain-damaged subjects and only a very restricted number of traumatically-brain-injured survivors. Without this type of information on brain-injured survivors, the clinical neuropsychologist is likely to be quite handicapped (Matarazzo & Herman, 1984).

Matarazzo et al. (1980) located only 11 studies that reported test-retest reliabilities for the Verbal, Performance and Full Scale IQ values of the WAIS. Table 2.2

lists these test-retest reliabilities in addition to other studies not included in the Matarazzo et al. (1980) review. The Matarazzo et al. (1980) article included a wide range of ages (mean ages 19 to 70), normal and clinical populations, and varying numbers of subjects per study (range 10 to 120). Test-retest intervals ranged from 1 week to 676 weeks. In addition, two studies reevaluated patients more than once.

Matarazzo et al. (1980) reviewed the results from all 11 studies and reported a remarkably high test-retest stability for the three WAIS IQ scores. The authors reported that a majority of correlation coefficients were in the .80's and .90's with the median correlation values for Verbal, Performance and Full Scale IQs at .89, .85, and .90, respectively. Studies of normal subject populations reported correlation coefficients averaging .82, .80, and .85 for Verbal, Performance and Full Scale IQs, respectively (Campbell, 1983). The Matarazzo et al. (1980) review also included studies that utilized psychiatric and mentally retarded subjects and reported average reliability coefficients of .91, .79, and .87, and .88, .91, and .90 for Verbal, Performance, and Full Scale IQs of the two groups, respectively. Reliability coefficients computed in various studies of elderly brain-damaged patients, chronic epileptics, and carotid endarterectomy patients averaged .89, .83, and .87 on Verbal, Performance and Full Scale IQs, respectively (Campbell, 1983). Overall, the lowest correlations for the entire range of studies reviewed by Matarazzo et al.

Table 2.2

WAIS IQ Test-Retest Reliabilities by Subject Population

Study/ Population	Age	<u>n</u>	Inter- val	VIQ	PIQ	FSIQ
Kangas Bradway, 1971 ^a normal adults	42	48	676	.70	.57	.73
Matarazzo et al., 1973 ^a normal job applicants	24	29	20	.87	.84	.91
Catron, 1978 normal college students	20	35	0	.86	.86	.88
Catron & Thompson, 1979 ^a normal college students	19	19	1	.91	.87	.94
		19	4	.87	.79	.83
		19	8	.72	.82	.74
		19	16	<u>.83</u>	<u>.85</u>	<u>.90</u>
				^b .85	.78	.84
Coons & Peacock, 1959 ^a psychiatric patients	33	24	1	.98	.96	.98
Kendrick & Post, 1967 ^a depressed and normal elderly	70	30	6 ^c 6 ^d 12 ^e	.90 .95 .89	.85 .76 .66	- - -
Klonoff et al., 1970 ^a chronic schizophrenics	47	42	416	.80	.58	.71
Brown & May 1979 ^a psychiatric patients	44	50	100	.91	.90	.92

Table 2.2 (cont'd)

<hr/>						
Wagner & Caldwell, 1979						
neurotics	30	19	229	.73	.76	.73
personality disorders	25	18	229	.85	.79	.91
schizophrenics	23	20	229	.92	.81	.84
Warner, 1983						
alcoholics	42	16	3	.82	.86	.85
non-alcoholics ^f	41	14	4	<u>.83</u>	<u>.76</u>	<u>.82</u>
				^b .86	.80	.85
Rosen et al., 1968 ^a						
mentally retarded	24	120	130	.87	.92	.88
Rosen et al., 1974 ^a						
educably retarded	29	50	186	.89	.90	.91
Dinning, et al., 1977						
mentally retarded	34	204	128	-	-	.90
Spitz, 1983						
adolescent retarded	17	42	177	-	-	.75
adult retarded	21	23	183	<u>-</u>	<u>-</u>	<u>.88</u>
				^b .88	.91	.86
Kendrick & Post, 1967 ^a						
brain-damaged elderly	70	10	6 ^g	.81	.94	-
			6 ^h	.82	.90	-
			12 ⁱ	.87	.92	-
Dodrill & Troupin, 1975 ^a						
chronic epileptics	27	17	35 ^j	.89	.71	.84
			70 ^k	.95	.78	.91
			105 ^l	.94	.74	.89
			35 ^m	.83	.74	.77
			70 ⁿ	.88	.78	.83
			35 ^o	.96	.92	.96

Matarazzo et al., 1979a carotid endarter- ectomy	62	17	20	.91	.85	.92
Wagner & Campbell, 1979 "organics"	24	39	229	<u>.83</u>	<u>.76</u>	<u>.78</u>
				^b .88	.75	.86

Note. Mean ages are in years. Mean test-retest intervals are in weeks.

^aMatarazzo et al. (1980). ^bMean values. ^cFirst to second test administration. ^dSecond to third test administration. ^eFirst to third test administration. ^fPsychiatric and medical patients. ^gFirst to second test administration. ^hSecond to third test administration. ⁱFirst to third test administration. ^jFirst to second test administration. ^kFirst to third test administration. ^lFirst to fourth test administration. ^mSecond to third test administration. ⁿsecond to fourth test administration. ^oThird to fourth test administration.

(1980) were .70 for Verbal IQ, .57 for Performance IQ, and .71 for Full Scale IQ. The lowest Verbal and Performance IQ correlation values were reported on normals retested 676 weeks apart (Kangas & Bradway, 1971). The lowest Full Scale IQ correlation value was calculated from a group of chronic schizophrenics tested 416 weeks apart (Klonoff et al., 1970).

Of the 11 studies reviewed by Matarazzo et al. (1980), only one involved "brain-damaged" individuals; a group of 10 elderly subjects suffering from diffuse brain damage (chronic brain syndrome) who were tested 3 times with test-retest intervals of 6 weeks (Kendrick & Post, 1967). Verbal IQ correlation coefficients ranged from .81 between the first to second testing to .87 between the first and the third testing. Performance IQ correlation coefficients ranged from .94 between the first and second testing to .90 between the second and the third testing. None of the groups included in the Matarazzo et al. (1980) review appeared to have sustained a traumatic-brain-injury. Of the remaining multiple reevaluations, Kendrick and Post (1967) retested depressed and normal elderly patients three times at 6 weeks intervals. Dodrill and Troupin (1975) retested chronic epileptics four times at 35 week intervals, respectively.

Based on their findings, Matarazzo et al. (1980) concluded that the three WAIS IQ values possessed test-retest reliabilities which were "statistically and clinically

robust" (p. 92) and demonstrated high stability in both clinical and normal samples. This was based on a majority of values in the .80's and .90's despite the fact that many of the subjects were clinically quite disturbed and retest intervals varied considerably (Matarazzo et al., 1980). The authors also reported that the test-retest stability of the WAIS appeared to be as high for one age level as for another.

Only six of the eleven studies reviewed by Matarazzo et al. (1980) reported the test-retest stability of all 11 WAIS subtests (see Table 2.3). While respectable, Matarazzo et al. (1980) reported that all subtest reliabilities were lower than the values for the three scales to which they contributed individually. The authors reported the median correlation coefficient for the total group of subtests was .76. Median values for all eleven subtests were: Information, .90; Comprehension, .69; Digit Span, .73; Arithmetic, .74; Similarities, .81; Vocabulary, .85; Picture Arrangement, .69; Picture Completion, .68; Block Design, .76; Object Assembly, .73; and Digit Symbol, .80.

Matarazzo et al. (1980) explained that the lower retest reliabilities for the subtests were not surprising since each subtest represented a smaller sample of behavior than did the WAIS taken as a whole as well as the restriction of range involved in the conversion of raw scores to scaled scores. In addition, test-retest correlations based on small sample sizes demonstrate large sampling error which

Table 2.3

WAIS Subtest Test-Retest Reliabilities by Subject Population

Study/ Population	Inf.	Comp.	D. Sp.	Arith.	Sim.	Voc.	P.A.	P.C.	B.D.	O.A.	D. Sym.
Matarazzo et al., 1973 ^a applicants	.75	.59	.73	.66	.58	.79	.70	.71	.72	.73	.87
Catron, 1978 students	<u>.90</u>	<u>.87</u>	<u>.62</u>	<u>.71</u>	<u>.82</u>	<u>.91</u>	<u>.77</u>	<u>.62</u>	<u>.86</u>	<u>.61</u>	<u>.76</u>
	^b .82	.73	.68	.69	.70	.85	.73	.66	.79	.67	.82
Coons & Peacock, 1959 ^a psychiatric	.94	.89	.84	.94	.94	.95	.88	.87	.88	.86	.92
Klonoff et al., 1970 ^a schizophrenics	.79	.59	.82	.52	.63	.24	.69	.54	.91	.59	.63
Wagner & Caldwell, 1979 neurotics	.89	.63	.81	.36	.35	.64	.25	.66	.68	.45	.51
personality disorders	.85	.85	.81	.65	.56	.89	.15	.62	.83	.46	.84
schizophrenics	.90	.71	.70	.70	.64	.66	.38	.64	.90	.88	.82
Warner, 1983 alcoholics	.92	.66	.53	.90	.75	.87	.63	.79	.87	.70	.81
non- alcoholics ^c	<u>.73</u>	<u>.81</u>	<u>.77</u>	<u>.62</u>	<u>.83</u>	<u>.89</u>	<u>.69</u>	<u>.34</u>	<u>.77</u>	<u>.79</u>	<u>.68</u>
	^b .86	.73	.75	.67	.67	.73	.52	.59	.83	.68	.74
Rosen et al., 1968 ^a retarded	<u>.78</u>	<u>.60</u>	<u>.69</u>	<u>.69</u>	<u>.68</u>	<u>.80</u>	<u>.85</u>	<u>.81</u>	<u>.87</u>	<u>.87</u>	<u>.88</u>
	^b .78	.60	.69	.69	.68	.80	.85	.81	.87	.87	.88
Dodrill & Troupin, 1975 ^a chronic	.90	.81	.75	.83	.68	.85	.27	.68	.69	.64	.87
epileptics	.90	.67	.81	.62	.88	.95	.67	.64	.72	.61	.67
	.85	.65	.61	.82	.88	.94	.72	.74	.67	.73	.74
	.90	.80	.55	.74	.70	.86	.59	.66	.63	.71	.79
	.96	.69	.44	.89	.82	.83	.38	.78	.79	.80	.80
	.94	.76	.87	.83	.81	.92	.78	.64	.82	.82	.89

Table 2.3 (cont'd)

Matarazzo
et al., 1979^a

endarter- ectomies	.88	.70	.57	.42	.83	.84	.64	.50	.76	.67	.67
-----------------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Wagner &
Caldwell, 1979

organics	<u>.82</u>	<u>.69</u>	<u>.80</u>	<u>.86</u>	<u>.49</u>	<u>.77</u>	<u>.60</u>	<u>.62</u>	<u>.70</u>	<u>.75</u>	<u>.66</u>
----------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

^b	.89	.72	.68	.66	.76	.87	.58	.66	.72	.72	.76
--------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Note. Abbreviations in the heading represent the following WAIS subtests: Information, Comprehension, Digit Span, Arithmetic, Similarities, Vocabulary, Picture Arrangement, Picture Completion, Block Design, Object Assembly, and Digit Symbol, respectively.

^aFound in Matarazzo et al. (1980). ^bMean values. ^cPsychiatric and medical patients.

also results in a decrease in subtest stability. Thus, while the IQ retest reliabilities were viewed as being very high, the subtest test-retest reliabilities were considered only "moderately good" (p. 93); requiring caution when evaluating a change in subtest scores in the individual patient (Anastasi; 1976, Matarazzo et al., 1970; Matarazzo & Herman, 1984).

Studies not included in the Matarazzo et al. (1980) review which reported additional test-retest reliability information are also listed in Tables 2.2 and 2.3. Using a mixed out-patient sample ($N = 96$) with retest intervals ranging from 1 to 10 years, Wagner and Caldwell (1979) found WAIS IQ reliabilities of .89, .90, and .87 for Full Scale, Verbal, and Performance IQs, respectively. The authors indicated that these values compared favorably with those obtained for institutionalized samples and were not far removed from the reliability (split-half) data reported by Wechsler (1955). These authors also reported that the subtests demonstrating the lowest overall test-retest reliabilities were Arithmetic, Similarities, Picture Completion, and Picture Arrangement. As noted above, comparatively low reliabilities have been reported for Picture Arrangement although not so for the other three. Table 2.3 presents the test-retest reliabilities of the WAIS subtests.

Wagner and Caldwell (1979) also divided their population into four diagnostic groups consisting of "organics" (p. 132) ($\underline{n} = 39$), neurotics ($\underline{n} = 19$), personality disorders (\underline{n}

= 18) and schizophrenics (\underline{n} = 20), and found the reliabilities to be remarkably high. The organic brain syndrome (organics) group was "sometimes" (p. 132) predetermined by evidence of a known brain injury prior to testing although the etiological nature of the rest of this subgroup was not reported. The organic subjects yielded Verbal, Performance and Full Scale IQ coefficients of .83, .76, and .78, respectively. Subtest correlation coefficients were also determined and ranged from the subtest Similarities (.49) to Arithmetic (.86). Verbal IQ correlation coefficients were .73, .85, and .92 for the remaining groups consisting of neurotics, personality disorders and schizophrenics, respectively. Performance IQs were .76, .79, and .81 for the three groups, and .73, .91, and .84 for Full Scale IQ values, respectively.

Warner (1983) calculated retest correlation coefficients on WAIS IQs and subtests for a group of alcoholics (\underline{n} = 16) and non-alcoholics (\underline{n} = 14) (which was comprised of an assortment of patients with medical and psychiatric diagnoses). Mean test-retest intervals for the alcoholic and non-alcoholic groups were 22.0 and 24.9 days, respectively. Mean ages for the two groups were 42.1 and 40.7 for the alcoholic and non-alcoholic groups, respectively. Correlation coefficients for the alcoholic groups were .82, .86, and .85 for the Verbal, Performance, and Full Scale IQ values, respectively. Subtest correlations ranged from .53 (Digit Span) to .92 (Information). Non-alcoholics averaged

Verbal, Performance and Full Scale correlation coefficients of .83, .76, and .82, respectively. Subtest correlation coefficients ranged from .34 (Picture Completion) to .89 (Vocabulary).

If retest reliability studies for the WAIS are few, those for the WAIS-R are even more rare. Wechsler (1981) presented test-retest (stability) coefficients for the three IQs and 11 subtests for two retest groups (see Tables 2.4 and 2.5). The test-retest interval ranged from two to seven weeks. Seventy-one individuals were reevaluated in the 25 to 34 year old range, and 48 individuals from the 45 to 54 year old age group. Verbal IQ correlation coefficients were .94 and .97 for the two chronological age groups, respectively. Performance IQ correlation coefficients were .89 and .90, and Full Scale IQ correlations were .95 and .96 for the two respective groups. The younger group subtest correlation coefficients ranged from .69 (Picture Arrangement) to .93 (Vocabulary) while the older group subtest correlation coefficients ranged from .67 (Object Assembly) to .94 (Information).

Snow, Tierney, Zorzitto, Fisher, and Reid (1989) also studied normals but used a group ($N = 101$) of geriatric subjects with a mean age of 67.1 years. The test-retest interval was an average of 1.1 years and resulted in retest reliabilities of .86 for Verbal IQ, .85 for Performance IQ, and .90 for Full Scale IQ. The authors reported that these reliability values were "high over a 1-year interval"

Table 2.4WAIS-R IQ Test-Retest Reliabilities by Subject Population

Study	Sample	Age	n	Inter- val	VIQ	PIQ	FSIQ
Wechsler, 1981	normals	25-34	71	(2-7)	.94	.89	.95
		45-54	48	(2-7)	.97	.90	.96
Warner, 1983	alcoholics	42	16	3	.91	.95	.96
	non- alcoholics ^a	41	14	4	.90	.84	.90
Ryan, et al., 1985	mixed ^b	37	21	38	.79	.88	.86
Snow, et al., 1989	normals	67	101	57	.86	.85	.90

Note. Mean ages are in years. Mean test-retest intervals are in weeks.

^aPsychiatric and medical patients.
neurologically-impaired patients.

^bPsychiatric and

(p. 425), but were not as high as those reported by Wechsler (1981) for younger subjects. Snow et al. (1989) suggested that Wechsler's (1981) values were higher because his retest interval was much shorter. Subtest test-retest reliability values ranged from .51 (Comprehension) to .91 (Digit Symbol).

Employing the WAIS-R, Ryan et al. (1985) assessed test-retest stability in a sample of 21 psychiatric and neurological patients. The mean test-retest interval was 38 weeks (range two to 144 weeks). Group mean age and education were respectively, 37.38 and 12.19 years. Included in the group were the following diagnoses: alcoholism with organic brain syndrome (OBS) (\underline{n} = 4); adjustment reaction (\underline{n} = 3); head injury (\underline{n} = 3); post-traumatic stress disorder (\underline{n} = 2); schizophrenia (\underline{n} = 2); Alzheimer Disease or possible disease (\underline{n} = 2); and drug induced psychosis (\underline{n} = 1). Verbal, Performance, and Full Scale IQs yielded stability coefficients of .79, .88, and .86, respectively. Subtest correlation coefficients ranged from .45 (Similarities) to .90 (Information).

Warner, (1983) also conducted test-retest procedures with the WAIS-R on alcoholics (\underline{n} = 16) and non-alcoholics (\underline{n} = 14) in the same manner noted previously. Verbal IQ correlation coefficients for the alcoholic and non-alcoholic groups were .91 and .90, respectively. Performance IQ was .95 and .84 with Full Scale IQ correlation coefficients of .96 and .90 for the alcoholic and non-alcoholic groups,

Table 2.5

WAIS-R Subtest Test-Retest Reliabilities by Subject Population

Study/ Population	Inf.	D.S.	Voc.	Arith.	Comp.	Sim.	P.C.	P.A.	B.D.	O.A.	D. Sym.
Wechsler, 1981 normals	.88 .94	.89 .82	.93 .91	.80 .90	.79 .82	.82 .86	.86 .89	.69 .76	.91 .80	.72 .67	.86 .82
Warner, 1983 alcoholics	.97	.84	.95	.78	.80	.81	.78	.75	.73	.83	.88
non- alcoholics ^a	.91	.84	.93	.81	.66	.85	.83	.86	.67	.34	.87
Ryan et al., 1985 mixed ^b	.90	.75	.78	.74	.76	.45	.81	.79	.75	.77	.74
Snow et al., 1989 normals	.81	.66	.71	.72	.51	.65	.65	.74	.84	.71	.91

^aPsychiatric and medical patients. ^bPsychiatric and neurologically-impaired patients.

respectively. Subtest coefficients for the alcoholics ranged from .73 (Block Design) to .97 (Information). Non-alcoholic subtest coefficients ranged from .34 (Object Assembly) to .93 (Vocabulary).

Noting that client characteristics and time limitations may result in the omission of subtests, Mittenberg and Ryan (1984) determined that elimination of any one WAIS-R subtest left Full Scale IQ reliability essentially unchanged at .97 (split-half reliability). In addition, the authors reported that subtraction of any combination of Performance subtests produced no further reduction of the Full Scale IQ reliability due to the high Verbal IQ reliability of .97 (split-half). The omission of more than one Verbal subtest, however, affected reliability moderately, but five verbal subtests would have to be deleted before reliability fell below .95. This is particularly important with neurologically-impaired patients since their deficits may make them noncompliant or physically restricted such that it is difficult, if not impossible, for them to take all the tests (Lezak & Gray, 1984).

In summary then, both the WAIS and WAIS-R IQs possess high to moderately high test-retest reliabilities, with lower reliabilities noted for the 11 subtests. The lower subtest reliabilities as described by Matarazzo and Herman (1984) reflect the combination of restriction of range when converting raw scores to scaled scores as well as smaller sampling of behavior relative to the Wechsler Scales in

total. From this review, it would appear that in most cases Verbal IQ reliabilities are greater than Full Scale IQ reliabilities, and Full Scale IQ reliabilities are greater than Performance reliabilities (Campbell, 1983). The consistently lower Performance IQ reliability coefficients may be due to the time limits and novel problem-solving nature inherent in these tasks. This is in contrast to the Verbal subtests which have a greater number of items from which to sample behavior as well as the more engrained, well-learned and over-rehearsed content inherent in those verbal tasks (Fisher, 1985; Mandleberg and Brooks, 1975; Ruesch, 1944).

It is clear from this review that there is a limited number of test-retest reliability studies completed on the WAIS and even fewer on the WAIS-R. In addition, it appears that those results which have been published have been severely compromised by methodological shortcomings. Furthermore, there have been few studies which have involved brain-damaged subjects and only a very restricted number of traumatically-brain-injured survivors. Thus, there is a need to supply base-rate information to the practitioner relative to the Wechsler Scales, consisting of test-retest reliabilities for brain-injured populations (Brown & May, 1979; Campbell, 1983; Matarazzo & Herman, 1984; Ryan et al. 1985; Seidenberg et al., 1981). Further data on the reliability of these tests and changes in test performances over time can enhance clear interpretation and judgments about these changes with confidence (Campbell, 1983).

Comparison of the WAIS and the WAIS-R

Following the revision of a test instrument, it is not unusual to examine the relationship between the revised edition and its predecessor. This is helpful to establish criterion validity and to determine whether the two forms are equivalent and/or interchangeable. This information is important when a determination needs to be made on an individual who has been tested with both the original test and its subsequent revision. Clearly, an investigator would want to know how the performance or results on one test compared to the performance or results of the other. If the two tests were found to be comparable, and therefore interchangeable, the influence of possible test-retest practice effects might be minimized by alternating the two tests in repeat or serial evaluations.

An examination of the relationship between the WAIS and the WAIS-R occurred following the revision of the WAIS. Wechsler (1981) presented correlation coefficients of scaled scores and IQs on the WAIS with those on the WAIS-R using 72 cases with an age range of 35 to 44 (see Table 2.6). All subjects were administered the WAIS-R and the WAIS in counterbalanced order with a test-retest interval of three to six weeks. Verbal, Performance, and Full Scale IQ correlation coefficients between the WAIS and WAIS-R were .91, .79, and .88, respectively. Correlations on subtests between the two test forms ranged from .50 (Picture Arrangement) to .91 (Vocabulary). This was quite similar to a

Table 2.6

WAIS/WAIS-R Correlations

	Wechsler, 1981	Prifitera & Ryan, 1983	Urbina, et al. 1982	Rabourn, 1983
Information	.87	.91	.90	.91
Digit Span	.86	.85	.82	.92
Vocabulary	.91	.94	.95	.96
Arithmetic	.85	.78	.88	.93
Comprehension	.72	.86	.82	.93
Similarities	.71	.74	.82	.86
Picture Completion	.63	.83	.80	.88
Picture Arrangement	.50	.80	.75	.84
Block Design	.85	.89	.82	.97
Object Assembly	.66	.80	.57	.98
Digit Symbol	.79	.94	.72	.98
Verbal IQ	.91	.96	.95	.96
Performance IQ	.79	.87	.86	.94
Full Scale IQ	.88	.93	.94	.95

study by Urbina, Golden and Ariel (1982) on a heterogenous sample of 35 females and 33 males ranging in age from 16 to 74. Correlations between the WAIS Verbal, Performance and Full Scale IQ and those on the WAIS-R were .95, .86, and .94, respectively.

Despite these correlation coefficients, Wechsler (1981) found the resulting WAIS Verbal, Performance and Full Scale IQ values to be 6.9, 8.0, and 7.5 points higher, respectively than the corresponding IQs on the WAIS-R (see Table 2.7). Using 32 psychiatric and vocational counseling patients, Prifitera and Ryan (1983) found the WAIS-R Verbal, Performance and Full Scale IQs to be 7.59, 7.06 and 7.75 lower, respectively, than the corresponding IQs on the WAIS. Correlations for these same IQ values were .96, .87, and .93, respectively.

Rabourn (1983) used a format that combined the WAIS and the WAIS-R by administering items consistent with both tests concurrently, and adding items in each test which were not common to both. This was done primarily to reduce errors due to practice effects common in test-retest methods. Rabourn found correlations between the WAIS and WAIS-R Verbal, Performance, and Full Scale IQ values ($N = 52$) to be .96, .94, and .95, respectively. Resulting WAIS IQ values were 6.2, 7.6, and 6.7 points higher than the WAIS-R. Similarly, Smith (1983) used naive subjects ($N = 70$) for both WAIS and WAIS-R administrations and found that the

Table 2.7

WAIS Minus WAIS-R Difference Scores

	Wechsler, 1981	Prifitera & Ryan, 1983	Urbina, et al. 1982	Rabourn, 1983
Information	1.10	1.03	0.85	1.20
Digit Span	0.60	0.15	0.24	0.80
Vocabulary	1.80	1.85	1.49	1.50
Arithmetic	1.00	1.00	0.97	1.00
Comprehension	1.80	1.75	1.29	2.00
Similarities	2.20	2.10	1.62	2.10
Picture Completion	1.80	1.09	1.21	1.40
Picture Arrangement	0.80	0.47	0.66	0.50
Block Design	1.00	0.88	0.79	1.40
Object Assembly	1.30	1.19	1.16	1.30
Digit Symbol	1.80	1.01	1.66	1.40
Verbal IQ	6.90	7.59	5.43	6.20
Performance IQ	8.00	7.06	5.31	7.60
Full Scale IQ	7.50	7.75	5.28	6.70

WAIS-R resulted in significantly lower estimates of intellectual ability than did the WAIS. The WAIS-R Verbal, Performance and Full Scale IQ values were 8, 9, and 9 points lower than the WAIS values, respectively.

Using a neurologically-impaired population (N = 114) similar in age, years of education, occupation, race, sex, etiology and location of cerebral dysfunction, Kelly, Montgomery, Felleman, and Webb (1984) also found the WAIS IQs to be higher than WAIS-R IQs. The WAIS group demonstrated mean scores 6.42, 3.54, and 5.79 points greater than the WAIS-R on Verbal, Performance, and Full Scale IQs, respectively. Differences between the two groups were statistically significant for the Verbal and Performance IQ mean values.

These studies suggest that while the correlations between the two Wechsler scales are high, different relationships may exist between the WAIS and the WAIS-R at different points on the intelligence distribution (Sattler, 1988). Therefore, the WAIS and WAIS-R are not equivalent (Smith, 1983; Warner, 1983) or interchangeable (Sattler, 1988). Thus, the two tests should not be used in the serial testing of individuals since differences between the WAIS and the WAIS-R will confound the results and subsequent interpretation of those scores.

In an attempt to understand why the two tests differ, Smith (1983) reported that Wechsler tended to attribute the differences between the WAIS and WAIS-R IQ scores to changes

in the populations on which the two scales were standardized and not to changes in the scales themselves. Rabourn (1983) cited societal influences on the test items, changes in the construction of the test, attitudes towards testing, and possible changes in the training of test administrators as possible variables contributing to the differences between WAIS and WAIS-R scores. Furthermore, Prifitera and Ryan (1983) suggested the influences of wide-spread exposure to TV and other media in the 28 years between the two sets of norms, as well as an overall increase in the educational level of the population, were accountable. In any event, it is clear that the use of these two psychometric instruments alternatively could present a major confound in any attempt to document recovery or measure the effects of practice over time (Warner, 1983). In addition, the two tests should not be used interchangeably in clinical neuropsychological assessment and research (Kelly et al., 1984).

Practice Effects

Repeat administrations of tests to assess recovery following head injury are common and have proved valuable in monitoring recovery in traumatically-brain-injured survivors (Brooks et al., 1984). However, it is possible that whenever a test is repeated increased familiarity with that test may result in an improved performance on that test. The improvement in performance which reflects the influence of learning and positive carryover as a result of having been

exposed to those tasks on a previous occasion is referred to as a practice effect (Seidenberg et al., 1981). Catron and Thompson (1979) have suggested the term "retest effect" (p. 352) because of the variety of other variables beyond mere practice that contribute to the effect. Hence, the term test-retest practice effect will be used for this study.

The concerns regarding test-retest practice effects have been expressed by several authors (Catron & Thompson, 1979; Matarazzo et al., 1980; Matarazzo & Herman, 1984; Warner, 1983). Without being cognizant of these effects for a given instrument, or the accuracy of an obtained score, it is quite difficult to determine whether changes in test performance over time in a population (such as TBI survivors) are due to treatment, unreliability of the test, spontaneous recovery, or to practice effects (Campbell, 1983). For instance, if a clinician used cut-off points as the sole basis for making judgments with respect to normality of brain functioning, the existence of test-retest practice effects may result in judging more individuals to be normal on successive administrations of a test than on the first test without intervention (Dodrill & Troupin, 1975; Matarazzo, Matarazzo, Wiens, Gallo & Klonoff, 1976). In fact, Dikmen et al. (1983) suggested that test-retest practice effects may possibly mask the true slowing of recovery over time when, in fact, practice effects may emulate continued cognitive improvement by influencing

scores to increase. The possibility of test-retest practice effects requires that considerable caution be exercised before concluding that subjects have recovered to normal functioning without corroborative evidence even when IQ scores are back to premorbid and/or average levels (Becker, 1977; Dodrill & Troupin, 1975; Matarazzo et al., 1980; Matarazzo & Herman, 1984).

The potentially deleterious effects of test-retest practice effects on clinical decisions also suggest that these effects need to be separated from true intellectual performance. This, in turn, suggests the need for a statistical procedure that would permit variation in test scores associated with a practice effect to be removed from variation that reflects true intellectual performance and natural recovery. Unfortunately, such modeling is not currently possible, and thus, teasing out practice effects from test scores becomes a logical as opposed to a statistical process.

Another difficulty related to practice effects noted previously, is that high test-retest correlations can mislead clinicians by implying similarity between the test-retest scores when, in fact, scores may increase upon retesting and only the rank order of scores are similar (Catron & Thompson, 1979). Thus, with extremely high test-retest reliabilities one could conclude erroneously that the Wechsler Scales might produce a score on retest nearly identical in absolute value to the score obtained in

the initial test for the same individual and be unaware of the substantial gain due to the test-retest effect alone (Matarazzo et al., 1980). This caution is underscored by Ryan et al. (1979) who reported that the WAIS-R is less reliable clinically than one might infer from the test-retest coefficients alone, and that large pretest to posttest changes must be interpreted in conjunction with information from other specialized assessment procedures (Matarazzo & Herman, 1984).

Test-retest practice effects may, in turn, also influence test-retest reliability coefficients differentially. If the initial and repeat tests are close temporally, the individual may remember some of the answers. This carryover or test-retest practice effect may yield a spuriously high retest correlation (Campbell, 1983; Cronbach, 1960; Derner et al., 1950; Freeman, 1962). Derner et al. (1950) suggested that the consistency of the practice effect, however, can be assessed by making retest measurements at varying time intervals. It is expected that a longer test-retest interval will result in a lowering of the reliability coefficient (Cronbach, 1960; Freeman, 1962) and that test-retest practice effects will diminish with time (Anastasi, 1968; Campbell, 1983; Tuma & Applebaum, 1980). If, however, test-retest effects differentially influence test scores rather than increase each individual score to the same degree, then the test-retest reliability coefficient will decline because the original rank order of the

individuals on the first test will change (Campbell, 1983; Meer & Baker, 1967).

In general, test-retest effects may be due to such factors as environmental and situational variables as well as emotional or affective factors (Seidenberg et al., 1981). Numerous authors (Lezak, 1983; Matarazzo, 1972; Quereshi, 1968; Steisel, 1951) suggest that tests that have a large speed component may be particularly susceptible to test-retest practice effects. Steisel (1951) theorized that in a retest situation the reaction times of a subject would tend to be faster than at the original testing. Since additional credit is allowed for speed on some tests, this would allow for higher scores in the retest situation. Karson, Pool and Freud (1957) reported that tasks involving manipulation of test materials might increase transfer from test to retest. Furthermore, Lezak (1983) theorized that tests which require an unfamiliar or infrequently practiced mode of response, or have a single-solution, particularly if it can be easily conceptualized once it was obtained, were more likely to show significant practice effects. In addition, Warner (1983) hypothesized that the subtests most susceptible to practice effects would also be those that were the least reliable.

Practice Effects on the Wechsler Scales

Given that the WAIS and WAIS-R lack equivalent, alternative forms and are not interchangeable, many concerns have

been expressed about the use of either the WAIS or the WAIS-R in test-retest procedures. Catron and Thompson (1979) suggested that the experience of taking the test once forever alters any subsequent test results; especially Full Scale and Performance IQ scores, and when intervals are less than 4 months, the Verbal IQ as well. In fact, Wechsler consistently cautioned that a gain in IQ of about 5 points from test to retest should, in general, be considered a practice or retest effect rather than a clinically meaningful change in actual IQ (Matarazzo et al., 1980). Thus, while the Wechsler IQs and subtests appear to be "robust" (p.92) psychometrically, some scores change sufficiently to warrant considerable caution in making clinical judgments in isolation (Matarazzo & Herman, 1984); as discussed below.

Matarazzo et al. (1980) analyzed 10 of the 11 studies noted previously for changes in WAIS Verbal, Performance, and Full Scale IQ values between test and retest. The median gains for Verbal, Performance, and Full Scale IQ values separately were 2, 8, and 5 IQ points, respectively; while corresponding mean gains were 2.38, 6.08, and 4.08 IQ points, respectively. In addition, Matarazzo et al. (1980) reported the actual gain or loss in means from one study to another ranged from -5 to +11 IQ points.

The pattern of retest gains was also found to be quite consistent with the Performance Scale demonstrating greater gains on retest than the Verbal Scale; a finding that is congruent with previous research (Campbell, 1983; Catron &

Thompson, 1979; Derner et al., 1950; Matarazzo, 1972). This pattern also holds true, despite the length of the retest interval (Catron & Thompson, 1979). Table 2.8 lists the WAIS IQ gains on retest of those studies in addition to, and including, the Matarazzo et al. (1980) review.

Catron (1978) studied 35 male college students who were administered the WAIS twice in immediate succession. Significant increases were noted with retest gains of 3.1, 14.2, and 8.3 points for Verbal, Performance, and Full Scale IQ scores, respectively. Catron concluded that the gain in these scores represented the maximum amount of gain one could expect to find on retest since all practice effects, and influences due to recency and insight would be at a premium. Kangas and Bradway (1971) retested normal individuals ($N = 48$) 13 years apart and found gains of 5.8, 11.3, and 8.5 points for Verbal, Performance, and Full Scale IQs, respectively. In addition, Catron and Thompson, (1979) retested 76 male college students with a retest interval of either 1 week, 1 month, 2 months, and 4 months apart, respectively ($n = 19/\text{interval}$). They found that as the test-retest interval increased, the retest gain scores decreased in a decelerating curve, although the gains in Performance IQ remained greater than those noted in Verbal IQ.

Matarazzo et al. (1980) also reported that increases were observed on all 11 subtests on retest to some degree.

Table 2.8

WAIS IQ Test-Retest Gains by Subject Population

Study/ Population	<u>n</u>	Age	Educ.	Inter- val	VIQ	PIQ	FSIQ
Kangas & Bradway, 1971 ^a normal adults	48	42	-	676	5.80	11.30	8.50
Matarazzo et al., 1973 ^a normal job applicants	29	24	14	20	5.60	4.90	5.50
Catron, 1978 normal college students	35	20	12+	0	3.10	14.20	8.30
Catron & Thompson, 1979 ^a normal college students	19	19	12+	1	4.74	11.37	8.00
	19	19	12+	4	1.79	9.79	5.68
	19	19	12+	8	2.27	8.74	5.42
	19	19	12+	16	<u>0.85</u>	<u>8.00</u>	<u>4.21</u>
					^b 3.45	9.75	6.51
Coons & Peacock, 1959 ^a psychiatric patients	24	33	-	1	2.60	8.60	5.00
Kendrick & Post, 1967 ^a depressed/ normal elderly	30	70	10	6	3.27	3.56	-
Klonoff et al., 1970 ^a chronic schizophrenics	42	47	10	416	4.91	7.86	6.64
Brown & May, 1979 ^a psychiatric patients	50	44	-	100	2.30	3.20	2.80

Table 2.8 (cont'd)

<hr/>							
Wagner & Caldwell, 1979							
neurotics	19	30	-	229	^c 1.00	1.00	1.00
personality disorders	18	25	-	229	^c 6.00	4.00	5.00
schizophrenics	20	23	-	229	^c 6.00	3.00	3.00
total outpatient ^d	96	25	12+	229	5.73	6.48	5.48
Warner, 1983							
alcoholics	16	42	12	3	3.80	8.60	6.30
non-alcoholics ^e	14	41	13	4	<u>2.40</u>	<u>7.10</u>	<u>4.60</u>
					^b 3.80	5.34	4.42
Rosen et al., 1968 ^a							
mentally retarded	120	24	-	130	1.00	2.70	1.80
Rosen et al., 1974 ^a							
educable retarded	50	29	-	186	2.00	2.00	2.00
Dinning et al., 1977							
adult retarded	204	34	-	128	-	-	1.70
Spitz, 1983							
adolescent retarded	42	17	-	177	-	-	1.00
adult retarded	23	21	-	183	<u> </u>	<u> </u>	<u>0.52</u>
					^b 1.50	2.35	1.40
Kendrick & Post, 1967 ^a							
brain-damaged elderly	10	70	10	6	0.70	0.70	-
Dodrill & Troupin, 1975 ^a							
chronic epileptics	17	27	13	35	-4.77	-3.76	-4.47
Matarazzo et al., 1979 ^a							
carotid endarterectomy	17	62	9	20	2.40	4.90	3.60

Wagner &
Caldwell, 1979
organics

39	24	12+	229	^c <u>4.00</u>	<u>4.00</u>	<u>4.00</u>
				^b 0.58	1.46	1.04

Note. Mean age and education are in years. Mean test-retest intervals are in weeks.

^aMatarazzo et al. (1980). ^b Mean values. ^cIQ test-retest gains are estimates derived from reported subtest gains.

^dThe total outpatient sample includes neurotics, personality disorders, schizophrenics and organics. ^ePsychiatric and medical patients.

Using test-retest WAIS subtest scores of 29 normal individuals, Matarazzo et al. (1973) found that changes do take place after a 20 week test-retest interval, albeit small ones. Table 2.9 lists the test-retest gains on WAIS subtests in addition to other reported gain scores. Mean changes in individual scaled scores for each of the 11 subtests in the 1973 study were 0.24 (Information), 1.93 (Comprehension), 0.41 (Arithmetic), 0.90 (Similarities), 1.00 (Digit Span), 0.90 (Vocabulary), 1.41 (Picture Completion), 0.76 (Block Design), 0.17 (Picture Arrangement), 0.69 (Object Assembly), and 1.00 (Digit Span).

In addition, these authors reported that in gain or loss difference scores on the six Verbal subtests, 63% of all 29 subjects fell between +1 to -1 point on retest. Fifty four percent of all subjects fell between +1 and -1 point on retest on the five Performance subtests. Nevertheless, without intervention of any kind, Matarazzo et al. (1980) stated some individuals changed as much as 3 to 7 points in subtest scaled scores from test to retest. The authors suggested that these changes may be due to either motivational or test taking differences, practice or retest effects, test unreliability, or some other as yet undiscovered factor. As such, Matarazzo et al. (1980) advocated that practitioners use caution in interpreting WAIS test-retest changes and suggested that a change of 3 to 5 points or more in subtest scaled scores and a change of 15 points or more in an IQ score may be interpreted as

Table 2.9

WAIS Subtest Test-Retest Gains by Subject Population

Study/ Population	Inf.	D.S.	Voc.	Arith.	Comp.	Sim.	P.C.	P.A.	B.D.	O.A.	D. Sym.
Matarazzo, et al., 1973 ^a normal applicants	0.24	1.00	0.90	0.41	1.93	0.90	1.41	0.17	0.76	0.69	1.00
Catron, 1978 normal college students	<u>0.08</u>	<u>0.37</u>	<u>0.03</u>	<u>1.48</u>	<u>0.72</u>	<u>0.32</u>	<u>1.71</u>	<u>2.86</u>	<u>1.14</u>	<u>3.00</u>	<u>2.03</u>
	^b 0.16	0.68	0.46	0.94	1.32	0.61	1.56	1.51	0.95	1.84	1.51
Coons & Peacock, 1959 ^a psychiatrics	0.60	0.50	0.30	0.20	0.60	0.90	0.80	1.60	0.90	2.60	0.80
Wagner & Caldwell, 1979 neurotics	-0.21	0.26	0.31	0.48	-0.15	0.52	-0.05	0.15	0.00	0.21	0.73
personality disorders	0.78	0.28	0.62	0.39	1.66	1.34	0.95	1.05	1.11	0.17	0.83
schizophrenics	0.10	0.55	0.10	0.45	1.30	0.00	0.15	1.10	0.35	0.35	0.55
outpatients ^c	0.35	0.38	0.52	0.49	0.95	0.92	0.47	0.51	0.63	1.03	0.56
Warner, 1983 alcoholics	0.10	0.80	0.30	0.60	0.50	1.20	1.40	1.10	1.60	1.40	1.00
non-alcoholics ^d	<u>0.80</u>	<u>-0.50</u>	<u>0.10</u>	<u>0.50</u>	<u>0.70</u>	<u>0.80</u>	<u>1.00</u>	<u>0.70</u>	<u>1.00</u>	<u>1.80</u>	<u>1.10</u>
	^b 0.36	0.32	0.32	0.44	0.79	0.81	0.67	0.88	0.79	1.08	0.79
Matarazzo et al., 1979 carotid endarterectomy	0.10	1.10	0.00	0.00	0.00	0.70	1.50	0.00	0.60	0.80	0.40
Wagner & Caldwell, 1979 organic	<u>0.56</u>	<u>0.38</u>	<u>0.79</u>	<u>0.57</u>	<u>0.98</u>	<u>1.38</u>	<u>0.67</u>	<u>0.12</u>	<u>0.84</u>	<u>0.49</u>	<u>0.36</u>
	^b 0.33	0.74	0.39	0.28	0.49	1.04	1.08	0.06	0.72	0.64	0.38

^aMatarazzo et al., (1980) study. ^bMean values. ^cThe total outpatient sample includes neurotics, personality disorders, schizophrenics, and organics. ^dPsychiatric and medical patients.

potentially clinically important. This assertion was also made with regard to use of the WAIS-R (Matarazzo & Herman, 1984). Combining both the 25 to 34 and 45 to 54 year old age groups from the test-retest samples for the WAIS-R, Matarazzo and Herman (1984) determined that the mean change upon retest after 2 to 7 weeks for 119 subjects (see Table 2.10) was a gain of 3.3, 8.4, and 6.2 points for Verbal, Performance and Full Scale IQ values, respectively, which were similar to the median gains noted in the review of the published test-retest studies above. It is also noteworthy to mention, as before, that these changes in IQ occurred despite rather significantly high test-retest reliability coefficients. Similarly, the WAIS-R subtests also demonstrated test-retest changes despite what would appear to be acceptable psychometric test-retest reliability (see Table 2.11). Retest changes on the WAIS-R subtests in the Matarazzo and Herman (1984) study reflected the following gains: 0.6 (Information), 0.4 (Digit Span), 0.2 (Vocabulary), 0.6 (Arithmetic), 0.2 (Comprehension), 0.9 (Similarities), 1.1 (Picture Completion), 1.3 (Picture Arrangement), 0.7 (Block Design), 1.9 (Object Assembly), and 0.9 (Digit Symbol).

In summary, Matarazzo et al's. (1980) "rules of thumb" regarding the interpretation of change in either subtest or IQ scores as noted above are based primarily on studies involving normal and psychiatric patients. His conclusions however, may not be entirely applicable to brain-damaged

Table 2.10

Test-Retest WAIS-R IQ Gains by Subject Population

Study/ Population	<u>n</u>	Age	Educ.	Inter- val	VIQ	PIQ	FSIQ
Warner, 1983							
alcoholics	16	42	12	3	4.80	8.50	7.00
non-alcoholic ^a	14	41	13	4	3.50	9.00	6.70
Matarazzo and Herman, 1984							
normals	119	-	-	2-7	3.30	8.40	6.20
Ryan et al., 1985							
mixed ^b	21	37		38	2.91	4.52	3.86

Note. Mean ages and education are in years. Mean test-retest intervals are in weeks.

^aPsychiatric and medical patients. ^bPsychiatric and neurologically-impaired patients.

Table 2.11

WAIS-R Subtest Test-Retest Gains by Subject Population

Study/ Population	Inf.	D.S.	Voc.	Arith.	Comp.	Sim.	P.C.	P.A.	B.D.	O.A.	D. Sym.
Warner, 1983											
alcoholics	0.50	1.20	0.20	1.10	0.80	0.70	1.40	1.00	1.10	1.40	1.50
non-											
alcoholics ^a	1.00	0.50	-0.20	1.20	-0.30	1.00	1.10	1.50	0.80	2.00	0.70
Matarazzo & Herman, 1984											
normals	0.60	0.40	0.20	0.60	0.20	0.90	1.10	1.30	0.70	1.90	0.90
Ryan et al., 1985											
mixed ^b	0.38	0.52	0.34	-0.15	0.91	1.38	0.76	0.76	0.58	0.90	0.48

Note. Mean ages and education are in years. Mean test-retest intervals are in weeks.

^aPsychiatric and medical patients. ^bPsychiatric and neurologically-impaired patients.

subjects (Campbell, 1983; Warner, 1983) as will be seen below.

Studies of Gain Scores Not Involving Normals

Studies reviewed by Matarazzo et al. (1980) involving normal subjects demonstrated average mean gains of 3.50, 9.01, and 6.21 points for Verbal, Performance and Full Scale IQ values, respectively. Similarly, studies involving psychiatric subjects as reviewed by Matarazzo et al. (1980) reported average gains of 3.27, 5.80, and 4.81 IQ points for Verbal, Performance and Full Scale, respectively, when subjects were retested with the WAIS. Test-retest studies by Warner (1983) demonstrated gains of 3.8, 8.6, and 6.3 IQ points on Verbal, Performance and Full Scale IQ values, respectively for alcoholics ($n = 16$) and gains of 2.4, 7.1, and 4.6 IQ points, respectively, for a group of non-alcoholics ($n = 14$) on the WAIS. In the same study Warner also used the WAIS-R to determine test-retest effects in both alcoholic and non-alcoholic groups. Gains in the alcoholic ($n = 16$) group were 4.8, 8.5, and 7.0 for Verbal, Performance and Full Scale IQ values, respectively while means of 3.8, 9.0 and 6.7 were noted for the IQ values of the non-alcoholics, ($n = 14$), respectively.

Studies that involved people with mental retardation, however, did not demonstrate such large gains. Of those reports reviewed by Matarazzo et al. (1980) average test-retest gains on Verbal, Performance, and Full Scale IQ

scores were 1.50, 2.35, and 1.90 points, respectively. Dinning, Andert and Hustak (1977) also used the WAIS to retest 204 retarded adults with a retest interval of 32 months and found the mean Full Scale IQ change on the WAIS to be only 1.7 points. Furthermore, Spitz (1983) retested adult mentally retarded subjects approximately 3.5 years after the initial WAIS and found the IQs changed by only about 1 point. Bell and Zubek (1960) concluded that while practice effects may be a factor in subjects with average intelligence, it is difficult to believe that it would be a significant factor in people with mental retardation. Furthermore, the authors suggested that test practice, unless allied with coaching, does not bring about the amount of gains reported in normals.

A study by Matarazzo et al. (1979) with subjects (N = 17) undergoing carotid endarterectomies demonstrated WAIS retest gains of 2.4, 4.9, and 3.6 points for Verbal, Performance and Full Scale IQ values, respectively. A serial study of epileptics (N = 17) by Dodrill and Troupin (1975) however, demonstrated a loss on retest with the WAIS of -4.7, -3.7, and -4.4 points for Verbal, Performance and Full Scale IQs values, respectively. A partial explanation for this drop was the use of anticonvulsants in nine of the subjects whose Full Scale IQ dropped 6.44 points. Nevertheless, no explanation was given for the other eight subjects who were not taking medications and who dropped an average of 2.25 points (Shatz, 1981). In fact, it wasn't until the

third administration that Performance IQs exceeded scores on the first test and not until the fourth evaluation that Verbal IQ and Full Scale IQ exceeded the first test scores.

In another serial study, Kendrick and Post (1967) evaluated brain-damaged elderly subjects ($n = 10$) three times. Changes in WAIS Verbal IQ resulted in a gain of only 0.70 points on the first retest (second administration) and a loss of 0.70 on the second retest (third administration). Similarly, Performance IQ improved by only 0.70 on the first retest and demonstrated a subsequent loss of 0.80 points on the second retest. Therefore, the possibility exists that practice effects or gains on retest with persons with known cerebral dysfunction do not occur in as a reliable fashion as typically seen in healthy subjects (Shatz, 1981).

Shatz (1981) has also suggested that practice effects differ in both magnitude and time course between those subjects who are neurologically intact and those who experience cerebral dysfunction. Research further suggests that retest changes may not occur to the same extent in all populations (Ryan et al., 1985). As demonstrated previously, normal subjects seem to show greater practice effects than psychiatric patients while individuals with cerebral dysfunction show minimal practice effects from a single retesting. In fact, those retest increases noted in IQ scores in subjects with cerebral dysfunction may possibly reflect improved cerebral functioning rather than the practice effects

characteristically seen in normal subjects or controls (Shatz, 1981) as will be demonstrated below.

In a study of 47 patients with vascular disease, the WAIS was administered and subsequently retested several months later (Duke, Bloor, Nugent & Majzoub, 1968). The patients were divided into a small vessel disease group (SVD), a large vessel disease operated group (LVD-O), and a large vessel disease non-operated group (LVD-N). The LVD-O group underwent carotid endarterectomies to permit an increase in blood flow. Results on retesting indicated that the SVD group made significant gains on all three IQs, whereas the LVD-O group made significant gains on Performance, and Full Scale IQ (see Table 2.12). The LVD-N group made no significant gains. The authors surmised that the surgery on the LVD-O group created a condition wherein the LVD-O group was able to achieve a practice effect by stopping the deterioration of accompanying vascular disease whereas the LVD-N was unable to do so.

Ivnik (1978) tested and retested patients with multiple sclerosis (MS) (\underline{n} = 14) as well as non-MS neurological controls (\underline{n} = 14). Using a mean test-retest interval of 3 years Verbal, Performance, and Full Scale IQs decreased due to the worsening of MS. It was noted that the control group demonstrated gains of 1.5, 3.1, and 2.0 IQ points on the respective IQ scales. Ivnik suggested that the test-retest differences probably represented true improvement in neuropsychological functioning for the control group by the fact

Table 2.12

Test-Retest Gains of Impaired and Control Populations on
WAIS IQ Values

Study	<u>n</u>	Age	Educ.	Inter- val	VIQ	PIQ	FSIQ
<hr/>							
Duke et al., 1968							
SVD	19	59		47	3.58	6.90	5.73
LVD-O	16	58		76	2.19	10.19	6.38
LVD-N	13	57		75	-1.83	-0.63	-1.08
Ivnik, 1978							
non-MS	14	37	12	138	1.50	3.10	2.00
MS	14	38	13	148	-3.60	-3.70	-3.70
Seidenberg et al., 1981							
improved epileptics	22	22	11	74	3.59	10.23	6.86
unimproved epileptics	25	22	11	78	-1.08	4.28	1.24
Drudge, et al., 1984							
normal adults	15	25	13	0			
head-injured adults	15	25	13	36	13.70	25.40	20.60
					^a (-9.6)	(-12.3)	(-11.2)
Becker, 1977							
enlisted controls	10	23	12	9	3.20	6.40	5.00
head-injured enlisted	10	20		9	7.00	15.40	12.30
					^a (-8.5)	(-17.8)	(-13.5)

Note. Mean ages and education are in years. Mean test-retest intervals are in weeks.

^aIQ points below normal controls at time of retest.

that several of the neurological disorders in that group involved neurosurgical procedures from which recovery of functioning over time was anticipated.

Seidenberg et al. (1981) examined the WAIS test-retest performances of a group of epileptics ($N = 58$). The results indicated significant retest increases only on Performance IQ and Object Assembly for a portion of the sample for whom seizure activity did not improve. With epileptics for whom seizures did improve, significant increases were noted on retest in 11 of the 14 WAIS measures. These improvements included Verbal, Performance, and Full Scale IQ values, and the subtests Information, Similarities, Comprehension, Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Digit Symbol. It was concluded that these retest improvements were not due to practice effects since the extent of the improvement succeeded those attributable to the influence of practice as discussed by Matarazzo (1972) but also the test-retest interval of more than 20 months was longer than the period of time the effects of practice were typically expected to operate (Seidenberg et al., 1981).

Drudge, Williams, Kessler, and Gomes (1984) evaluated TBI survivors ($n = 15$) twice with the WAIS; once at 2.6 months and again at 11.5 months, post-injury. All changes between the initial and second tests were significant except for Digit Span. Retest gains were 13.7, 25.4, and 20.6 points for Verbal, Performance, and Full Scale IQ values,

respectively. Despite these changes on the second test, the TBI patients were inferior to controls tested only once on 9 of the 14 various WAIS scales and subtests. By testing controls only once, not twice, learning or practice effects were not controlled for, however the authors assumed that the improvement in the TBI subjects was the result of the recovery process. The assumption was based on the knowledge that TBI patients are most deficient in learning and memory; both of which are instrumental for test-retest practice effects. These cognitive deficits in the face of a nine month test-retest interval also suggested a reasonable reduction in the effects of incidental memory. This was also consistent with Schau, O'Leary and Chaney (1980) who suggested that practice effects would not likely be detectible over a one year period of time.

Becker (1977) evaluated and then re-evaluated a group ($n = 10$) of male enlisted patients with closed head injuries at approximately 2 and 13 weeks, post-injury, respectively. This impaired group was compared with a matched group ($n = 10$) of enlisted men who had no history of head injury, brain disease, or psychiatric diagnosis and who were tested at approximately the same times. Becker reported that the overall improvement shown by the closed head injured patients were not significantly greater than the control subjects leading the author to conclude that the "improvement" noted on retesting was attributable to practice effects and the experience of prior test taking. Despite

this contention, the head injury group's test-retest gains were more than twice the test-retest gains of the control group reflecting greater change relative to the much lower scores of the head injury group on initial testing than those of the control group. The head-injured group's initial Verbal, Performance, and Full Scale IQ values were 12.3, 26.8, and 19.8 lower than the matched control group, IQ values, respectively. Even with the large test-retest gains demonstrated by the head-injured group, IQ values were still 8.5, 17.8, and 13.5 lower than the matched group on Verbal, Performance, and Full Scale measures respectively, at retest. Therefore, test-retest gains were not equivalent and may have varied for reasons other than practice effects including genuine recovery.

Mandleberg and Brooks (1975) compared the three year post-injury follow-up of WAIS scores of TBI patients (N = 40) differing in the number of times (one to four) the WAIS had been given previously. The authors found no evidence for test-retest practice effects insofar as the three year WAIS scores were not related to the number of previous testings. Practice effects were also absent at 5 and 10, months post-injury as well. The authors suggested that the lack of evident practice effects in TBI subjects ran counter to findings in non-brain-injured groups. They surmised that in the latter, improved retest scores could be a function of late learning, increased test familiarity, and decreased

levels of anxiety, whereas those same factors were less powerful in the TBI subjects they had evaluated.

In a second study, Mandleberg (1976) compared the WAIS scores of TBI patients ($n = 51$) evaluated 4 times within a 30 month period, post-injury with those TBI subjects ($n = 98$) who had no prior exposure. Mandleberg reported that prior exposure to the WAIS did not significantly enhance IQ scores of the former group. This finding was congruent with Levin et al. (1982) who observed that severely impaired patients who were evaluated at 6 to 12 month intervals frequently denied having seen the test materials previously.

This is not to say, however, that practice effects do not occur in patients with cerebral dysfunction, rather that they do not manifest themselves to the same extent as do normal subjects (Shatz, 1981). In fact, multiple administrations of the Wechsler Scales over a period of several months probably will produce significant practice effects in patients with relatively mild injury, but not in severely injured patients who are serially tested at widely separated intervals (Levin et al., 1982).

Factors Which Influence Practice Effects

Just as test-retest practice effects influence various subject populations differently, other factors are thought to influence practice effects. The first of these variables may be age. Eisdorfer (1963) reported a gain of only .19 total scaled score points when 47 subjects (mean age 65

years) were retested 39 months later with the WAIS. In another sample of 41 subjects (mean age 74), Eisdorfer found a decrease of 1.05 total scaled scored points also at a 39 month follow-up. Rhudrick and Gordon (1973) reported test-retest data on 86 elderly patients (mean age 72) with the WAIS with a one to two year follow up. The results demonstrated a decline on Verbal IQ and a mean increase of 2.58 points in Performance IQ. This led Shatz (1981) to conclude that while practice effects cannot be totally ignored in the elderly, the extent of practice effects on this group are not as robust as the 5 point improvement in younger subjects suggested by Matarazzo et al. (1980). In fact, Matarazzo and Herman (1984) in their review of the WAIS-R standardization group suggested that age did not appear to influence retest gains or loss in the age groups 25 to 35, or 45 to 54; a fact that was also supported by Ryan et al. (1985) in their retest study of the WAIS-R in a mixed outpatient sample with a mean age of 37.38.

Education may also influence practice effects, although studies investigating this variable are scarce. Ryan et al. (1985) found that gain or loss on retest WAIS-R Full Scale IQ was strongly associated with years of education which suggested that prior learning history may influence the amount of incidental memory that occurs between the first and second testing. It has also been reported (Matarazzo & Herman, 1984) that mean Full Scale IQ values are typically higher in those who have completed more years of education.

Additionally, it has been suggested (Grafman, Salazar, Weingartner, Vance & Amin 1986; Steisel, 1951) that post-injury cognitive performances show greater correlation with pre-injury intelligence and that brighter individuals may benefit more from a retest situation than those less bright.

Different types of cerebral dysfunctioning and course of neurological processes are also likely to effect test-retest practice. Impairment of learning ability as a result of brain damage must be seen as a manifestation of the underlying pathology and may vary according to the extent and focus of the brain damage (Matarazzo et al., 1979; Shatz, 1981). Patient populations in some studies also provide certain limitations not likely to reveal the practice effects as do other groups (Dodrill & Troupin, 1975). These authors have suggested that very healthy individuals may do so well on testing initially that there is little room for test-retest practice effects and very impaired patients maybe so disturbed that they cannot profit from the initial testing to have performances improve on the second. The effects of drugs have not been fully investigated either, although one study (Dodrill & Troupin, 1975) suggested medications may, to some degree, minimize test-retest practice effects by making the subjects less able to profit from their test-taking experiences.

Length of the test-retest interval is another important variable related to test-retest practice effects (Catron & Thompson, 1979; Shatz, 1981) and one which appears to be

intimately linked with test-retest reliability. As the test-retest interval increases, practice effects typically decrease (Anastasi, 1968; Derner et al., 1950). This may reflect a concomitant decrease not only in the test-retest reliability but a lack of familiarity with the test instrument on retest. Decreasing the test-retest interval will increase the reliability coefficient by facilitating memory effects on the second test administration (Matarazzo et al., 1979; Warner, 1983). Increasing the retest interval allows a greater amount of "mental growth" or improvement to take place (Catron & Thompson, 1979) and in particularly long intervals, subjects may cross from one age group to another with the likely effect of decreasing the consistency in IQ scores. Shatz (1981) in turn, has suggested that an examination of either raw or scaled scores as opposed to age weighted scores be performed when interpreting individual or group performances over time since increases in IQ scores may be due to an aging artifact (Klonoff et al., 1970).

Obviously then, test-retest practice effects are not the same for all populations nor are the influences of those variables noted previously. Therefore, research must be encouraged to collect and publish data in samples of patients with known cerebral dysfunction to establish the necessary foundation for the evaluation of WAIS-R test-retest practice effects in neuropsychological assessment (Shatz, 1981). In this regard, Shatz has suggested the use of non-impaired control groups be avoided in studies using

cerebrally-impaired subjects since the former may obscure the improvements of the brain-injured group.

Intellectual Recovery in TBI Survivors

While it is important to understand the extent to which individuals recover from intellectual impairment due to head injury, relatively few studies have been conducted in this area until recently (Brooks, 1984). In the 1930's and 1940's a few researchers such as Conkey (1938), Ruesch (1944), Ruesch and Moore (1943) and Tooth (1946) investigated the nature and occurrence of intellectual impairment recovery. It was not until the 1970's that interest in intellectual impairment and recovery in the TBI survivor was rekindled (Brooks, 1984). In their review of the published literature on intellectual recovery following traumatic-brain-injury, Levin et al. (1982) identified 17 reports; 10 of which employed the WAIS. Unfortunately, many of these studies demonstrated the methodological difficulties described previously. In addition, only four studies involved retesting with the WAIS (Becker, 1977, Mandleberg, 1975; Mandleberg & Brooks, 1975; and Mandleberg, 1976). Since then, only a few reports investigating intellectual recovery in TBI patients with the WAIS have appeared in the literature (Diller & Ben-Yishay, 1983; Drudge et al., 1984; Tabaddor et al., 1984). At the time that the research was reviewed, no studies with the WAIS-R related exclusively to the intellectual recovery of TBI patients had been

published. From this limited sample and other literature related to intellectual recovery, the following information has been obtained.

Following traumatic-brain-injury, cognitive functions can be seriously compromised (Campbell, 1983). Despite the variety of insult and severity of injury, certain patterns of intellectual deficits manifest themselves in the recovering TBI survivor. In general, patients with left hemisphere damage obtain significantly lower Verbal IQs in comparison to their own Performance IQs although both may be suppressed (Bornstein, 1983; Tabaddor et al., 1984). Patients for whom either right hemisphere damage or bilateral damage has occurred; patient's Performance IQ scores will typically be significantly lower in comparison to their own Verbal IQ scores although both may be decreased (Becker, 1977; Bornstein, 1983; Drudge et al., 1984; Dye, Saxon & Milby, 1981; Fisher, 1985; Mandleberg, 1976; and Tabaddor et al., 1984). Nevertheless, it would appear that many more cases appear to follow the latter pattern (lower Performance scores) rather than the former as a result of the more frequent and diffuse nature of traumatically-induced brain-injury.

This pattern of greater impairment in Performance IQ than Verbal IQ has been thought to reflect the overlearned, well established, and more resilient nature of the verbal information requested in the Verbal section of the Wechsler Scales through simple, readily elicited responses (Fisher,

1985; Mandleberg & Brooks, 1975; Ruesch, 1944). In contrast, the Performance IQ subtests rely on the integration of many complex functions including motivation, attention, sustained effort, perception, dexterity, perception and speed (Brooks, 1975; Mandleberg & Brooks, 1975); any one of which might be in itself sufficiently impaired to depress Performance skills (Mandleberg & Brooks, 1975). In addition, Seidenberg et al. (1981) reported that new learning situations inherent in the Performance sections are also more vulnerable to the effects of brain damage.

Recovery of intellectual deficits found in TBI survivors also reveal rather characteristic patterns; namely, that the return of intellectual abilities is rapid and the final level is reached earlier for Verbal than Performance, or non-verbal skills (Bond & Brooks, 1976; Diller & Ben-Yishay, 1983). This may be due to the fact that Verbal subtests on the Wechsler scales are less impaired initially and hence return quicker than those functions tested by the non-verbal subtests (Bond & Brooks, 1976; Eson et al., 1978). In turn, test-retest improvements on the Performance sections are greater relative to the Verbal measures whose changes demonstrate minimal increases (Diller & Ben-Yishay, 1983; Seidenberg et al., 1981). Greater gains in the Performance subtests may, of course, reflect the concept of regression-to-the-mean where, upon repeat testing, there is a greater tendency for more impaired scores to converge towards the mean in the retest situation (Eisdorfer, 1963).

Therefore, in areas of greater brain impairment, there may be greater cognitive improvement observed with the repetition of the test (Lezak, 1983).

A review by Diller and Ben-Yishay (1983) suggested that recovery of neuropsychological functions tends to be lawful; that is, individuals tend to maintain their standing as it regards performances on neuropsychological tests. In essence, those individuals whose performances are low after injury are more likely to stay low relative to the total group. The overall rate of return of intellectual functions also appears to be more rapid during the early stages and slows up considerably thereafter (Diller & Ben-Yishay, 1983). In general, researchers have suggested most changes occur during the first year after injury with the most rapid improvement of all skills within 3 to 6 months (Bond, 1975; Bond & Brooks, 1976; Diller & Ben-Yishay, 1983; Mandleberg and Brooks, 1975). This period of rapid improvement is also followed by a period of decelerating rate of improvement until asymptote is reached suggesting a quadratic trend in the recovery pattern (Dodrill & Troupin, 1975; Eson et al., 1978). Typically Verbal skills "plateau" around 6 months whereas Performance abilities reach plateau by approximately 13 months (Mandleberg & Brooks, 1975).

Despite the fact that significant gains may be made within one year, performances may remain impaired when compared to others or themselves premorbidly (Drudge et al., 1984), suggesting that TBI patients actually recover only a

portion of their initial loss over time (Tabaddor et al., 1984; Wagner & Caldwell, 1979). Therefore, the continued presence of residual deficits would represent the remaining, non-recovered portion of the initial loss (Dikmen et al., 1983). Thus, the notion of a true and total recovery may be tenuous particularly in the absence of concomitant positive behavior changes, and good estimates of pre-injury intellectual functioning of both verbal or non-verbal abilities despite normal test scores (Matarazzo et al., 1980; Matarazzo & Herman, 1984; Ruesch, 1944).

The extent to which an individual recovers also seems to be influenced by many of the same factors which may influence practice effects; namely age (Bond & Brooks, 1976; Brooks, 1986; Carlsson, von Essen & Lofgren, 1968; Lewin, Marshall & Roberts, 1979; Lundholm, Jepsen & Thornval, 1975; Heiskanen & Sipponen, 1970; Smith, 1974), premorbid intelligence (Bond & Brooks, 1976; Grafman et al., 1986; Williams et al., 1984), severity of injury (Bond & Brooks, 1976; Carlsson et al., 1968; Dye et al., 1981; Klove & Cleeland, 1972; Lundholm et al., 1975; Smith, 1961) and educational attainment (Finlayson, Johnson & Reitan, 1977; Grafman et al., 1986; Parsons & Prigatano, 1978). Other possible variables that may influence recovery and/or neuropsychological test performance include sex, socioeconomic level and occupation, handedness, duration of injury and experiences between test administrations, alcohol consumption, severity of brain damage, and psychiatric history (Heaton, Baade &

Johnson, 1978; Parsons & Prigatano, 1978). Another important variable interrelated with those factors noted above is memory (Fisher, 1985). Since many head-injured patients experience marked difficulties in memory as well as intellectual tasks, the capacity to efficiently problem-solve and learn new information, both verbally and non-verbally will be severely compromised (Finlayson & Block, 1982; Fisher, 1985).

In this regard, Levin, Grossman and Kelly (1976) confirmed findings in previous studies which demonstrated that persistent impairment of recognition memory continues several months or more following trauma. These authors also stated, however, that short-term recognition memory is less vulnerable to head injury than is recognition memory spanning longer intervals; a pattern also suggested by the investigation of recall memory. Brooks (1974) reported that the recovery of memory is dependent upon other factors other than time elapsing since the injury, and suggested that recovery may be completed within six months. Although rate of learning in patients with TBI was still lower than in controls, and group differences increased at each trial, Lezak (1979) indicated that simple learning does show an improvement pattern in TBI populations but at a much reduced rate. Lezak's research also suggested that few TBI patients improve their ability for verbal learning within the first three years. Thus, the ability to profit from previous exposure to test materials in a TBI population may be

diminished relative to normals suggesting that test-retest practice effects may not be so robust. They may increase, however, when shorter intervals are used between the test and retest.

Noting that very few studies have investigated the recovery of intellectual functions in surviving TBI populations with the WAIS, and literally none with the WAIS-R, there is a need for information about intellectual changes as measured by the WAIS-R in the aforesaid clinical population (Ryan et al., 1985). It is also necessary to stress the need to study the natural history of recovery before assuming that increased scores on repeat testing demonstrate more than a test-retest practice effect as demonstrated when normal subjects are retested (Newcombe, 1982). In this regard, it would be extremely important to distinguish between the process of recovery of intellectual functioning over time and the possible, concomitant process of test-retest practice effects (Brooks et al., 1984). Such a study might then contribute information helpful to the interpretation of scores specific to those individuals with TBI on whom certain medico-legal or psycho-vocational determinations may be based. Test-retest data as well as a set of curves that would document the kind of changes expected for a particular combination of test and time intervals would therefore be most beneficial (Catron & Thompson, 1979).

Few studies have attempted to determine the level of recovery in TBI patients by testing the same patients on a

number of occasions (Levin et al., 1982). Those who have used these procedures have encountered methodological problems in trying to separate the effects of test-retest practice and those due to genuine recovery (Brooks et al., 1984). As has been stated previously, the problem is not solved by the simple use of non-impaired retest control groups (Brooks et al., 1984; Shatz, 1981) since the cognitive levels and rates of change in the intellectual performances of the head-injured and control groups are always likely to be different at different stages during recovery.

Warner (1983) has reported that one way studies assessing long-term cognitive recovery in patient populations have avoided the problem of test-retest practice effects is by using retest intervals approximating a year or more (Drudge et al., 1984; Schau et al., 1980). Assuming one used a lengthy test-retest interval for a group of impaired control subjects; items from the test at the time of the second administration should be less familiar. If it could also be assumed that the scores from the first test administered to both an impaired experimental group and a similarly impaired control group were the same, (or nearly the same), a combined sequential and cross-sectional design could be a way of both charting recovery (Mandleberg & Brooks, 1975) and identifying the extent of test-retest practice at the same time (Brooks et al., 1984).

In this way, changes from the first test to the last test administration in both the control and experimental

groups could be used to measure the magnitude of recovery expected within a protracted period of time. Differences in the changed scores between the two groups, however, may yield objective findings that may indicate that those differences, if significant, are partly due to practice effects in addition to the natural course of recovery.

In summary, repeat neuropsychological assessment can be an important vehicle through which to document recovery from a neurological condition. However, when administering the same assessment tool in a repeated fashion, the clinician must be cognizant of the influences of repeated testing on obtained test scores. In addition, one must be aware of the error variance inherent in such testing and the reliability of the test itself.

A review of the literature reveals a few reports that address these important issues. As noted previously, high test-retest reliability coefficients in tests such as the Wechsler Scales may mislead clinicians by implying that test scores will be similar upon retest when, in fact, retest gains may be evident due to test-retest practice effects alone. In this regard, research involving normals has demonstrated gains of several IQ points despite high test-retest reliability coefficients. Other studies involving non-normals, however, do not demonstrate the same gains as an apparent result of cerebral dysfunction. Improvements, when noted, in neurologically-impaired groups are presumed to be due to improved cerebral functioning

and/or treatment effects rather than to test-retest practice effects assumed to operate in normal populations.

Since very few studies involve brain-injured subjects, and even then a restricted number of TBI survivors, this study will attempt to provide data that will establish the necessary foundation for the evaluation of test-retest practice effects and reliability of the WAIS-R in addition to monitoring recovery over a one year period of time.

From the literature review, it is apparent that many of the methodological shortcomings evident in previous studies must be avoided in this study. Employing a large number of subjects of a select group of TBI survivors will be paramount then in reducing contamination that might arise from a combination of other clinical populations and making decisions regarding statistical inferences with greater confidence. In addition, employing a control group of similarly-matched TBI subjects is imperative so as not to obscure subtle changes in the experimental group that might possibly occur when utilizing a non-impaired control group. Also evident from the literature review is the importance of screening subjects on pre-injury conditions on variables that might otherwise confound or compromise cognitive efficiency or recovery. For similar reasons a matched control group is also important in controlling the effects of many other injury and post-injury variables that might possibly bias outcome results.

The related literature has also revealed that in order to minimize test-retest practice effects in a control group, a reasonable test-retest interval must be chosen. A review of the literature has supported an interval of approximately one year during which practice effects would be minimized or considered marginal. In this manner, the improvements or gains demonstrated by the control group should reflect the extent of true recovery in intellectual functioning within a one year period of time. Scores in the experimental group which then exceed the control groups' scores within the same time period (assuming potentially confounding variables are held constant) would demonstrate increases in scores due to test-retest practice effects as a result of multiple test administrations. These gains would also reflect increases beyond those anticipated from natural recovery alone as observed in the control group.

Previous research has also indicated that recovery may involve greater improvements in Performance IQ than in Verbal IQ and the improvements in Verbal IQ may be completed more quickly when compared to Performance IQ. Based upon the review of the literature, groups of TBI subjects would be expected to show improvements, however, the increased familiarity of one group to a particular test over another may lead to greater obtained scores in the former group which in the past may have been misinterpreted and related strictly to recovery. In addition, variables thought to influence cognitive recovery are also considered to

influence practice effects. Therefore, control of those variables may minimize the effects these factors may have in influencing or biasing this study's results. A careful explanation of the methodological operations employed in this study with reference to these concerns are noted in the chapter to follow.

CHAPTER III

Methodology

Introduction

Due to the continuing need to monitor cognitive and intellectual recovery following traumatic-brain-injury, information regarding test-retest changes is essential to the accurate interpretation of observed performances. At the present time, such information is virtually absent. Without it, clinicians charting recovery from head injury will be unable to determine whether such changes in scores from test to retest are due to the recovery process, psychometric unreliability, or test-retest practice effects (Campbell, 1983).

While it is generally assumed that most surviving TBI subjects demonstrate some type of improvement early in their recovery, it is unclear to what extent changes measured over time are due to the recovery process and/or to the effects of repeat testing. Therefore, it is important to determine: 1) the role of test-retest practice effects on test scores in a population of recovering TBI survivors; 2) the stability or change in test scores obtained from test to retest; and 3) the test-retest reliability of the WAIS-R. As noted

earlier, this determination has been primarily logically, rather than statistically based.

This investigation is important then, since it will attempt to distinguish between the process of recovery and the effects of test-retest practice. The differentiation between recovery and the effects of multiple repetitions of the WAIS-R is a question of substantial clinical import. This is due to the pressure placed upon clinicians to continuously monitor recovery and to provide an accurate interpretation of obtained scores upon which medico-legal-rehabilitation determinations are frequently based. Improving the accuracy of the interpretation would, therefore, have important ramifications for professionals, families, and patients, alike.

Based upon a review of the literature, it is anticipated that test-retest practice effects will result, although perhaps not as profoundly as has been demonstrated in normal controls. Therefore, it is expected that if variables identified as potentially confounding are held constant, an experimental group of recovering TBI survivors who have taken the WAIS-R twice as often as a control group of recovering TBI survivors, will demonstrate significantly greater IQs and subtest scores, and that these greater scores will be due to test-retest practice effects.

It is also expected that both the experimental and control groups will demonstrate recovery within a one year period of time following their head injuries. However, this

recovery may not be without some form of test-retest practice effect. Differences between WAIS-R scores obtained at initial testing and scores obtained at one year post-injury will, under the assumption that potentially confounding variables have been identified and held constant, reflect a test-retest practice effect (if any) in addition to the patients' natural recovery. In this regard, it is anticipated that IQ and subtest scaled score means of the experimental group will be greater than the control group at one year post-injury due to the excessive exposure to, and familiarity with, the test instrument. In addition, it is expected that the pattern of recovery in WAIS-R IQ values of the experimental group will be curvilinear. That is, a period of rapid improvement will be followed by a decelerating rate of improvement suggesting a quadratic trend in the recovery pattern.

To test these hypotheses, archival data was retrieved on TBI patients who were evaluated on a longitudinal basis at approximately 2, 4, 8, and 12 months post-injury, respectively (see Appendix A). Test-retest gains or practice effects were then investigated by making comparisons of test scores between the experimental group and a control group of similar adult TBI survivors who were tested only twice with the WAIS-R at approximately 2 and 12 months post-injury, respectively (see Appendix A). The extent of recovery within each of the experimental and control groups over time

was also investigated in addition to determining test-retest correlations and changes in internal consistency.

In order to better understand the procedures employed in the present study, this chapter will address methodological issues. Included are the subject population, sample selection, data collection, characteristics of the sample, instrumentation, design, threats to validity, the hypotheses, data analyses, and limitations of the analyses.

Subject Population

All of the subjects involved in this study were either admitted to and/or received treatment from rehabilitation facilities in Michigan as a result of a traumatic-brain-injury. The primary site for subject recruitment was an 80 bed rehabilitation hospital which supported an Adult Brain Injury Program for treatment of both inpatients and outpatients. Secondary sites included: (1) a 300 bed acute care facility which also supported a rehabilitation unit for the assessment and treatment of adult head injuries; and (2) a residential, rehabilitation facility supplying treatment to recovering TBI survivors subsequent to acute care hospitalization.

Subjects were diagnosed as having had a traumatic-brain-injury on the basis of a neurological examination, presence of coma secondary to the head injury, neuroradiological findings (i.e., computerized axial tomography), or a combination of any of the findings noted above. The general

purpose of the rehabilitation at those facilities which provided subjects for this study was to restore the subject's cognitive, physical, and vocational potential to the maximum extent possible following impairment of these areas due to an accident or other form of trauma.

Sample Selection

As part of each patient's assessment and treatment, all patients were referred by a physician to a staff psychologist for neuropsychological testing on an ongoing basis. This service was performed to monitor and assess the rate and degree of recovery following head injury, and to make recommendations regarding treatment, guardianship, discharge and/or vocational planning. Since the WAIS-R was not published prior to 1981, it was determined that only those patients whose traumatic-brain-injuries occurred on or following 01/01/81, and who had only received testing with the WAIS-R, would be considered for inclusion in this study. Consistent with the WAIS-R manual (Wechsler, 1981) only those individuals who were 16 or older at the time of their evaluations were included for study.

Due to the inordinate amount of time and cost to conduct a study of this nature prospectively, a retrospective study investigating archival data appeared warranted. Files of patients who were either considered or designated as TBI survivors were reviewed proforma to determine which patients would become subjects for this study. Brain-injured

patients were assigned to one of two groups where applicable. Assignment to either the experimental or the control group was contingent upon the frequency of tests administered as well as the times post-injury the subjects were tested with the WAIS-R. Assignment to the experimental group required evidence of four test administrations of the WAIS-R at approximately 2, 4, 8, and 12 months post-injury, respectively. Criterion for assignment to the control group consisted of two test administrations with the WAIS-R at approximately 2 and 12 months post-injury, respectively.

In an effort to avoid methodological short-comings reported in the neuropsychological literature and to meet subsequent recommendations for research (e.g., Levin et al., 1982), subjects were excluded who evidenced either: (1) a prior and significant head injury; (2) a premorbid psychiatric history; (3) pre-injury epilepsy; (4) penetrating head injury (e.g., gunshot wound); (5) chronic alcoholism; (6) anoxia; (7) meningitis; (8) mental retardation or other congenital disability; or (9) other neurological impairment (i.e., multiple sclerosis). These variables were reported to result in potential confounds in both recovery and intellectual status over time.

Since random assignment of subjects to groups was not possible due to the retrospective nature of the study, attempts were made to ensure that the two groups were as comparable as possible in terms of demographic and injury-related variables. In addition, testing periods were

selected to fall within specified ranges so as to eliminate other confounds created by unequal test-retest intervals. Thus, subjects were also excluded if any of the following occurred: (1) Verbal and/or Performance IQs were either below 60 or exceeded 110 at the time of the first test (pretest) administration; (2) severity of injury, as determined by length of coma, exceeded 35 days; (3) the first test (pretest) administration occurred more than three and one-half months (105 days) post-injury; (4) the last test (posttest) administration occurred earlier than eight and one-half months post-injury and/or later than 18 months post-injury; and (5) no fewer than six months separated the first test (pretest) from the last test (posttest) administration. Therefore, the stated testing times of 2, 4, 8, and 12 months are approximations to the actual testing times and should be treated as such.

Data Collection

Subjects in both the experimental and control groups were administered the WAIS-R by psychologists trained in its administration and experienced in traumatic-brain-injury assessment and rehabilitation. As above, the experimental group consisted of subjects evaluated at four time periods: approximately 2, 4, 8, and 12 months post-injury. The control group consisted of subjects evaluated at two time periods: approximately 2 and 12 months post-injury, respectively.

Data obtained for this study were transferred to a Data Summary Form (see Appendix B). The Data Summary Form was designed to provide a format for the systematic collection of information germane to a subject's head injury and recovery. This form also allowed for the inclusion of information regarding pre-injury status, medical information, and demography on which to compare the two groups more closely in addition to the required psychometric data. As noted on the Data Summary form, variables were included that either empirically or logically were thought to confound test results and/or intellectual recovery. Therefore, an attempt was made to include as many of those variables as possible in order to identify and control for their influences.

While the majority of data collection occurred at the rehabilitation facilities mentioned previously (notably the primary rehabilitation hospital), data collection also required the inclusion of data from all other professional agents or agencies who assessed and/or treated each subject's intellectual status post-injury. Clinical information derived from the assessment and/or treatment of each respective subject while at the rehabilitation facilities were retrieved from the respective Psychology Department files. This was also true for both the inpatient and outpatient files found in the respective Medical Record Departments. Clinical information derived from the assessment and treatment of each subject by an agent or agency

other than the rehabilitation facilities noted above was also retrieved from either the inpatient and/or outpatient files in the respective Medical Record Departments where available.

However, clinical information derived from the assessment and/or treatment of a subject by another agent or agency other than the rehabilitation facilities noted above and not available in the respective Medical Record Departments required additional procedures. These procedures involved a statement of informed consent as well as a signed authorization to release information for the purpose of this study. Unless legal guardianship had been determined for the subject, he or she was considered appropriate to make decisions regarding consent and release of that requested information. The procedures which follow in obtaining, using, and treating archival data in a retrospective study of this nature were approved by the Michigan State University Committee on Research Involving Human Subjects (UCRIHS) (see Appendix C).

In instances where informed consent was required, a letter describing the purpose of this study and need for data collection was sent to either the subject or his/her guardian's most recent address (see Appendix D) in addition to a release of information form delineating the data requested for this study (see Appendix E). A return envelope was also enclosed to enable the subject and/or their guardian to return the forms which, in turn, were sent

to the respective agent or agencies. Failure to return the respective form by either the subject or guardian was followed by a phone call to either the subject or the guardian to insure receipt of the materials, to answer questions of concern, and/or to ask permission to send a second letter if the first was lost. Failure to return the consent form following this procedure was respected as a refusal to participate. In total, requests for participation were sent to 23 potential subjects of which 15 responded affirmatively.

Information obtained from the files of the respective Psychology and/or Medical Record Departments, or received from an agent or agency other than the rehabilitation facilities noted above were recorded on the Data Summary Form accordingly. All data obtained from these procedures were coded and recorded by serial number to protect and insure patient confidentiality and anonymity. The use of patient names was not relevant to this study except in requesting information from other agencies vis a vis informed consent procedures.

Transformation of data (e.g., age at time of injury, estimated preinjury IQ) was computed by hand and entered on the Data Summary Form. Raw score test data from the WAIS-R was converted to scaled scores where necessary and recorded. When required, calculations of IQs were also completed according to the WAIS-R manual and entered accordingly on the Data Summary Form. Procedures also included the

prorating of IQ scores based on missing items from the WAIS-R as permitted by the WAIS-R manual.

Characteristics of the Sample

Roughly 1000 (all) files from the three rehabilitation facilities were reviewed. Eighty-five TBI subjects were judged to meet inclusion criteria for this study. Seventy-three subjects were identified from the primary rehabilitation hospital. The remaining 12 subjects were divided evenly among the two, secondary rehabilitation sites. Sixty-five of the subjects were male and 20 were female. Ages ranged from 16 to 76 with a mean age of 27.6 years. Average education level obtained by the subjects was 11.8 years with a mean grade point average (GPA) of 2.06.

Severity of head injury, as evidenced by length of coma, averaged 4.6 days suggesting a severe, traumatic-brain-injury overall. Other severity of injury indices, including the Los Amigos Cognitive Functioning Levels (Malkumus, 1974) and the Disability Rating Scale (Rappaport, Hall, Hopkins, Belleza & Cope, 1982), were 5.2 and 10.6, respectively. These scores, on average, suggested that while the subjects were alert at the time of their rehabilitation admission, many may have demonstrated disorientation, confusion, verbal inappropriateness, agitation in response to external demands, and physical impairments suggestive of moderately-severe disabilities.

Acute care hospitalization immediately following trauma averaged 29.5 days with subsequent rehabilitation hospitalization requiring an additional 56.5 days on average. A mean of 262.6 days of rehabilitative therapy was provided to the subjects from the time of their injuries including acute care hospitalization. Subjects who received psychotropic medications required an average length of 90.6 days to control agitation during the early phase of recovery. Anticonvulsant medications were also employed an average of 252.0 days to prevent or control post-traumatic seizures in at-risk subjects.

On average, subjects were initially tested at 58.2 days post-injury and demonstrated Verbal, Performance, and Full Scale IQ means of 84.0, 78.2, and 80.2, respectively. Based on the mean Full Scale IQ, subjects were functioning, in general, in the low average range of intelligence. Final testing with the WAIS-R took place an average of 374.0 days post-injury; approximately 10 1/2 months following the first evaluation with the WAIS-R.

Instrumentation

Instrumentation consisted of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Wechsler, 1981). As noted previously, the WAIS-R is composed of 11 subtests; six verbal and five nonverbal. The verbal and nonverbal sections may be administered separately or together to yield a Verbal, Performance, and Full Scale IQ.

The WAIS-R was standardized on 1,880 white and nonwhite Americans, equally divided with respect to gender and selected to be representative of the U.S. late adolescent and adult populations during the 1970s (Sattler, 1988). Accordingly, the demographic characteristics used to obtain a stratified sample were age (16-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-69, and 70-74), sex, race (white = 1,664, black = 192, and Asians plus native Americans = 24), geographic region (Northeast, North Central, South, and West), education, and urban-rural residence (Sattler, 1988).

The WAIS-R is intended for use with individuals sixteen years of age and older. Normative data is provided through age seventy-four, although individuals older than seventy-four may be tested using the norms for ages seventy to seventy-four. WAIS-R examiners need to be thoroughly trained in the use of individually administered intelligence scales. This also requires that the examiners need to adhere carefully to the specific directions given in the test manual for each subject. In addition, specific rules regarding starting, discontinuance, probing, and repetition of items must be addressed (Sattler, 1988).

Scoring WAIS-R responses requires careful study of the scoring criteria, scoring guidelines, and scoring examples in the WAIS-R manual (Sattler, 1988). Raw scores for each subtest were converted to scaled score equivalents ($\bar{x} = 10$, $SD = 3$) for each of the nine age groups in the standardization group. In this way, subject scores can be directly

compared to that of age peers. However, for this study, the scaled scores for each of the eleven subtests in the battery were based on a general reference group consisting of five hundred normal subjects used in the standardization sample between the ages of twenty and thirty-four. This is consistent with the recommendation of Shatz (1981) who suggested that an examination of scaled scores is preferable to age-weighted scores since increases in scores may be due to an aging artifact.

IQ scores are derived by converting raw scores to scaled scores based on the general reference group noted above. The scaled scores are summed, and the sum and the subject's age are used to find IQs for the Verbal, Performance, and Full Scale sections in the WAIS-R manual. IQs for each section employ a mean of 100 and a standard deviation of 15. Age-corrected subtest scaled scores are not used to calculate IQs, but are used only when making subtest interpretations and comparisons. The WAIS-R manual also makes provisions for prorating IQ values when a subtest is omitted from either the Verbal and/or the Performance sections of the WAIS-R.

The WAIS-R provides highly reliable IQs (Sattler, 1988) as reported previously in Chapter Two. The standard errors of measurement (SE_m) in IQ points based on the average of nine age groups, are 2.53 for Full Scale IQ, 2.74 for the Verbal IQ, and 4.14 for the Performance IQ (Sattler, 1988).

Evidence of construct validity has been provided by: (1) the level and pattern of intercorrelations between WAIS-R subtests and scales; (2) the observation that scores on the WAIS-R are distributed in a manner consistent with theoretical explanations; and (3) the results of factor analysis applied to the intercorrelations of the 11 subtests (Sattler, 1988). A review of studies by Sattler (1988) also suggested satisfactory concurrent validity with other intelligence tests, picture vocabulary tests, achievement measures, and years of education.

Regarding the WAIS-R's validity, Wechsler (1981) stated,

"A body of evidence both rational and empirical, attests the validity of the Wechsler Adult Scale as a measure of global intelligence. Inasmuch as the WAIS-R measures the same abilities as the 1955 WAIS and its predecessor, the Wechsler-Bellevue Intelligence Scale, and overlaps considerably in test content with these earlier forms of the Wechsler Adult Scale, studies with the WAIS and W-B can be considered relevant to this revision." (p.49)

Design

The research design for this study was conceived as having the form of an incomplete split-plot, repeated measures design as seen in Figure 3.1.

C

E

I

c

c

s

t

an

F

su

st

th

3.

ne,

of

inj

Fig.

		Time (months)			
		2	4	8	12
Groups	Experimental	X	X	X	X
	Control	X			X

Figure 3.1. Research Design.

In this study, time (or repeated measures) was partially-crossed with the grouping variable (experimental or control). Specifically, groups represented a between-subjects factor and time a within-subjects factor (i.e., time at pretest = 2 months post-injury; 4 months; 8 months; and posttest = 12 months post-injury). IQ values for the Full, Verbal, and Performance scales, as well as all WAIS-R subtest scaled scores, served as dependent variables.

Using Campbell and Stanley's (1963) configuration, this study employed a quasi-experimental design referred to as the "non-equivalent control" (p. 47) design seen in Figure 3.2. The X in Figure 3.2 represents exposure to the experimental variable; namely, the two additional administrations of the WAIS-R at approximately four and eight months post-injury, respectively.

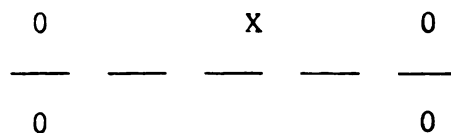


Figure 3.2. Non-equivalent Control Design.

The Os in Figure 3.2 represent pre- and posttests with the WAIS-R at approximately 2 and 12 months post-injury, respectively. The dashed line in Figure 3.2 using Campbell and Stanley's (1963) convention, represents comparison groups not equated by random assignment.

Threats to Validity

The lack of random assignment means that the experimental and control groups cannot be assumed to be equivalent at the pretest (Campbell & Stanley, 1963), i.e., the effects of potentially confounding variables cannot be expected to be evenly distributed across the two groups. Such an imbalance threatens the internal validity of this study and allows for (at best) weak causal statements to be made.

Campbell and Stanley (1963) defined internal validity by asking whether an experimental intervention results in a specific effect. If the effect can be attributed solely to the intervention the study is presumed to have internal validity. This is in contrast to extraneous variables which, if not controlled by the experimental design, might produce effects confounded with the effect of the experimental stimulus. These extraneous variables have been categorized into eight classes by Campbell and Stanley (1963) including: (1) history; (2) maturation; (3) testing; (4) instrumentation; (5) statistical regression; (6) selection of subjects; (7) experimental mortality; and (8) selection-maturation interactions.

Campbell and Stanley (1963) defined external validity as the degree to which the experimental effect can be generalized (e.g., to various populations, settings, treatment variables). These authors' reported threats to external validity or representativeness that include the following: (1) reactive or interaction effects of testing; (2) interaction effects of selection biases and the experimental variable; (3) reactive effects of experimental arrangements; and (4) multiple-treatment interference.

The lack of random assignment implies that the two groups may differ prior to treatment intervention. However, the availability of pretreatment information on potentially confounding variables means that it is possible to empirically determine whether groups are equal on the potentially confounding variables. The more similar the two groups on pretest variables (i.e., psychometric testing at two months post-injury), the more effective the control group becomes (Campbell & Stanley, 1963). This is particularly true as it regards intellectual performance in this study. Campbell and Stanley (1963) also stated that for purposes of internal validity, the status of this design controls for the main effects of history, maturation, testing, instrumentation, selection, and mortality.

One type of information is the pre-injury characteristics of the subjects who are the victims of head injury and the variables that may have contributed to placing subjects at greater risk to TBI. A second type is the physiological

profile which characterizes TBI survivors not only by the type of injuries, but also by the treatment and recovery which the subjects experienced. In addition, the retrospective nature of the data collection and the detailed nature of current medical records maximizes a researcher's control over several nonmedical-type variables. Thus, an attempt was made to make the two groups as similar as possible through their recruitment and demography.

This notwithstanding, deliberate attempts were made to further limit other threats to internal validity as noted by the examples below. While intrasession history could not be controlled for in all situations, an attempt was made to delineate similarities between the two groups. One way of doing this was by determining the length of both the acute and rehabilitation hospitalizations as well as the length of continuous therapy to explain some of the events which occurred between measurements. The effects of maturation were also controlled for by attempting to restrict some variables such as the age of individuals who were injured and subsequently tested. An attempt was also made to control the time span between testings, and therefore maturation within subjects. Pretesting with the WAIS-R using standardized techniques reported earlier was expected to restrict the effects of testing as well as instrumentation. It was assumed that all examiners followed appropriate administrative and scoring procedures. Selection of

subjects also was made partially on physiological determinants (e.g., length of coma) which yielded a more homogeneous group of both experimental and control subjects.

Campbell and Stanley (1963) stated that one general concern regarding the internal validity of the design (i.e., the assumption that the independent variable is responsible for variation in the dependent variable) was the problem of regression-to-the-mean. However, if both groups demonstrate similar deviations from the mean at pretest (two months post-injury), then the regression-to-the-mean phenomenon should be similar for both groups. This would help to rule out regression-to-the-mean as a threat to internal validity.

If the means of the two groups are substantially different at pretest, then Campbell and Stanley (1963) state that the occurrence of unwanted regression effects is insured and it becomes predictable that the two groups will differ on the posttest altogether independent of any effects of X. In that case, statistical control of the variation associated with these unwanted effects (e.g., analysis of covariance) should be considered. However, such a statistical analysis is unlikely to remove all of the undesirable effects associated with pretreatment differences among groups. In this study the process of selecting subjects was expected to produce groups which were approximately equivalent on a host of potentially confounding variables.

Despite these attempts to improve internal validity, limitations of this study still included the failure to

control for or assess the degree to which other variables affected recovery and practice effects. Such factors included, for example, indices of the subjects' emotional status at the time of testing, the time of day testing took place, as well as the subjects' attitudes toward testing and/or the test examiner. Pre- and posttests of the subjects' memory abilities and learning capabilities were also important variables not taken into consideration as it regards outcome from recovery or repeat testing. In addition, concerns regarding the quality and quantity of treatment given to subjects between testing was not accounted for.

Campbell and Stanley (1963) also emphasized that interaction(s) may exist between selection factors and other variables such as history, maturation and testing. The selection of groups, however, was made on a similar basis in that minimally, all individuals sustained a severe traumatic-brain-injury and were referred for intellectual testing to chart recovery. As noted previously, other attempts were made to make the groups homogeneous. While recovery time post-injury, maturation, history, and pretesting may be similar, an effort to explain a pretest-posttest gain specific to the experimental group may need to hypothesize an interaction between these variables and the selection differences that distinguish the experimental from the control group.

As suggested by Campbell and Stanley (1963), an interaction between selection and maturation may be tenable

even when groups are equal on a pretest. If, in a group of recovering TBIs, individuals in the experimental group experienced more treatment during their rehabilitation or showed greater spontaneous recovery than individuals in the control group, such an interaction may exist. The reasons for experiencing more treatment (testing) may be due to chance, patient availability, funding sources, or severity of injury necessitating greater need for monitoring recovery to name but a few. In addition, testing may have been used to prescribe treatment such that greater areas of deficit received more cognitive retraining which might have further increased obtained scores on retest. However, the design employed in the present study does not permit interactions of this nature to be investigated statistically. Thus, absolute statements about test-retest practice effects are not possible; rather, only weak causal statements are permissible.

Threats to external validity (i.e., the generalizability of research findings to and across populations of subjects and settings) are likewise exacerbated by the absence of random assignment of patients to both groups. This is because internal validity is a prerequisite for external validity. It is also unclear to what extent the interaction between testing and treatment threatens external validity. The nature of the proposed study, however, suggests that pretesting may affect subsequent testing (or treatment in this case) as a result of prior exposure resulting in

test-retest effects. However, this interaction would seem to be generalizable to that proportion of the universe (e.g., neuropsychological testing, rehabilitation, and psycho-medico-legal determinations) where repeated measures (assessments) are common.

The interaction of selection and treatment may also affect external validity to the extent that there was difficulty in obtaining subjects meeting the parameters required for inclusion in the population. This may be due to other factors which did not result in repeat testing such as the number of physician referrals, overall patient improvement, or funding for testing as examples. Also, attrition of subjects over time may be a confounding variable. A number of authors (e.g., Brooks et al., 1984) have reported attrition rates of 30% in similar studies which may have resulted in fewer subjects who were available. While all subjects were recruited from rehabilitation facilities in Michigan, it is unclear how representative this sample is of other TBI survivors or rehabilitation facilities elsewhere in the United States. It should be emphasized, however, that the rehabilitation facilities from which this sample was recruited were ones which espoused state-of-the-art technology and treatment. However, the nature of the subject population suggests that the results for the sample used in the study can be generalized.

Reactivity to testing and the testing arrangements are also a possible confound to generalizability. In cases in

which subjects were tested as part of medical treatment repeat intellectual assessment may have been perceived as part of the role patients assumed while hospitalized or under the care of their physician, and hence generalizability was increased. If, on the other hand, patients looked forward to future repeat testings to try to improve their scores, or resented having to return for more tests, generalizability would be compromised. Generalizability is also likely to be compromised due to multiple-treatment interference. That is, the multiple testing (or treatment in this case) applied to the same subjects will contaminate results since the effects of such treatments are not erasable (Campbell & Stanley, 1963).

Hypotheses

The primary hypotheses tested in this study were as follows:

Hypothesis one.

$$H_0: u_e - u_c = 0$$

$$H_1: u_e - u_c > 0$$

Hypothesis two.

$$H_0: u_e - u_c = 0$$

$$H_1: u_e - u_c > 0$$

Hypothesis three.

$$H_0: u_e - u_c = 0$$

$$H_1: u_e - u_c > 0$$

Hypothesis four.

$$H_0: u_e - u_c = 0$$

$$H_1: u_e - u_c > 0$$

Hypothesis five.

$$H_0: u_e - u_c = 0$$

$$H_1: u_e - u_c > 0$$

Secondary hypotheses include the following:

Hypothesis one.

$$H_0 = p_1 - p_2 = 0$$

$$H_1 = p_1 - p_2 \neq 0$$

Hypothesis two.

$$H_0 = \psi \text{ quadratic} = 0$$

$$H_1 = \psi \text{ quadratic} > 0$$

Data Analyses

Before formal hypothesis testing, descriptive statistics were used to summarize demographic characteristics including pre-injury, injury, and post-injury variables. In addition, preliminary analyses were completed to test for differences

between the experimental and control groups on potentially confounding variables that may have influenced the results of this study. Preliminary analyses of these variables employed two-sample, t-tests for independent groups (two-tailed; $\alpha = .05$).

In order to test whether there were any differences between the experimental and control group at the time of the pretest (two months post-injury), two-sample, t-tests for independent groups were used to compare performances on the Verbal, Performance and Full Scale IQ values (two-tailed, $\alpha = .05$). All t-tests in the study utilized separate variance estimates by way of the Welch-Aspin Test (Marascuilo & Serlin, 1988). This test is appropriate when the statistical assumption of homogeneity of variances is questionable. If mean differences were significant, then a strategy to remove extraneous variation due to the pretest such as analysis of covariance was to be used. Fortunately, this last option was not necessary. Similar analyses utilizing two-sample, t-tests for independent groups were also completed on the WAIS-R subtest values between the control and the experimental group.

Because of the a priori nature of the primary research questions, all of the major statistical analyses involving the primary hypotheses were planned and all tests were directional ($\alpha = .05$). This is important for the following reasons: (1) When properly guided by sound theory or previous research, directional tests are more appropriate

than non-directional tests since a review of research relevant to the said proposal should lead to a prediction for how the results should turn out (Glass & Hopkins, 1984); and (2) other things being equal, the statistical test associated with a directional hypothesis will have greater power for rejecting a false H_0 than a two-tailed test. This increases the likelihood of detecting a treatment effect. The planned analyses employed t-statistics, and consisted of both between- and within-subject tests. The analysis associated with a particular dependent variable was also considered to be one experiment, since the various dependent variables were to be analyzed singly.

Within-subject effects as postulated in hypotheses one and two were examined by using matched-pair t-tests. In this regard, the changes observed between the pretest (two months post-injury) and posttest (12 months post-injury) were tested separately in the control and experimental groups for both the IQ and subtest scaled score values.

Between-subject effects as postulated by hypotheses three and four employed two-sample, t-tests for independent groups. This statistical test was used to test for differences between the experimental and control group test scores at posttest (12 months post-injury) for all IQ and subtest scaled score values. A similar planned comparison was used for hypothesis five which tested the significance of an interaction; namely, the presence of test-retest gains thought to exist due to practice effects. In this regard

the differences between the pre- and posttest scores of the experimental group were compared with the pretest-posttest differences in the control group for all IQ and subtest scaled score values.

Secondary hypotheses were not directional. In order to test secondary hypothesis one, z-tests were used to test differences among reliability coefficients. Specifically, Pearson correlations of the pre- and posttest administrations were computed separately for the experimental and control groups and the groups were compared on these correlations using Fischer's two-sample, z-test involving r to Z transformed correlations (Glass & Hopkins, 1984). This test provided information on whether the test-retest correlations from pretest to posttest were the same for the experimental and control groups.

In order to test secondary hypothesis two, a test of linearity for changes in the IQ values over four test times in the experimental group were made. Since testing and resulting IQ values were obtained at unequal intervals of time trend coefficients were derived by hand following Kirk (1982). Estimated linear and quadratic contrasts were then tested for significance using the F-test.

Secondary hypothesis three utilized a gross estimate of the WAIS-R'S internal consistency by calculating the median Pearson correlations among all possible subtest correlations over all four test administrations. Descriptive information

was also provided regarding the variance for both the experimental and control groups over time.

Limitations of the Analyses

Statistically, the greatest limitation to the data analyses described previously is the use of multiple t-tests. Glass and Hopkins (1984) stated that whenever more than one t-test is made, the probability of one or more type-I errors (i.e., rejecting the null hypothesis when it is true) increases. That is, alpha, the probability of making type-I error becomes quite large as the number of groups increases. The dependency among the t-tests also makes this problem more complex inasmuch as it is impossible to determine the actual value of alpha for several different, non-independent t-tests (Glass & Hopkins, 1984). Thus, falsely significant results are likely and hence confound the strength of statements relative to this study's outcome.

CHAPTER IV

Analysis of Results

Introduction

Analysis of the data generated by this study is presented in this chapter. Four areas of interest are addressed. The first section concerns a summary of data collected on discrete demographic variables including pre-injury, injury, and post-injury characteristics in both the experimental and control groups. Comparisons between the two groups were made on those variables thought to have a potentially significant effect on both neuropsychological test performance and/or recovery following TBI. The second section consists of a description of the sample on both demographic and experimental continuous variables also thought to impact the course of recovery and psychometric results. A comparison of these variables between the two groups at pretest further evaluated the ability of the control group to hold constant confounding variables. The third and fourth sections formally test the primary and secondary hypotheses, respectively.

Demographic Data

A total of 85 subjects were appropriate for inclusion in this study; 47 of which were in the experimental group while the remaining 38 comprised the control group. Recall that the validity of the statistical results are predicated on the ability to identify and hold constant, potentially confounding variables. Thus, wherever possible, attempts were made to collect as much data as was available on all subjects. However, the retrospective nature of this study created a situation in which complete information on some subjects was not available in the existing medical records. Therefore, data analysis did not always encompass the complete sample of 85 subjects, and the number of subjects in the analyses vary. The number (or percentage) of subjects used in each analysis is noted in the tables to follow.

In keeping with the need to hold constant confounding influences, the groups were compared on variables thought to have a potentially significant impact on neuropsychological test performance and recovery. No statistically significant differences between the groups were found. Table 4.1 presents information derived from the collection of discrete variables on the sample of 85 subjects.

Consistent with the national population of people with head injury (Rimel & Jane, 1983), the sample was largely male (76.5%). The sample was also predominately Caucasian (94.1%). Right-handedness was prevalent (85.9%) with

Table 4.1

Comparison of Experimental and Control Group Subjects
on Pre-Injury Demographic Variables

	Percent		
	Experimental	Control	Overall
Sex			
Male	78.7	73.7	76.5
Female	21.3	26.3	23.5
Race			
Caucasian	100.0	86.8	94.1
Black	-	7.9	3.5
Hispanic	-	2.6	1.2
Other	-	2.6	1.2
Handedness			
Right	87.2	84.2	85.9
Left	12.8	5.3	9.4
Ambidextrous	-	2.6	1.2
(Missing)		7.9	3.5
Marital Status			
Single	61.7	63.2	62.4
Married	31.9	26.3	29.4
Divorced	6.4	10.5	8.2
Academic Obtainment			
Pre-High School	6.4	23.7	14.1
High School Drop Out	12.8	13.2	12.9
G.E.D.	2.1	10.5	5.9
High School Degree	72.3	44.7	60.0
Associates Degree	2.1	-	1.2
Bachelors Degree	4.3	5.3	4.7
(Missing)		2.6	1.2
Special Education			
Classification			
Learning Disabled	4.3	5.3	4.7
Emotionally Impaired	2.1	-	1.2
Not Classified	91.5	81.6	87.1
(Missing)	2.1	13.2	7.1

Table 4.1 cont'd.

Reported Alcohol Abuse			
Yes	53.2	31.6	43.5
No	40.4	63.2	50.6
(Missing)	6.4	5.3	5.9
Reported Drug Abuse			
Yes	17.0	21.1	18.8
No	74.5	73.7	74.1
(Missing)	8.5	5.3	7.1
Occupational Endeavors			
Professional - Technical	4.3	5.3	4.7
Farmer - Farm Manager	2.1	-	1.2
Managers - Office	8.5	2.6	5.9
Clerical - Sales	-	2.6	1.2
Craftsman - Foreman	4.3	5.3	4.7
Operators	8.5	5.3	7.1
Service Work	17.0	7.9	12.9
Laborers	27.7	23.7	25.9
Housekeeping	4.3	2.6	3.5
High School Student	10.6	23.7	16.5
College Student	6.4	7.9	7.1
Disabled	-	2.6	1.2
Unemployed	2.1	7.9	4.7
Retired	4.3	2.6	3.5

Note. Overall percentages were calculated on a total N = 85. Experimental and control percentages were calculated on ns of 47 and 38, respectively.

left-handed individuals and ambidextrals representing 10.6% of the sample. Ambidextrous individuals (1.2%) were represented in the control group only. Handedness was not known because of incomplete records on 3.5% of the sample. A majority of subjects (62.4%) were single with the remainder either married (29.4%) or divorced (8.2%). These figures are consistent with Rimel and Jane (1983) who also reported that the incidence of injuries involving singles is 20 percent higher than the population base as a whole.

In general, a large percentage of subjects (60%) had obtained at least high school degrees while approximately 14% were still in high school at the time of their injuries. An additional 5.9% of the sample had either completed an Associate or a Bachelor's degree. Nearly 13% of the total sample had dropped out of high school while another 5.9% had completed G.E.D.s. Educational obtainment could not be determined in one of the cases as a result of missing information. 5.9% of the sample were served by special education programs as identified and treated by their schools' respective educational planning committees. Reportedly, 4.7% of these individuals were classified as learning disabled with a smaller percentage (1.2) identified as emotionally impaired. A little more than seven percent of the total sample had missing information regarding special education services with the proportion being higher in the control group than the experimental group.

Another set of variables that may potentially compromise recovery and psychometric test results were histories of either alcohol or substance abuse. Nearly half of the sample reportedly abused alcohol while approximately 19% of the sample reportedly abused other substances (e.g., marijuana, cocaine).

Table 4.1 also reveals that 23.6 percent of all subjects were in some form of educational endeavor at the time of their injuries. A large percentage of subjects (54.1%) were involved in occupations such as craftsmen, operators, general laborers, service workers, and housekeepers. Thirteen percent of the sample was involved in professional-technical, farming, managerial, or sales work. The remaining 9.4% were either disabled, unemployed or retired.

With regard to the cause of the injuries sustained, Table 4.2 indicates that 70.6% of the sample was involved in motor-vehicle accidents while an additional 18.9% was involved in either motorcycle or car-pedestrian accidents. This data is consistent with Rimel and Jane (1983) who also reported that more than half of all head injuries are related to traffic accidents. The remaining subjects (10.6%) either sustained falls or were victims of assault. As a result of the cause of their accidents, an overwhelming number of subjects (71.8%) qualified for no-fault insurance coverage for rehabilitation care subsequent to their accidents. This, in turn, suggested that a large number of subjects received potentially unlimited finances with which

Table 4.2

Comparison of Experimental and Control Group Subjects
on Injury and Post-Injury Variables

	Percent		
	Experimental	Control	Overall
Source of Trauma			
Motor Vehicle Accident	78.7	60.5	70.6
Motorcycle	4.3	10.5	7.1
Pedestrian-Car	6.4	18.4	11.8
Fall	6.4	7.9	7.1
Assault	4.3	2.6	3.5
Insurance Coverage			
Auto No-Fault	76.6	65.8	71.8
Private	8.5	18.4	12.9
Government (Medicaid)	12.8	15.8	14.1
(Missing)	2.1	-	1.2
Documented Use of Alcohol at Time of Injury			
Yes	36.2	26.3	31.8
No	51.1	44.7	48.2
(Missing)	12.8	28.9	20.0
Type of Injury			
No Skull Fracture	63.8	68.4	65.9
Non-Penetrating Skull Fracture	31.9	26.3	29.4
Penetrating Skull Fracture	4.3	-	2.4
(Missing)		5.3	2.4
Medication Induced Coma			
Yes	2.1	7.9	4.7
No	97.9	86.8	92.9
(Missing)		5.3	2.4

Table 4.2 cont'd.

Intracranial Pressure Monitoring			
Yes	23.4	31.6	27.1
No	76.6	55.3	67.1
Unknown	-	7.9	3.5
(Missing)		5.3	2.4
Neurosurgical Intervention			
Yes	14.9	26.3	20.0
No	85.1	65.8	76.5
(Missing)		7.9	3.5
Initial CT Scan Results			
Not Completed	2.1	5.3	3.5
Right-Sided Involvement	21.3	18.4	20.0
Left-Sided Involvement	14.9	21.1	17.6
Bilateral Involvement	17.0	21.1	18.8
Intraventricular Involvement	6.4	7.9	7.1
Normal	38.3	18.4	29.4
(Missing)		7.9	3.5
Side of Weakness or Hemiparesis			
None	38.3	34.2	36.5
Right	34.0	26.3	30.6
Left	17.0	13.2	15.3
Bilateral	8.5	5.3	7.1
Unknown	2.1	15.8	8.2
(Missing)		5.3	2.4
Post-Injury Seizure Activity			
Yes	10.6	15.8	12.9
No	89.4	78.9	84.7
Unknown		2.6	1.2
(Missing)		2.6	1.2
Use of Anticonvulsant Medication			
Yes	46.8	52.6	49.4
No	53.2	42.1	48.2
(Missing)	-	5.3	2.4

Use of Psychotropic
Medication

Yes	19.1	7.9	14.1
No	80.9	84.2	82.4
(Missing)		7.9	3.5

to fund rehabilitation treatment services. The balance of the subjects were supported by either private (12.9%) or governmental (14.1%) insurance programs. Insurance coverage could not be determined in one case. Of those subjects for whom records were available (20% missing), over 30% of all injuries involved alcohol. This proportion of alcohol-related injuries is less than figures reported by Rimel and Jane (1983), but consistent with Adams and Putnam (1989) who reported 26% of their subjects were under the influence of a non-prescribed substance at the time of their accident. These reports, nevertheless, suggest the need for further driver and alcohol education to the general public as it regards drinking and driving.

A large proportion of the subjects (65.9%) sustained closed head injuries without skull fracture. Approximately 30% of the subjects sustained non-penetrating skull fractures in addition to their head injuries. These figures are consistent with information provided by Rimel and Jane (1983). Approximately two and one-half percent of the sample could not be confirmed as having had a skull fracture.

According to medical records, only a small percentage (4.7%) of these individuals required drug-induced comas. Approximately 27% of the subjects did require some form of intracranial pressure monitoring to measure brain tissue swelling. Neurosurgical intervention such as the removal of a hematoma was required in approximately 20% of all cases, and is consistent with data reported by Rimel and Jane (1983). When CT scans of the brain were completed, nearly 30% of the findings were normal. This was in contrast to those on whom right (20%), left (17.6%), bilateral (18.8%) and intraventricular (7.1%) damage was identifiable. CT scans were either not done or were missing on 7.0% of the sample. Of those subjects on whom CT scan records were available, no mass effect was noted.

As a result of the head injuries incurred, 53% of the sample demonstrated some form of hemiparesis or motor weakness. Information about motor functioning was missing or could not be identified for 10.6% of the cases. Additionally, 12.9% of the subjects demonstrated seizure activity following their head injuries. Seizure data were missing for 1.2% of the total group, and another 1.2% could not be confirmed. Furthermore, 49.4% of all subjects for whom information was recorded, required the administration of an anticonvulsant medication such as Dilantin as either intervention for or prophylaxis of seizure activity. In

addition, 14.1% of the reported sample required some form of psychotropic medication such as Valium or Haldol principally during the acute care hospitalization as treatment for agitation and/or confusion.

Preliminary Data

A key goal of the subject selection was to generate a sample that was homogeneous with respect to various potentially confounding variables. However, the lack of random assignment of subjects to treatments makes it likely that the groups may differ on variables not being investigated and weakens the internal validity of the study. The presence of potentially confounding variables thought to impact test results and recovery following TBI, however, was assessed for the experimental and the control groups. To test the equivalence between the two groups, two-sample, t-tests for independent groups were employed ($\alpha = .05$). Nonsignificant results suggest, but do not ensure, that the groups are equivalent on that variable.

Comparisons between the experimental and the control group on important demographic variables are presented in Table 4.3. Mean ages of the subjects at the time of TBI were not significantly different between the experimental ($\bar{x} = 29.17$ years) and the control groups ($\bar{x} = 25.84$ years). In addition, pre-injury variables such as years of education, cumulative grade point average, and IQ scores were not significantly different between the two groups. Similarly,

TABLE 4.3

Comparison of Demographic Variables Between Experimental
and Control Groups

	EXPERIMENTAL				CONTROL				<i>p</i>
	<u>M</u>	<u>±</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>±</u>	<u>SD</u>	<u>n</u>	
Age at Injury	29.17	±	13.87	(47)	25.84	±	12.75	(38)	.253
Total Education	11.96	±	1.88	(47)	11.67	±	1.65	(37)	.451
Most Recent G.P.A. ^a	2.03	±	0.61	(40)	2.12	±	0.74	(26)	.630
Class Rank ^b	33.91	±	17.32	(23)	35.80	±	25.66	(10)	.835
Pre-Injury IQ ^c	101.05	±	10.96	(17)	100.33	±	19.14	(3)	.954
Estimated FSIQ ^d	102.47	±	3.95	(47)	101.85	±	3.46	(37)	.451
Length of Coma	4.86	±	6.56	(47)	4.32	±	5.50	(36)	.681
Length of Acute Care Hospitalization	31.68	±	16.82	(47)	26.66	±	19.42	(36)	.221
Time From Injury to Rehabilitation Hospitalization	32.21	±	17.90	(47)	30.55	±	19.91	(29)	.715
C.F.L. ^e Time of Rehabilitation Hospitalization	5.35	±	0.90	(47)	5.16	±	1.08	(27)	.458
D.R.S. ^f at Time of Rehabilitation Hospitalization	10.44	±	3.35	(45)	11.28	±	3.36	(14)	.422
Length of Rehabilitation Hospitalization	59.08	±	33.06	(47)	53.05	±	42.11	(34)	.490
Length of Continuous Therapy	275.19	±	113.51	(47)	246.78	±	127.38	(37)	.291

Table 4.3 cont'd.

Length of Anticon- vulsant Medications	248.00	<u>+</u> 125.59	(22)	257.05	<u>+</u> 152.80	(18)	.841
Length of Psychiatric Medications	53.11	<u>+</u> 82.76	(9)	203.33	<u>+</u> 233.01	(3)	.380

Note. Two-tailed probabilities were determined for all comparisons. Age at injury and total education are in years. All lengths of coma, hospitalizations, therapy and medications are in days. M = mean; SD = standard deviation; n = sample size; p = probability level.

^aGrade point average obtained or derived from available high school transcripts.

^bRetrieved from academic transcripts, the pre-injury IQ was the score determined by psychometric testing regardless of the instrument administered. ^cFormula equation

suggested by Karzmark, Heaton, Grant and Matthews (1985). ^dRancho Los Amigos Cognitive

Functioning Levels (Malkumus, 1974). ^eRappaport Disability Rating Scale (Rappaport, Hall, Hopkins, Belleza, & Cope, 1982).

post-injury variables including hospitalization length and treatment duration were the same statistically for both groups. In fact, a review of Table 4.3 suggests that none of the continuous, demographic variables that were identified and examined as potentially confounding to both test results and recovery were statistically different between the two groups using separate variance estimates. This suggests (but does not guarantee) that the two groups were equivalent at the beginning of the study and implies greater confidence in the comparisons between the control and experimental groups.

It is important to note, however, that while the means of the variables are similar, the variances (i.e., standard deviations) for some variables appear to differ (e.g., class rank, length of psychiatric medications). Thus, statements of equivalence refer to means and not to variances.

In this study, the larger variance was often associated with the smaller control group. This suggests several possibilities. First, the control group may not be as homogeneous with respect to variability as the experimental group despite nonsignificant findings using t-tests. Second, the lack of similarity (based on variance estimates) between the two groups suggests that the degree to which the control group can hold constant certain confounding variables is lessened by its greater variability; and hence confidence in the results is weakened. It is also possible that the probability of making a Type I error is greater

than the nominal probability ($\alpha = .05$) due to the heterogeneity of variance particularly in this instance where the larger sample (experimental group) has the smaller variance.

In order to control for other effects such as testing, instrumentation, and regression effects, tests of similarity between the experimental and control groups were made at the time of the pretest (two months post-injury). This was accomplished by comparing performances on the Verbal, Performance, and Full Scale IQ values of the WAIS-R using the statistical test noted previously. No significant differences were found ($p > .05$). Table 4.4 presents IQ means for both groups. Mean Verbal, Performance, and Full Scale IQ values were 83.89, 77.04, and 79.63 respectively, for the experimental group. Control group means were 84.28, 79.68, and 81.05 for Verbal IQ, Performance IQ, and Full Scale IQ values, respectively. The mean times during which both the experimental and control groups were first tested post-injury were not significantly different (54.46 vs. 63.02 days).

Similar analyses were computed for all WAIS-R subtests at pretest as well. Table 4.4 presents the mean subtest values for both groups. Six of the 11 subtest scaled score means appeared lower in the experimental group than in the control group. Nevertheless, no significant differences were noted between the experimental and the control groups ($p > .05$) among any of the eleven WAIS-R subtests.

Table 4.4

Comparison of WAIS-R IQ and Subtest Scaled Scores Between
Experimental and Control Groups at Pretest

	Experimental				Control				p
	<u>M</u>	<u>±</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>±</u>	<u>SD</u>	<u>n</u>	
Information	5.91	±	1.77	(47)	5.64	±	2.31	(37)	.565
Digit Span	7.19	±	2.55	(47)	8.05	±	2.79	(37)	.149
Vocabulary	7.04	±	1.89	(47)	6.65	±	1.82	(32)	.367
Arithmetic	7.51	±	2.20	(47)	7.10	±	2.10	(37)	.397
Comprehension	7.71	±	3.06	(46)	7.67	±	2.28	(37)	.944
Similarities	6.55	±	2.12	(47)	6.94	±	2.65	(37)	.466
Picture Completion	7.17	±	2.50	(47)	7.15	±	2.46	(38)	.982
Picture Arrangement	6.06	±	1.89	(46)	7.05	±	2.51	(38)	.054
Block Design	6.80	±	2.92	(45)	7.28	±	2.49	(36)	.415
Object Assembly	5.71	±	2.59	(45)	6.30	±	3.06	(38)	.357
Digit Symbol	4.70	±	1.63	(40)	5.24	±	2.29	(33)	.259
Verbal IQ	83.89	±	9.02	(47)	84.28	±	11.09	(38)	.860
Performance IQ	77.04	±	10.20	(47)	79.68	±	12.51	(36)	.298
Full Scale IQ	79.63	±	8.65	(45)	81.05	±	11.48	(38)	.532
Time Post-Injury ^a	54.46	±	18.86	(47)	63.02	±	35.59	(37)	.191

Note. Two-tailed probabilities were determined for all comparisons.

^aTime post-injury is in days.

It is important to restate the situation that variance differences surrounding ability-related variables (e.g., Full Scale IQ means) suggests that statements about group similarity cannot be made unequivocally. Rather, one group may be more homogeneous than another and this difference may lessen the overall effect of the control group in controlling for particular effects since one group will be expected to vary more about its mean than another. This uncertainty effects prognostication.

Tests of Primary Hypotheses

Hypothesis One: Full Scale, Verbal, and Performance IQ means for the experimental group at 12 months post-injury will be greater than their respective means at 2 months post-injury. This same hypothesis is advanced for the control group.

Table 4.5 presents WAIS-R Full Scale, Verbal, and Performance IQ means at 2 and 12 months post-injury, respectively for both the experimental and control groups. The Full Scale IQ mean for the experimental group increased from 79.63 to 93.95 (a difference of 14.31 points), while the control group Full Scale IQ increased from 81.05 to 92.39 (a difference of 11.34 points). The change from pretest to posttest was significant for both groups ($p < .001$) with the greatest change noted in the experimental group. Table 4.5 also includes effect sizes (Glass and Hopkins, 1984) which are standardized mean differences. Effect sizes provide information regarding the practical

Table 4.5

Experimental and Control Group Changes in WAIS-R IQ Mean Values
From Pretest to Posttest

		Pretest			Posttest					Effect
	<u>n</u>	<u>M</u>	<u>±</u>	<u>SD</u>	<u>M</u>	<u>±</u>	<u>SD</u>	Difference	<u>p</u>	size ^a
<hr/>										
Verbal IQ										
Experimental	(47)	83.89	<u>±</u>	9.02	93.12	<u>±</u>	9.53	9.23	.000*	1.02
Control	(38)	84.28	<u>±</u>	11.09	92.65	<u>±</u>	12.62	8.36	.000*	.75
Performance IQ										
Experimental	(47)	77.04	<u>±</u>	10.20	97.00	<u>±</u>	11.11	19.95	.000*	1.96
Control	(38)	79.68	<u>±</u>	12.51	93.73	<u>±</u>	13.24	14.05	.000*	1.12
Full Scale IQ										
Experimental	(47)	79.63	<u>±</u>	8.65	93.95	<u>±</u>	8.92	14.31	.000*	1.66
Control	(38)	81.05	<u>±</u>	11.48	92.39	<u>±</u>	12.36	11.34	.000*	.99

Note. One-tailed probabilities were determined for all comparisons. In all cases it was hypothesized that the posttest mean would be greater than the pretest mean.

^aeffect size = standardized mean difference, or $\frac{Y_{\text{posttest}} - Y_{\text{pretest}}}{s_{\text{pretest}}}$

*p < .001.

significance of the results and are used only when tests are significant. Thus, the difference of 14.31 points in the experimental group Full Scale IQ is a change of 1.66 standard deviations and represents a substantial change in score from pretest to posttest.

Significant changes were also noted on Verbal and Performance IQ means for both the experimental and control groups as well ($p < .001$) with the experimental group demonstrating the larger changes. Verbal IQ means increased from 83.89 to 93.12 (a difference of 9.23 points; effect size of 1.02) in the experimental group while the control group Verbal IQ increased from 84.28 at pretest to 92.65 (a difference of 8.36; effect size of .75) at posttest. The experimental group Performance IQ increased from 77.04 to 97.00 (a change of 19.95 points; effect size 1.96); there was a change of 14.05 points in the control group Performance IQ from 79.68 at pretest to 93.73 at posttest (effect size = 1.12). Thus, all changes from pretest to posttest were significant ($p < .001$) and in the expected direction. Therefore, the research hypothesis was empirically supported.

Hypothesis Two: Subtest scaled score means for the experimental group at 12 months post-injury will be greater than their respective subtest scaled score means at 2 months post-injury. The same hypothesis is advanced for the control group.

Tables 4.6 and 4.7 present changes in WAIS-R subtest scaled score means at 2 and 12 months post-injury for the experimental and control groups, respectively. Changes in all subtest scaled score means for both groups were significant ($p < .001$) and in the expected direction. The extent of change in the experimental group ranged from a difference of 0.75 scaled score points in Vocabulary to 3.71 scaled score points in Picture Arrangement. Effect sizes ranged overall from .39 to 1.97. The smallest change in subtest scaled scores means in the control group was also on Vocabulary with a mean increase of 1.00 scaled score points (effect size .56). The largest increase in the control group was that of Block Design with a change of 2.34 scaled score points and an effect size of .94. Effect sizes were generally the largest in the Performance section of the WAIS-R. Effect sizes also appeared larger, as a group, in the experimental group than in the control group. Since all subtest scaled score mean changes were in the expected direction and were significantly different ($p < .001$), the research hypothesis was supported.

Hypothesis Three: Full Scale, Verbal, and Performance IQ means of the experimental group will be greater than the Full Scale, Verbal, and Performance IQ means of the control group at one year post-injury.

Table 4.6

Experimental Group Changes in WAIS-R Subtest Scaled
Score Mean Values from Pretest to Posttest

	<u>n</u>	<u>Pretest</u>			<u>Posttest</u>			Difference	<u>p</u>	Effect size
		<u>M</u>	<u>±</u>	<u>SD</u>	<u>M</u>	<u>±</u>	<u>SD</u>			
Information	(47)	5.91	±	1.77	7.80	±	2.37	1.89	.000*	1.07
Digit Span	(47)	7.19	±	2.55	8.97	±	2.47	1.78	.000*	.70
Vocabulary	(44)	7.02	±	1.93	7.77	±	1.77	0.75	.000*	.39
Arithmetic	(47)	7.51	±	2.20	8.82	±	2.59	1.31	.000*	.60
Comprehension	(45)	7.68	±	3.09	9.68	±	2.35	2.00	.000*	.65
Similarities	(47)	6.55	±	2.12	9.23	±	2.66	2.68	.000*	1.26
Picture Completion	(47)	7.17	±	2.50	10.80	±	2.13	3.63	.000*	1.45
Picture Arrangement	(46)	6.06	±	1.89	9.78	±	2.92	3.71	.000*	1.97
Block Design	(47)	6.80	±	2.92	9.59	±	2.66	2.78	.000*	.96
Object Assembly	(42)	5.76	±	2.48	9.04	±	2.30	3.28	.000*	1.32
Digit Symbol	(40)	4.70	±	1.63	7.25	±	2.35	2.55	.000*	1.56

Note. One-tailed probabilities were determined for all comparisons. In all cases it was hypothesized that the posttest mean would be greater than the pretest mean.

*p < .001.

Table 4.7

Control Group Changes in WAIS-R Subtest Scaled Score Mean Values
from Pretest to Posttest

	<u>n</u>	<u>Pretest</u>			<u>Posttest</u>			Difference	<u>p</u>	Effect size
		<u>M</u>	<u>±</u>	<u>SD</u>	<u>M</u>	<u>±</u>	<u>SD</u>			
Information	(37)	5.64	<u>±</u>	2.31	7.13	<u>±</u>	2.49	1.48	.000*	.64
Digit Span	(37)	8.05	<u>±</u>	2.79	9.24	<u>±</u>	2.78	1.18	.000*	.42
Vocabulary	(31)	6.58	<u>±</u>	1.80	7.58	<u>±</u>	2.11	1.00	.000*	.56
Arithmetic	(37)	7.10	<u>±</u>	2.10	8.70	<u>±</u>	2.58	1.59	.000*	.76
Comprehension	(37)	7.67	<u>±</u>	2.28	8.72	<u>±</u>	2.53	1.05	.000*	.46
Similarities	(37)	6.94	<u>±</u>	2.65	8.70	<u>±</u>	2.81	1.75	.000*	.66
Picture Completion	(38)	7.15	<u>±</u>	2.46	9.28	<u>±</u>	2.37	2.13	.000*	.87
Picture Arrangement	(35)	7.08	<u>±</u>	2.54	9.28	<u>±</u>	2.40	2.20	.000*	.87
Block Design	(38)	7.28	<u>±</u>	2.49	9.63	<u>±</u>	2.46	2.34	.000*	.94
Object Assembly	(35)	6.40	<u>±</u>	3.05	8.65	<u>±</u>	2.41	2.25	.000*	.74
Digit Symbol	(33)	5.24	<u>±</u>	2.29	7.30	<u>±</u>	2.18	2.06	.000*	.90

Note. One-tailed probabilities were determined for all comparisons. In all cases it was hypothesized that the posttest mean would be greater than the pretest mean.

*p < .001.

Table 4.8 presents the Full Scale, Verbal, and Performance IQ means for both the experimental and control groups at 12 months post-injury. The time from injury to posttest were not significantly different between the experimental and control groups. The Verbal IQ mean for the experimental group was 93.12; a difference of 0.47 points from the control Verbal IQ mean of 92.65. A difference of 3.27 points was noted between the experimental Performance IQ mean of 97.00 and that of the control Performance IQ mean of 93.73. Control and experimental group Full Scale IQ means differed by 1.56 points. Statistical comparisons between the two groups did not demonstrate a significant difference among any of the three IQ means ($p > .05$). Therefore, the research hypotheses associated with the tests of the Full Scale, Verbal, and Performance IQ means in Table 4.8 were not supported, i.e., the experimental group was not significantly greater than the Full Scale, Verbal, and Performance IQ means of the control group at one year post-injury.

Hypothesis Four: All subtest scaled score means in the experimental group will be greater than all subtest scaled score means of the control group at one year post-injury.

Table 4.8 also contains the subtest scaled score means for both the experimental and control groups at 12 months post-injury. Only two subtests, however, Comprehension and Picture Completion proved to differ significantly between

Table 4.8

Comparison of WAIS-R IQ and Subtest Scaled Scores Between Experimental and Control Groups at Posttest

	Experimental				Control				p	Effect size ^a
	<u>M</u>	<u>±</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>±</u>	<u>SD</u>	<u>n</u>		
Information	7.80	±	2.37	(47)	7.13	±	2.46	(38)	.102	
Digit Span	8.97	±	2.47	(47)	9.10	±	2.87	(35)	.415	
Vocabulary	7.77	±	1.77	(44)	7.36	±	2.12	(36)	.178	
Arithmetic	8.82	±	2.59	(47)	8.65	±	2.56	(38)	.380	
Comprehension	9.69	±	2.32	(46)	8.71	±	2.50	(38)	.034*	.39
Similarities	9.23	±	2.66	(47)	8.68	±	2.78	(38)	.179	
Picture Completion	10.80	±	2.13	(47)	9.28	±	2.37	(38)	.001**	.64
Picture Arrangement	9.74	±	2.90	(47)	9.10	±	2.47	(37)	.140	
Block Design	9.59	±	2.66	(47)	9.63	±	2.46	(38)	.474	
Object Assembly	8.93	±	2.31	(44)	8.64	±	2.67	(37)	.308	
Digit Symbol	7.06	±	2.29	(46)	6.86	±	2.36	(38)	.350	
Verbal IQ	93.12	±	9.53	(47)	92.65	±	12.62	(38)	.425	
Performance IQ	97.00	±	11.11	(47)	93.73	±	13.24	(38)	.114	
Full Scale IQ	93.95	±	9.92	(47)	92.39	±	12.36	(38)	.265	
Time at Posttest ^b	373.12	±	43.59	(47)	375.28	±	75.47	(38)	.438	

Note. One-tailed probabilities were determined for all comparisons. In all cases it was hypothesized that that means of the experimental group would be greater than the mean of the control group.

^aeffect size = $(Y_e - Y_c)/S_c$. ^bTime at post-injury is in days.

*p < .05 ** p < .001.

the two groups ($p < .05$), and in the expected direction using a one-tailed test of significance. Effect sizes for Comprehension and Picture Completion were .39 and .64, respectively. The remaining subtests were not significantly different ($p > .05$). Therefore, the research hypothesis was not supported for nine out of 11 subtests since the means of the experimental group were not greater than the subtest scaled score means of the control group at one year post-injury. The research hypothesis was, however, supported for the WAIS-R subtests Comprehension and Picture Completion.

Hypothesis Five: Pretest-posttest differences in IQ and subtest scaled score means over time in the experimental group will be greater than the corresponding differences in the control group.

Differences between the pretest and posttest means of the experimental and control groups are presented in Table 4.9. That is, the average pretest to posttest difference of the experimental group was compared to the average pretest to posttest difference of the control group on the same variables. These "differences of differences" analyses correspond to an interaction. It will be noted that the change in Performance IQ in the experimental group from pretest to posttest was in the expected direction and significantly larger than the pretest to posttest change of the Performance IQ of the control group ($p < .05$) with an effect size of .50. The Verbal and Full Scale IQ changes in the experimental group appeared larger and in the expected

Table 4.9

Comparisons of WAIS-R IQ and Subtest Change Score Means from Pretest to Posttest

	Experimental				Control				p	Effect size
	<u>M</u>	<u>±</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>±</u>	<u>SD</u>	<u>n</u>		
Information	1.89	±	1.43	(47)	1.48	±	1.67	(37)	.117	
Digit Span	1.78	±	2.06	(47)	1.18	±	2.37	(37)	.115	
Vocabulary	0.75	±	1.16	(44)	1.00	±	1.18	(31)	.184	
Arithmetic	1.31	±	1.40	(47)	1.59	±	2.08	(37)	.246	
Comprehension	2.00	±	2.66	(45)	1.05	±	2.24	(37)	.042*	.42
Similarities	2.68	±	1.97	(47)	1.75	±	2.57	(37)	.038*	.36
Picture Completion	3.63	±	2.29	(47)	2.13	±	2.37	(38)	.002**	.63
Picture Arrangement	3.71	±	2.61	(46)	2.20	±	2.06	(35)	.002**	.73
Block Design	2.78	±	1.93	(47)	2.34	±	2.37	(38)	.177	
Object Assembly	3.28	±	2.09	(42)	2.25	±	2.68	(35)	.035*	.38
Digit Symbol	2.55	±	1.98	(40)	2.06	±	1.98	(33)	.149	
Verbal IQ	9.23	±	5.99	(47)	8.36	±	9.13	(38)	.308	
Performance IQ	19.95	±	9.43	(47)	14.05	±	11.82	(38)	.007**	.50
Full Scale IQ	14.31	±	6.48	(47)	11.34	±	9.58	(38)	.053	
Time Post-Injury ^a	318.65	±	49.50	(47)	314.75	±	86.75	(37)	.404	

Note. One-tailed probabilities were determined for all comparisons. In all cases it was hypothesized that change score means of the experimental group would be greater than all the change score means of the control group.

^aTime post-Injury is in days.

* $p < .05$. ** $p < .01$.

direction than the control group, but the differences between the change score means of the two groups were not statistically significant ($p > .05$). Overall, IQ changes from pretest to posttest ranged from 8.36 points in Verbal IQ of the control group to 19.95 points in Performance IQ of the experimental group. In addition, it should be noted that the time which elapsed between pretest and posttest between the two groups was not significantly different ($p > .05$) and reflected a pretest to posttest interval of approximately 10 1/2 months.

Similar analyses were performed on the differences in the change score means of the subtests in the two groups. Nine of the eleven subtests in the experimental group demonstrated changes in the expected direction with changed scores means in the experimental group exceeding change score means of the control group. However, only five of the nine subtests reflecting these larger changes were significantly different from the control group ($p < .05$). These significant changes were in the subtests Comprehension, Similarities, Picture Completion, Picture Arrangement, and Object Assembly. Effect sizes for the respective subtests were .42, .36, .63, .73, and .38. The two subtests in which changed score means were greater in the control group than in the experimental group were Vocabulary and Arithmetic. Differences between the two groups on these two subtests were not significantly different nor were the other four subtests ($p > .05$). Subtest change score means ranged from

0.75 points in Vocabulary in the experimental group to 3.71 points in Picture Arrangement also in the experimental group.

It should also be noted here that the variation in the experimental group was less than the control group in all but two measures. Thus, these results suggest greater homogeneity among the experimental group relative to the degree of change which occurred from pretest to posttest in the control group.

The comparison of pretest to posttest averages of the experimental and the control groups represents an interaction effect which is most fully conveyed by graphs. Figure 4.1 graphically displays the interaction between time and treatment effects for the Performance IQ means of the experimental and control groups. The presence of a disordinal interaction suggested that the treatment effect (additional testing) in the experimental group had a greater influence on the outcome in the experimental group Performance IQ value than the control group within the same period of time.

Figures 4.2 through 4.6 display the significant interactions between time and treatment for the WAIS-R subtests as well. Disordinal interactions as discussed above are evident in the subtests Similarities, Picture Arrangement, and Object Assembly. Ordinal interactions were noted in the subtests Comprehension and Picture Completion. It will be noted that these latter two subtests were not

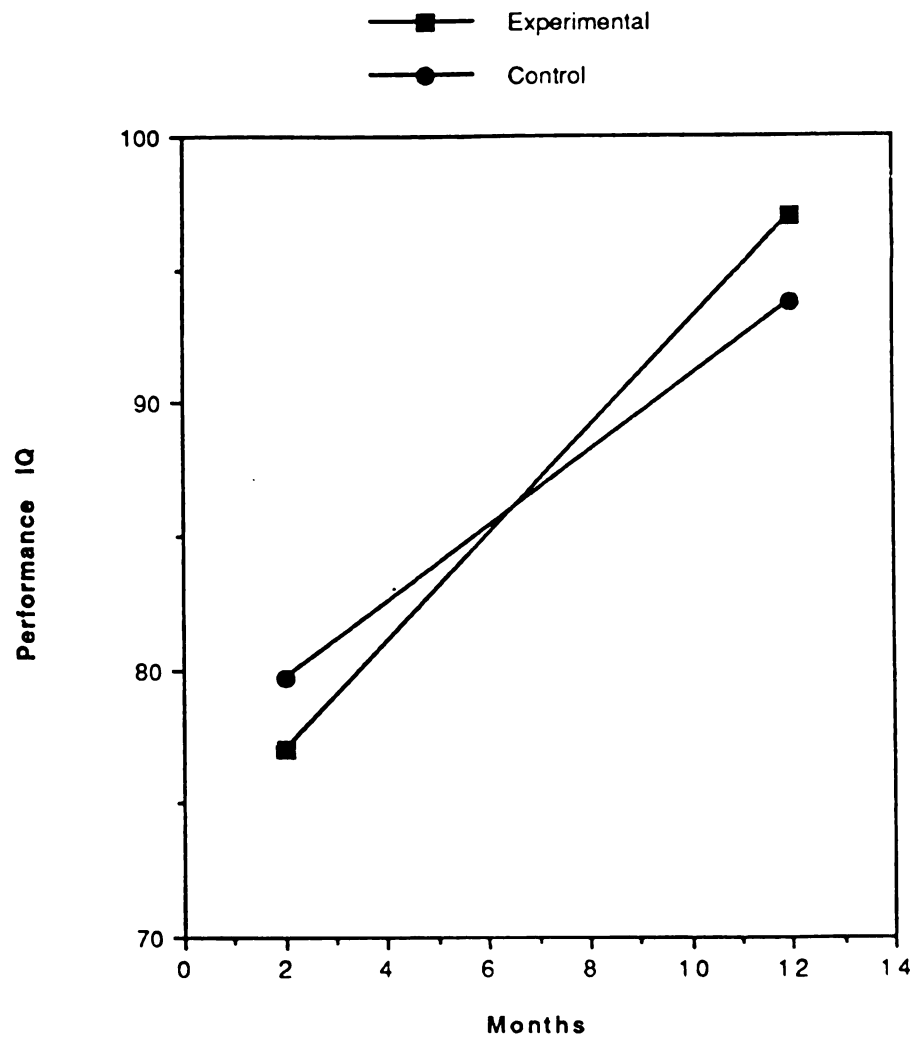


Figure 4.1. Interaction between time and treatment effects for Performance IQ means.

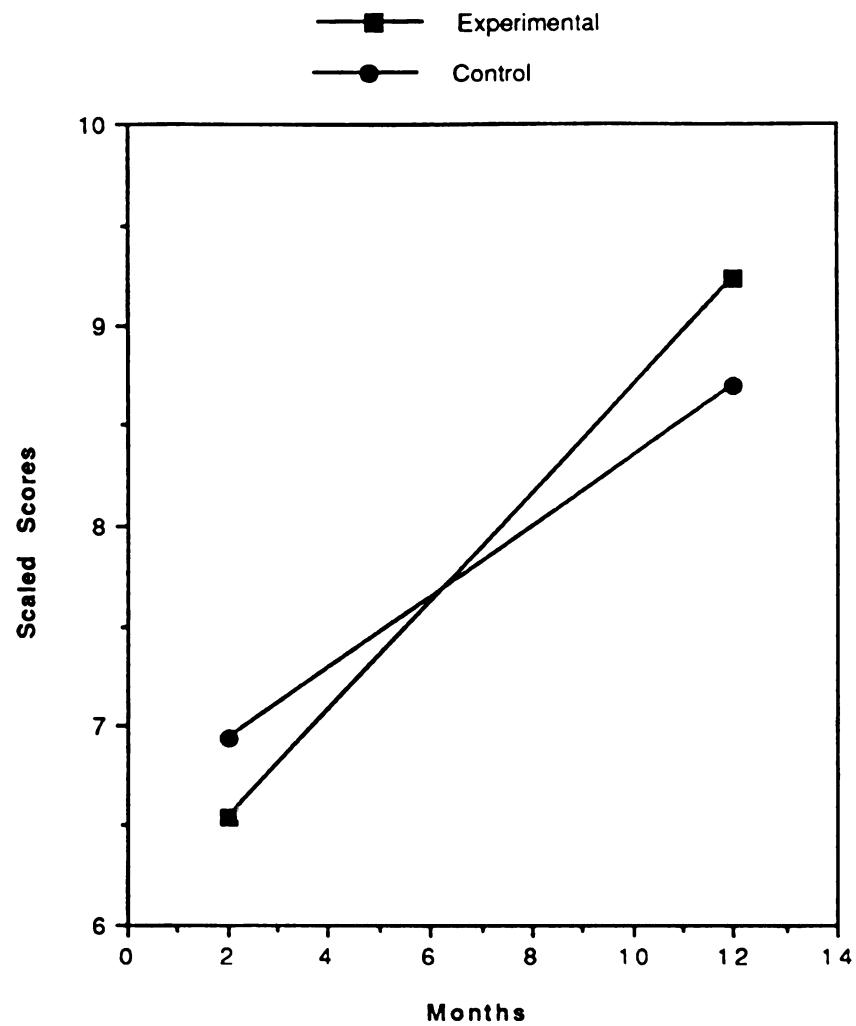


Figure 4.2. Interaction between time and treatment effects for Similarities scaled score means.

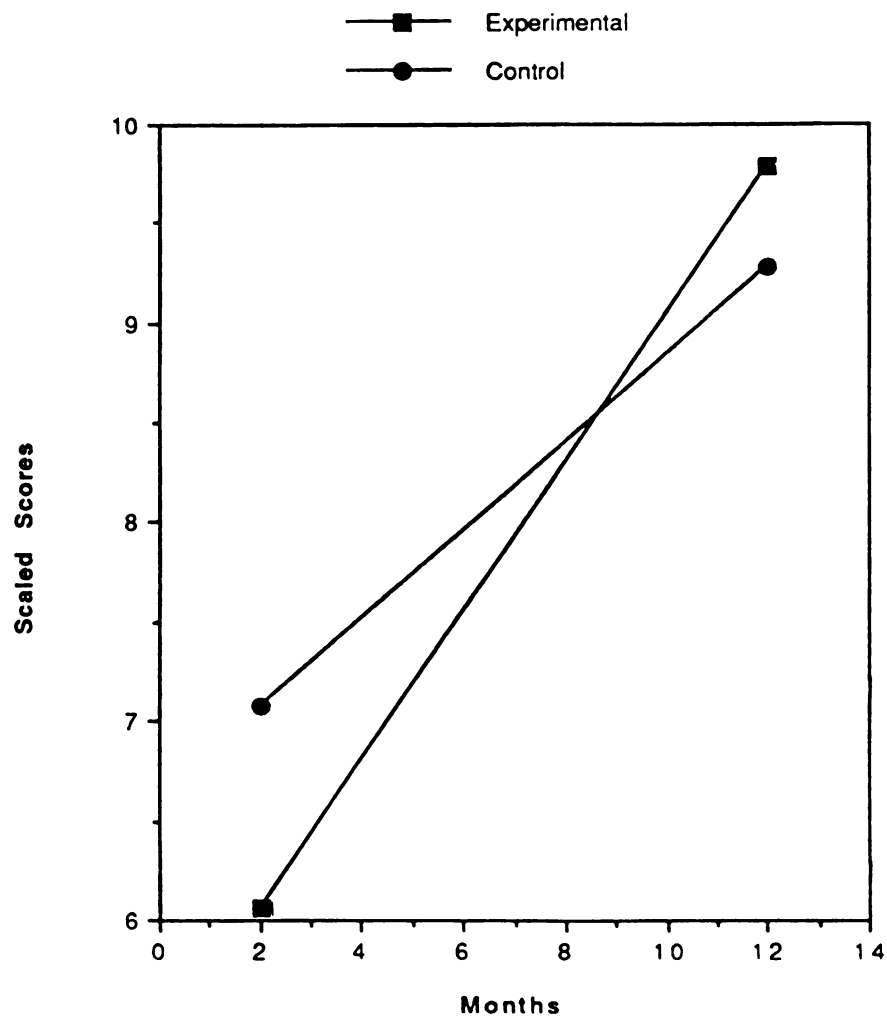


Figure 4.3. Interaction between time and treatment effects for Picture Arrangement scaled score means.

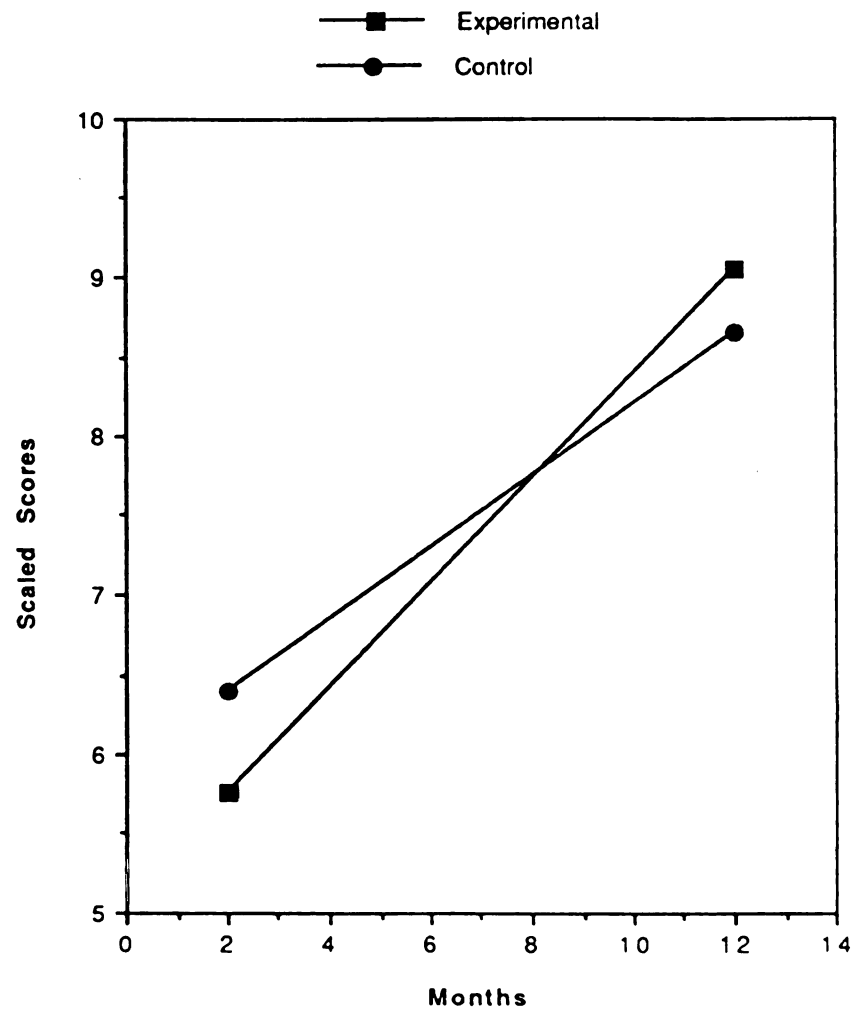


Figure 4.4. Interaction between time and treatment effects for Object Assembly scaled score means.

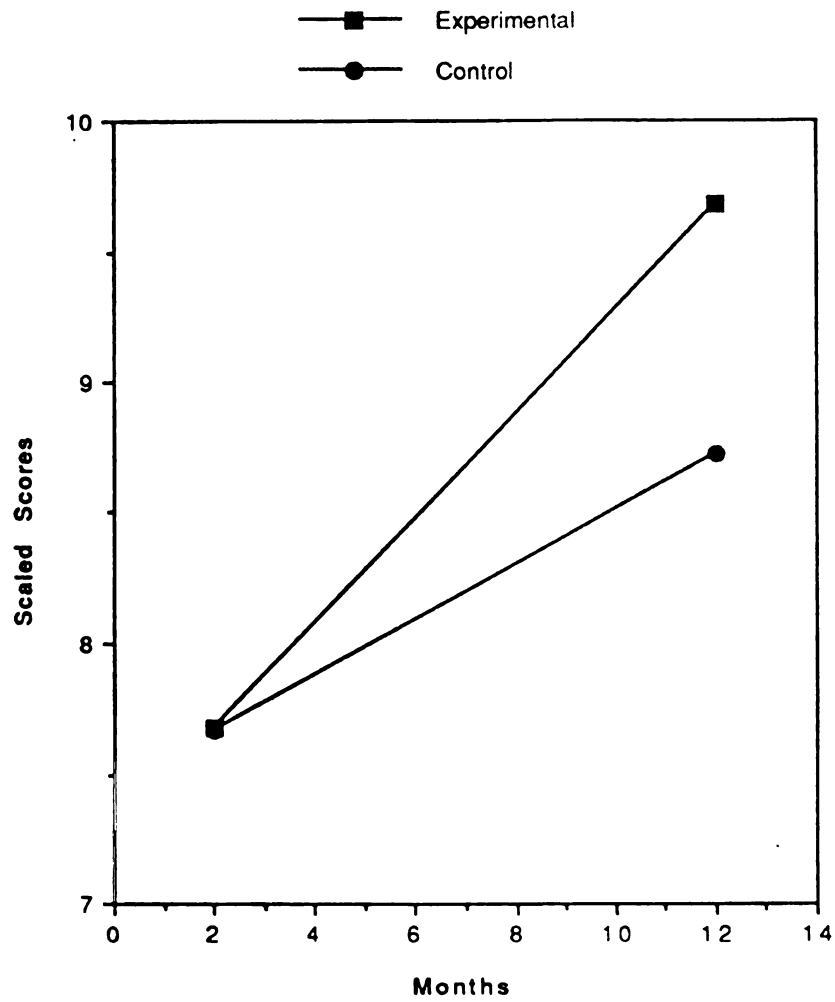


Figure 4.5. Interaction between time and treatment effects for Comprehension scaled score means.

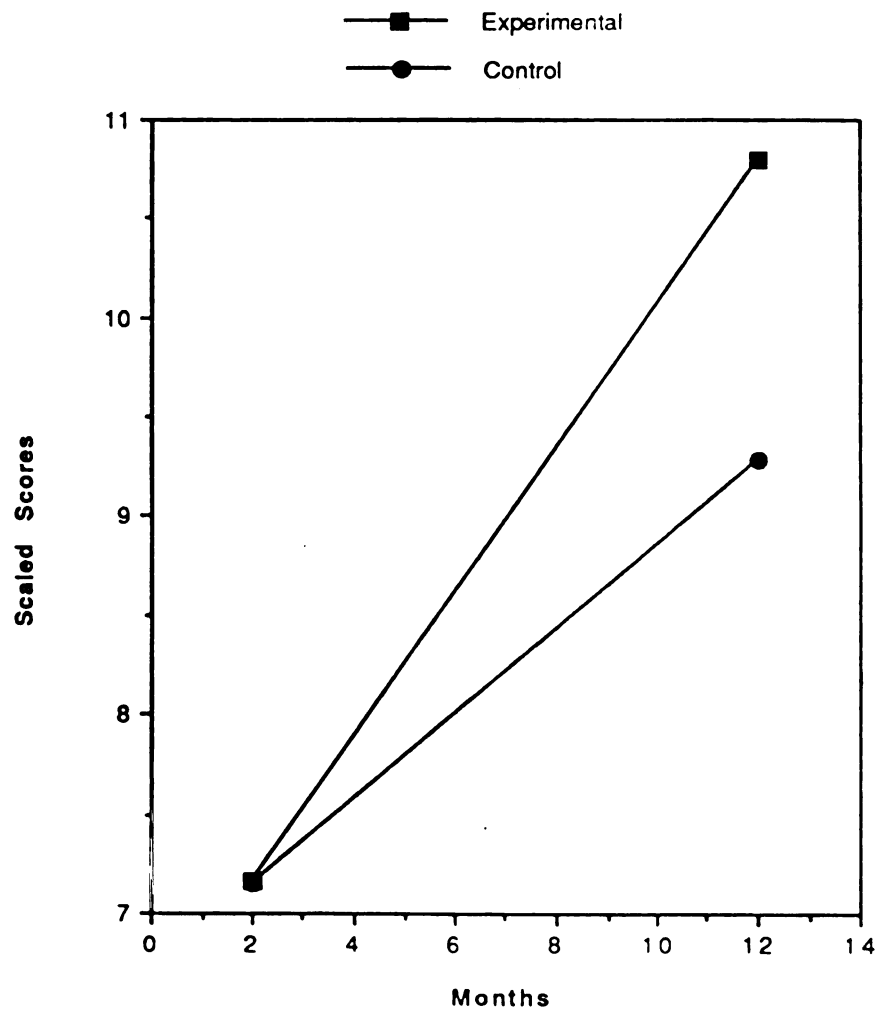


Figure 4.6. Interaction between time and treatment effects for Picture Completion scaled score means.

distinguishable (graphically) at pretest yet demonstrated considerable gain beyond the control group over time with additional testing.

Since pretest posttest differences between the experimental and control groups were statistically different for six of the variables, the research hypothesis was partially supported. Thus, the possible effects of test-retest exposure were noted in Performance IQ and the subtests Comprehension, Similarities, Picture Completion, Picture Arrangement and Object Assembly of the experimental group. These same effects were not evident as determined by pretest posttest differences in the eight remaining subtests and IQ values.

Tests of Secondary Hypotheses

Hypothesis One: Test-retest reliabilities on the WAIS-R for the experimental group will differ from the test-retest reliabilities of the control group at one year post-injury for Full Scale, Verbal and Performance IQs and all subtests.

Test-retest reliabilities of the WAIS-R IQs and subtests between the pretest and posttest administrations are listed in Table 4.10. Test-retest reliabilities appeared lower for the Performance IQ and higher for Verbal IQ on the WAIS-R for both groups. The weakest test-retest reliabilities were in the subtests Picture Arrangement (.48) and Picture Completion (.51) for the experimental group, and Picture Completion (.51) and Object Assembly (.53) for the control

group. The highest test-retest reliabilities were noted in the subtests Arithmetic (.84) and Vocabulary (.80) of the experimental group, and Vocabulary (.82) and Information (.75) in the control group. Test-retest reliabilities of the subtests ranged from .48 (Picture Arrangement) to .84 (Vocabulary); both of which were in the experimental group. Test-retest reliabilities of the WAIS-R IQs ranged from .58 (Performance IQ) in the control group to .79 (Verbal IQ) in the experimental group. Comparisons between the test-retest reliabilities of the experimental and control groups using Fisher's two-sample z-test are also presented in Table 4.10.

Test-retest reliabilities calculated in the experimental group appeared to be larger than the test-retest reliabilities calculated for the respective IQs and subtests in the control group. Nine of the possible 14 test-retest reliabilities seemed larger in the experimental group than the control group, but only one subtest, Arithmetic, had a significantly greater ($p < .05$) test-retest reliability than its control group analog. The test-retest reliability calculated on the subtest Picture Completion was the same for both the experimental and control groups (.51). The test-retest reliabilities for the remaining subtests Vocabulary, Comprehension, Picture Arrangement, and Digit Symbol reliabilities did not differ statistically between the experimental and control groups ($p > .05$).

Since only one test-retest reliability was significantly different between the experimental and control groups, the

Table 4.10

Comparisons Between the Pretest/Posttest Correlations
of the Experimental and Control Groups

	Experimental		Control		<u>z</u>
	<u>r</u>	<u>n</u>	<u>r</u>	<u>n</u>	
Information	.799	(47)	.759	(37)	0.457
Digit Span	.663	(47)	.637	(37)	0.227
Vocabulary	.807	(44)	.828	(31)	-0.306
Arithmetic	.842	(47)	.620	(37)	2.173*
Comprehension	.551	(45)	.569	(37)	-0.130
Similarities	.680	(47)	.558	(37)	0.858
Picture Completion	.519	(47)	.519	(38)	0.000
Picture Arrangement	.480	(46)	.653	(35)	-1.119
Block Design	.765	(47)	.541	(38)	1.825
Object Assembly	.619	(42)	.539	(35)	0.507
Digit Symbol	.553	(40)	.609	(33)	-0.337
Verbal IQ	.793	(47)	.710	(38)	0.894
Performance IQ	.611	(47)	.580	(38)	0.212
Full Scale IQ	.764	(47)	.679	(38)	0.808

Note. Two-tailed probabilities were determined for all comparisons. It was hypothesized in all cases that test-retest reliabilities of the experimental group would differ from the test-retest reliabilities of the control group because the experimental group would reflect the more recent influences of additional testing. Since the subjects might remember some of the answers, the correlation between the two tests would not be independent and a high correlation would result. r = correlation coefficient; z = Fisher's z; a mathematical transformation of r.

* $p < .05$.

research hypothesis was not supported. Therefore, with the exception of the subtest Arithmetic, the test-retest reliabilities on the WAIS-R for the experimental group did not differ from the test-retest reliabilities of the control group at one year post-injury using a test-retest interval of approximately 10 1/2 months.

Hypothesis Two: Changes in Full Scale, Verbal, and Performance IQ means in the experimental group will be quadratic over a one year period post-injury.

Table 4.11 presents the WAIS-R subtest scaled score and IQ means over four separate test administrations during a period of approximately one year since time of injury.

A review of Table 4.11 indicates that the changes in Verbal, Performance, and Full Scale IQs appear to be in a positive direction over all four test administrations. Noting the changes in the respective Verbal subtests, the change from one test administration to the next remains in a positive direction, however the amount of change appears to slow down with successive testing. Between the third and fourth test administrations the Verbal subtests, Digit Span and Arithmetic actually demonstrated a decrease in average scores.

Performance subtests appeared to reveal larger changes from one test administration to another than did the Verbal subtests, but these too, slowed down with successive testing. No Performance subtest, however, appeared to decrease in its mean score over time. The larger increases

Table 4.11

Change in WAIS-R IQs and Subtest Scaled Scores Over Four Test Administrations

	Months Post-Injury											
	2			4			8			12		
	<u>M</u>	<u>±</u>	<u>SD</u>	<u>M</u>	<u>±</u>	<u>SD</u>	<u>M</u>	<u>±</u>	<u>SD</u>	<u>M</u>	<u>±</u>	<u>SD</u>
Information	5.91	±	1.77	7.19	±	2.23	7.61	±	2.21	7.80	±	2.37
Digit Span	7.19	±	2.55	8.91	±	2.43	9.12	±	2.55	8.97	±	2.47
Vocabulary	7.04	±	1.89	7.46	±	1.74	7.73	±	1.60	7.79	±	1.79
Arithmetic	7.51	±	2.20	8.23	±	2.32	8.95	±	2.38	8.82	±	2.59
Comprehension	7.71	±	3.06	8.82	±	2.13	9.27	±	2.07	9.69	±	2.32
Similarities	6.55	±	2.12	8.19	±	1.93	8.91	±	2.56	9.23	±	2.66
Picture Completion	7.17	±	2.50	9.31	±	2.28	10.14	±	2.05	10.80	±	2.13
Picture Arrangement	6.06	±	1.89	8.21	±	2.52	8.63	±	2.23	9.74	±	2.90
Block Design	6.80	±	2.92	8.55	±	2.47	9.34	±	3.04	9.59	±	2.66
Object Assembly	5.71	±	2.59	7.93	±	2.86	8.78	±	2.57	8.93	±	2.31
Digit Symbol	4.71	±	1.65	5.87	±	2.13	6.50	±	2.06	7.09	±	2.28
Verbal IQ	83.89	±	9.02	89.93	±	8.17	92.82	±	9.44	93.12	±	9.53
Performance IQ	77.04	±	10.20	87.85	±	11.06	92.44	±	11.73	97.00	±	11.11
Full Scale IQ	79.63	±	8.65	88.08	±	8.84	92.02	±	9.92	93.95	±	9.92
Time Post-Injury ^a	54.46	±	18.86	115.74	±	20.51	263.78	±	26.41	373.12	±	43.59

^aTimes post-Injury are in days.

noted in the early testings also reflected lower performances initially when compared to the Verbal subtests. This is typical in TBI WAIS-R test patterns.

Trend analyses were performed on IQ means to assess whether the changes noted over time were linear or curvilinear. Since time of testing led to unequal intervals over one year post-injury, trend coefficients were derived by hand following Kirk (1982, Appendix C). All statistical hypotheses were nondirectional. F-tests for linear and quadratic effects over time were computed (Glass & Hopkins, 1984). Table 4.12 summarizes the trend analyses for both linear and quadratic trend components for each of the three IQ variables.

The analysis of trend components demonstrated the presence of a statistically significant linear ($p < .001$) as well as a quadratic ($p < .025$) trend in all three WAIS-R IQ means over time. This is supported by Figure 4.7 which suggested that the IQ values increase initially in a linear fashion, and subsequently the change in IQ means begin to slow-up resulting in a curvilinear pattern of change over time.

A measure of explained variance, h^2 , (eta-squared) was also computed for all statistically significant trends. The proportion of variation associated with time in the Verbal, Performance, and Full Scale IQs due to a linear effect was 74%, 87%, and 82%, respectively. That is, 74% of the variation due to time for Verbal IQ was attributable to the

Table 4.12

Trend Analysis Summary of WAIS-R IQs by Time in the Experimental Group

<u>Source</u>	<u>SS</u>	<u>df</u>	<u>MS</u>	<u>F</u>	<u>p</u>	<u>a_h²</u>
VIQ ^b by Time	2588.31	3				
linear	1920.31	1	1920.31	23.45	0.001***	.74
quadratic	553.57	1	553.57	6.76	0.025*	.21
remainder ^c	114.43	1	100.00	1.39	0.250	
Error	15065.14	184	81.87			
Total	17653.46	187	94.40			
PIQ ^d by Time	10316.14	3				
linear	8932.27	1	8932.27	73.23	0.001***	.87
quadratic	817.56	1	817.56	6.70	0.025*	.08
remainder ^c	556.31	1	556.31	4.64	0.050*	.08
Error	22443.48	184	121.95			
Total	32759.63	187	175.18			
FSIQ ^e by Time	5680.55	3				
linear	4630.07	1	4630.07	52.90	0.001***	.82
quadratic	782.88	1	782.88	8.94	0.010**	.14
remainder ^c	267.60	1	267.60	3.05	0.100	
Error	16105.40	184	87.52			
Total	21785.95	187	116.50			

^a $h^2 = SS_{\text{effect}} / SS_{\text{Time}}$ = measure of practical significance. ^b VIQ = Verbal IQ. ^c This is an aggregate of the higher order trends: $SS_{\text{remainder}} = SS_{\text{Total}} - SS_{\text{error}} - SS_{\text{linear}} - SS_{\text{quadratic}}$.

^d PIQ = Performance IQ. ^e FSIQ = Full Scale IQ.

* $p < .05$. ** $p < .01$. *** $p < .001$.

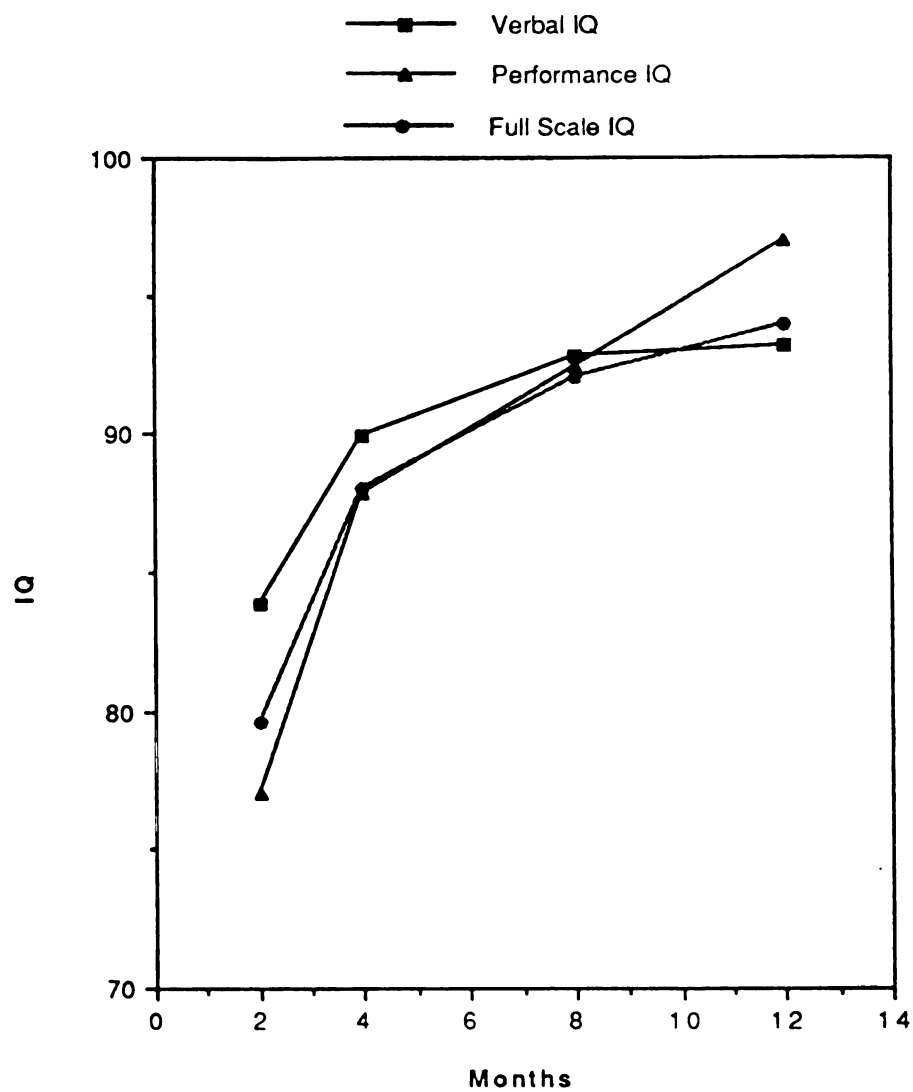


Figure 4.7. Change in Verbal, Performance, and Full Scale IQ means over time.

linear trend. The proportion of variation due to a quadratic effect in Verbal, Performance, and Full Scale IQs was 21%, 8%, and 14%. Thus, the variation in recovery appears to be primarily linear and secondarily quadratic.

Figure 4.7 demonstrates the changes in Verbal, Performance, and Full Scale IQ means over the course of four test administrations within a one year period of time. From this figure, it will be noted that the Performance IQ mean value begins lower than the Verbal IQ mean value at the first test administration, approximates the Verbal IQ by the third testing, and then surpasses Verbal IQ by the fourth and final test administration.

Since trend analysis demonstrated both a statistical significant linear and quadratic trend for IQ means over time, the research hypothesis was supported. Therefore, changes in IQ mean values are curvilinear (as well as linear) within a one year period of recovery for all IQ means in the experimental group.

Hypothesis Three: Indices of internal consistency in the WAIS-R for the experimental group will change over the four test administrations.

In order to obtain a global estimate of the WAIS-R internal consistency, the following procedure was employed. All possible pairwise (Pearson) correlations among the eleven subtests for each time were computed and the median correlations found. Using this method, median correlations of .48, .43, .44, and .38 were found for the first through

fourth test administrations, respectively. These correlations suggest only moderate consistency among the subtest performances based on scaled scores. It is also interesting to note that the median correlations demonstrated a decrease over time suggesting that variables influencing test results, possibly including test-retest practice effects, may have differentially affected the subjects.

CHAPTER V

Discussion and Conclusions

Summary

Procedures to investigate test-retest practice effects, reliability, and stability of the WAIS-R (Wechsler, 1981) over time are in short supply. While a limited number of studies have been completed on its predecessor, the WAIS (Wechsler, 1955), psychologists have and will continue to practice under a handicap until a considerable amount of actuarial base-rate information is collected and published on normal as well as patient (i.e., neurologically impaired, developmentally disabled) populations (Matarazzo & Herman, 1984). Without such information, an understanding of the influences of repeat administrations on neuropsychological tests will be difficult, implying that the interpretation of test scores and test score changes may be in error (Campbell, 1983). This is particularly important in charting intellectual and cognitive recovery in traumatically-brain-injured survivors when greater specificity and accuracy in determining change over time is central to rehabilitation and/or vocational planning, as well as medico-legal determinations.

Consistent with the extremely limited research on test-retest practice effects, reliability, and stability with either the WAIS or WAIS-R (Matarazzo & Herman, 1984), there is a dearth of published research on either instrument investigating test-retest practice effects, reliability, and stability with recovering traumatically-brain-injured (TBI) survivors. Without this type of information, it would be extremely difficult, if not impossible, to discern whether observed changes upon re-evaluation were due to treatment effects, spontaneous recovery, test-retest practice effects, or test unreliability (Campbell, 1983). Therefore, it was the purpose of this study to provide a data base from which changes observed on the WAIS-R over time in a population of recovering TBI patients could be examined.

In particular, this study attempted to: a) provide descriptive information regarding the stability of test scores of individuals who were recovering from TBI within a one year period of time; b) determine the possible existence, extent, and magnitude of test-retest practice effects resulting from repeated administrations of the WAIS-R over time; and c) provide information concerning the test-retest reliability of the WAIS-R through serial testing.

In order to provide the information for this data base, and to empirically test the research questions noted previously in Chapter One, an experimental group consisting of 47 adult TBI survivors tested four times with the WAIS-R at approximately 2, 4, 8, and 12 months post-injury was

compared with a matched control group of 38 similar TBI survivors. The control group subjects were tested only twice with the WAIS-R at approximately 2 and 12 months post-injury. Because of the retrospective nature of this study, randomization was not possible. Therefore, attempts were made to match the two groups as closely as possible on a number of variables thought to have the potential to confound intelligence testing and/or cognitive recovery. In addition, efforts were made to avoid methodological short-comings as reported in the neuropsychological literature (e.g., Levin et al., 1982).

Utilizing a quasi-experimental design referred to as the "non-equivalent control" (p. 47) design (Campbell & Stanley, 1963) groups represented a between subjects factor, and time a within subjects factor. Verbal, Performance, and Full Scale IQ values as well as all WAIS-R subtest scaled scores served as dependent variables. The two additional administrations of the WAIS-R at approximately 4 and 8 months post-injury, respectively were considered the treatment condition.

Test-retest gains or practice effects were then investigated by making comparisons of tests scores between the experimental and control groups in the manner noted previously in Chapter Three. The extent of recovery within the experimental and control groups over time were also investigated as well as the determination of test-retest correlations and changes in internal consistency. The

section to follow attempts to qualify these results given the limitations of this study.

Discussion

The omnibus research question in this study asked whether there was a test-retest practice effect that occurred with multiple testing of the WAIS-R, in addition to the course of natural recovery, in surviving traumatically-brain-injured individuals. In order to answer this larger question, primary and secondary research questions were addressed and tested with five primary and three secondary hypotheses, respectively. Accordingly, discussion of the results are organized in order of the hypotheses tested.

Hypothesis One.

In order to assess whether test-retest practice effects occur due to multiple testings with the WAIS-R, it was necessary to first determine whether TBI survivors demonstrated changes in intelligence measures over time on a single retest. The results of testing hypothesis one suggested that significant changes ($p < .001$) do occur and in a positive (gain) direction between two and twelve months, post-injury. Specifically, the control group demonstrated mean gains of 8.36, 14.05, and 11.34 points in Verbal, Performance and Full Scale IQs, respectively. These gains in IQ measures were, in general, consistent with IQ gains

reported by Becker (1977) and Drudge et al. (1984) who also studied head-injured subjects, but who employed a single retest with the WAIS.

In fact, many of the changes which occurred upon the single retest of the control group in this study, demonstrated characteristics common to other studies incorporating TBI populations (Diller & Ben-Yishay, 1983; Drudge et al., 1984; Mandleberg & Brooks, 1975). One such characteristic was the observation that at the time of the initial testing, all IQ values were below average levels of intelligence. Most notable was the greater impairment in Performance IQ mean (\bar{x} = 79.68) relative to the Verbal IQ mean (\bar{x} = 84.28). This pattern was consistent with research previously reported (Becker, 1977; Bornstein, 1983; Drudge et al., 1984; Dye et al., 1981; Fisher, 1985; Mandleberg, 1976; Tabbador et al., 1984). This between-scales discrepancy, which is typically observed during acute stages of head injury, is attributed to the resilient, language-mediated aspects of the Verbal section (Drudge et al., 1984; Fisher, 1985; Mandleberg & Brooks, 1975; Ruesch, 1944), in contrast to the more vulnerable, problem-solving skills of the Performance section (Brooks, 1975; Drudge et al., 1984; Mandleberg & Brooks, 1975). As noted previously, diminution of any one of these latter skills may depress Performance skills (Mandleberg & Brooks, 1975).

Another observation noted in this study and common to other reports involving TBI subjects was the larger

test-retest gain observed in the Performance IQ mean (\bar{x} = 14.05; effect size = 1.12) when compared to the Verbal IQ mean gain (\bar{x} = 8.36; effect size = .75). In addition, the Performance IQ mean (\bar{x} = 93.73) surpassed the Verbal IQ mean (\bar{x} = 92.65) upon a single retest. One explanation for these dramatic test-retest gains is that the changes in Performance IQ means are larger relative to Verbal IQ since the Verbal measures demonstrate only minimal increases on retest because the verbal abilities are more resilient to head injury and are less impaired when tested initially (Diller & Ben-Yishay, 1983; Seidenberg et al., 1981). Other explanations for these changes suggest that although the Performance measures maybe more sensitive to brain-injury in the acute stage of recovery, they become less so as the time between testing and the traumatic event increases (Drudge et al., 1984). Furthermore, greater gains in the Performance subtests may reflect the phenomenon of regression-to-the-mean where, upon repeat testing, there is a greater tendency for more impaired scores to converge towards the mean (Eisdorfer, 1983). Thus, in areas of greater impairment (Performance IQ in the acute state) there may be greater cognitive improvement observed with the repetition of the test (Lezak, 1983).

The significant gains reported previously for the control group retested only once resulted in all IQ means falling within the average range of intelligence (Verbal IQ, \bar{x} = 92.65; Performance IQ, \bar{x} = 93.73; Full Scale IQ, \bar{x} =

92.39). While these changes do not mean that these individuals returned to normal (Dodrill & Troupin, 1975; Matarazzo et al., 1976), the mean IQ scores were within the expected range of functioning for the general population. Therefore, the test-retest gains were interpreted as improvements in scores from previously low average or borderline levels of intellectual functioning. Furthermore, these improvements were considered to be largely attributable to recovery rather than from artifacts of multiple testing.

Support for this inference stems from the fact that the gains noted in the control group IQ means (i.e., VIQ = 8.36; PIQ = 14.05; FSIQ = 11.34 points) far exceeded the test-retest gains (i.e., VIQ = 3.30; PIQ = 8.40; FSIQ = 6.20 points) attributed to practice effects by Matarazzo and Herman (1984) and others (Ryan et al., 1985; Warner, 1983) in normal as well as clinical populations when using the WAIS-R. The improvements in Verbal, Performance, and Full Scale IQ means in the control group also exceeded the test-retest gains reported in Chapter Two in normal, psychiatric, alcoholic, medical, mentally retarded, and other neurologically-impaired (non-TBI) populations on the WAIS.

These retest gains are particularly important when the test-retest intervals are considered. The lengthier test-retest interval used in the control group should have reduced this group's familiarity with the test. This is in

contrast to those subjects in the previously reported studies for whom test-retest intervals were much shorter, and hence familiarity and retest gains would be expected to be larger. In addition, it has been reported that traumatically-brain-injured survivors are frequently deficient in learning (Lezak, 1979) and memory abilities (Brooks, 1976); both of which are functions instrumental for test-retest practice effects to occur (Drudge et al., 1984). These cognitive deficits, when augmented by the average test-retest interval in this study (10 1/2 months), should have greatly reduced the likelihood of appreciable incidental learning in the control group (Drudge et al., 1984). Support of this kind was also provided by Schau et al. (1980) who reported that test-retest practice effects would not be detectable over a one year period of time.

Since repetition of a test may lead to an increased familiarity and improved performance on that test, retest practice effects cannot be summarily ruled out as a factor contributing to the gains on a single retest. However when the same test is given multiple times, as in the experimental group, the retest effects may become more prominent when compared to a single retest within the same time frame.

Essentially, all of the characteristics noted above for the control group were the same for the experimental group. The only differences were that the magnitude of the retest gains between two and twelve months post-injury appeared

greater in the experimental group for the Verbal, Performance and Full Scale IQ means ($\bar{x} = 9.23$, effect size = 1.02; $\bar{x} = 19.95$, effect size = 1.96; $\bar{x} = 14.31$, effect size = 1.66), and levels of recovery were higher (Verbal IQ, $\bar{x} = 93.12$, Performance IQ, $\bar{x} = 97.00$; and Full Scale IQ, $\bar{x} = 93.95$) than the control group.

Given the fact that both groups were not statistically different ($p > .05$) at pretest the greater magnitude of scores and retest gains in the experimental group were thought to be attributable to test-retest practice effects. These results suggest that while TBI patients may demonstrate a return to average intelligence within one year of injury on a single retest with the WAIS-R, additional retesting may inflate those test scores accordingly. In addition, gains on a single retest are thought to reflect recovery, however, the degree to which test-retest practice effects take place upon a single retest is unclear and perhaps mitigated by other factors such as the test-retest interval. Thus, the clinician is cautioned about making interpretations regarding recovery of intellectual functioning when the WAIS-R is administered more than twice within a one year period of time.

Hypothesis Two.

Consistent with the changes noted in the WAIS-R IQ mean values, both the control and experimental groups demonstrated significant changes ($p < .001$) in a positive

(gain) direction on all Verbal and Performance subtests upon retest. Specifically, the control group demonstrated mean gains of 1.00 to 2.34 scaled score points. The magnitude of the test-retest gains noted on these subtest scaled scores were also consistent with the test-retest gains on WAIS subtests reported by both Becker (1977) and Drudge et al. (1984) who employed head-injured subjects with a single retest of the WAIS.

Like the composite IQs, the test-retest gains demonstrated among the subtests were the largest in the Performance section. Retest gains in the control group Performance subtests ranged from 2.06 to 2.34 scaled score points with effect sizes ranging from .74 to .94. Verbal subtests in the control group ranged from 1.00 to 1.59 scaled score points with effect sizes ranging from .42 to .76. Within the control group, Block Design demonstrated the largest gain on retesting, followed by Object Assembly, Picture Arrangement, Picture Completion, Digit Symbol, Similarities, Arithmetic, Information, Digit Span, Comprehension, and Vocabulary.

Retest gains in the experimental group were larger than the control group with the greatest gains also noted in the Performance section. Gains ranged from 2.55 to 3.71 on Performance subtests. Effect sizes ranged from .96 to 1.97. Verbal subtest gains ranged from .75 to 2.68 scaled score points with effect sizes ranging from .39 to 1.26. The order of subtests for the experimental group varied somewhat

from the pattern of the control group although five of the six largest gains also occurred on the Performance subtests. The order of gains from largest to smallest were Picture Arrangement, Picture Completion, Object Assembly, Similarities, Block Design, Digit Symbol, Comprehension, Information, Digit Span, Arithmetic, and Vocabulary.

While the explanations for the larger gains in Performance IQ relative to Verbal IQ noted previously may also hold for the Performance subtests, an additional explanation is possible. The tendency for Performance subtests to improve on retest may be partially explained by the fact that all Performance subtests are timed tests, therefore, a possibility exists of earning bonus points for speed; that is, correctly completing the tasks as quickly as possible (Catron, 1978; Steisel, 1951). Additional reasons for greater improvements in Performance subtests may also include developing strategies and forming Gestalts which also result in higher scores.

It should also be noted that the retest gains on the subtest Vocabulary were consistently the smallest in both the experimental and control groups. Similar findings regarding Vocabulary were also reported by Catron (1978), but with normal college students. This lends further support to the earlier contention that the Verbal subtests, and hence total Verbal IQ, are more resilient to head trauma and demonstrate fewer gains as a result. Therefore,

Vocabulary may represent the subtest which is least susceptible to practice effects.

Impaired at the time of the initial testing, all subtests in both groups were either at or within one standard deviation of the WAIS-R mean (\bar{x} = 10; SD = 3) at retest. The retest mean gains on the subtests in both groups were also larger than the gains noted in studies reported previously with the WAIS-R that employed normal as well as clinical (non-TBI) populations (Matarazzo & Herman, 1984; Ryan et al., 1985; Warner, 1983). Subtest gains in the control group were two to five times larger than the retest gains noted in the published studies above, however, the experimental group subtest gains were larger than the control group. In addition, the test-retest subtest gains in this study were greater than the gains reported in studies employing the WAIS on normal, psychiatric, medical, alcoholic, and other neurologically-impaired (non-TBI) populations.

Since the subtest gains demonstrated on a single retest in this study far exceeded the test-retest gains thought to be due to practice effects alone in the studies noted previously, the excess retest gains here were interpreted as improvements due to actual recovery. This was further supported by the Becker (1977) and Drudge et al. (1984) studies that demonstrated retest gains of similar strength. The similarity of magnitude among the studies employing TBI subjects also suggested that improvements upon retest cannot

simply be attributable to test-retest practice effects since they, too, demonstrated greater gains than those reported in studies of various normal and clinical (non-TBI) populations.

However, when the WAIS-R is repeated on more than one occasion and within the same 10 1/2 month period, gains in subtest scaled scores may become inflated due to retest practice effects resulting in increases in total scaled score points and larger IQ means. Caution must, therefore, be exercised when interpreting changes in individual subtest scaled scores when the WAIS-R is given more than two times within a 10 1/2 month period post-injury.

In summary, since the WAIS-R demonstrated significant changes ($p < .001$) on all dependent variables (both IQ and subtest scaled scores) between pretest and posttest, it could be alleged that in studies with recovering TBI subjects the WAIS-R lacks what Matarazzo et al. (1980) referred to as high clinical reliability; that is, the absence of a meaningful change in scores from initial test to retest. In contrast, the changes noted above are meaningful in that the gains demonstrated even on a single retest are significantly large such that they probably cannot be explained by test-retest practice effects alone. Since the test-retest gains in this study exceeded those retest gains reported to occur in normal and non-TBI populations where test-retest practice effects were thought to be maximum, the most plausible explanation to account for

these additional gains would be that they reflect the amount of recovery that has taken place since the time of the initial testing. However, those gains that result from additional, multiple testings and are in excess of those gains reported in the control group are likely to represent the accumulation of retest practice effects due to additional testings. The current results suggest that additional practice effects may result in increases of approximately 1, 6, and 3 IQ points in Verbal, Performance and Full Scale IQ scores, respectively.

Hypothesis Three.

The question tested by hypothesis three asked whether IQ means at one year post-injury would differ between the group of subjects tested four times and the group tested only twice with the WAIS-R. The basis for the hypothesis stemmed from the understanding that, if test-retest practice effects existed, the practice effects would be of such a magnitude as to be evident in the IQ means of the experimental group when compared to the control group IQ means at posttest.

While a trend in this direction was observed with the experimental group appearing to demonstrate larger IQ means at posttest than the control group, none of the differences were statistically significant ($p > .05$). As will be discussed below, the sizeable variability surrounding the posttest means was sufficient to reject tests of significance.

Hypothesis Four.

Like hypothesis three, hypothesis four asked whether subtest scaled scores would differ between the experimental and control group at one year post-injury. As noted above, a trend was observed in which the experimental group appeared to demonstrate larger means at one year post-injury than the control group. Specifically, the experimental group means were higher on nine of the eleven subtests but only Comprehension and Picture Completion reached statistical significance. Digit Span and Block Design appeared to demonstrate larger means in the control group at one year post-injury, but were not significantly different.

Despite this trend of greater experimental than control group means at posttest, the differences between posttest means was not statistically significant. Variances surrounding the mean values in the control group, however, appeared larger in eight of the eleven subtests. Only the variance in the subtests Arithmetic, Picture Arrangement, and Block Design appeared greater in the experimental group.

The greater variability surrounding the control group means may be due to smaller sample sizes; however, another reason for the variance may be the fact that the subjects in the control sample did not have the benefit of multiple testing. As a result, the control group demonstrated less consistency in their level of performance since test-retest practice effects due to retest familiarity was minimal. In

other words, the experimental sample which had opportunities to become quite familiar with the WAIS-R, may have demonstrated greater convergence of scores since the samples' performances may have created a ceiling (Dikman et al., 1983; Schau et al., 1980) or homogenizing effect wherein further improvements, and hence variability, in scores were reduced (see Figure 4.8).

This homogenization effect is likely to occur as the result of the subjects' inability to improve their obtained scores beyond a certain threshold. This, in turn, forces scores to congregate around a particular score and variability is reduced. For example, extremely bright, neurologically-intact individuals may do so well on some tests initially, that there is little room for test-retest practice effects to take place. Hence, the variability among those subjects' scores is reduced making the group more homogeneous in appearance.

In contrast to the healthy, neurologically-intact subjects above, TBI subjects demonstrate gains in scores due to a combination of true recovery and probable test-retest practice effects. However, just as in the example above, these subjects may eventually experience a homogenization effect as well. This most likely stems from the TBI subjects' inability or incapacity to exceed certain levels or thresholds of performance. The limits or ceiling imposed on the degree to which these subjects can progress is probably influenced by several factors including the

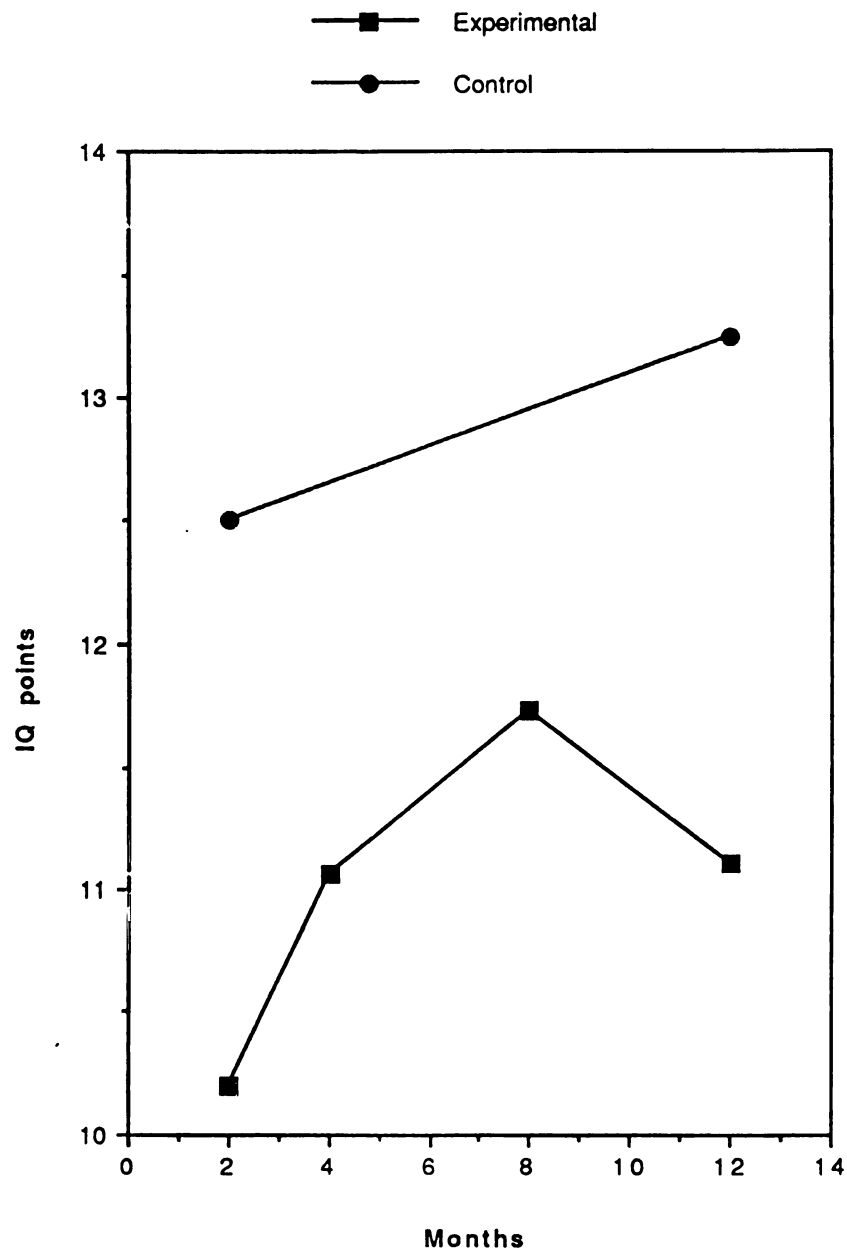


Figure 4.8. Changes in Performance IQ variance over time.

subjects' innate intellectual abilities premorbidly, the subjects current capabilities due to the limited amount of recovery that has taken place within a prescribed period of time since injury, memory and the ability to benefit from exposure to the test, as well as the limits and distribution of scores in the test itself. For these reasons, TBI subjects tested repeatedly may reach these thresholds quicker than TBI subjects tested less frequently and hence, the consistency among scores becomes greater. This, in turn, reduces variability and increases the homogeneity of the scores for the group tested more frequently.

The occasion for the subtests Comprehension and Picture Completion to yield significantly different ($p < .05$) means between the two groups with the considerable variability noted among all test scores certainly suggests the possibility of a test-retest practice effect. The fact that these two subtests may have generated test-retest practice effects is not unexpected. Comprehension and Picture Completion were among two of five subtests noted to demonstrate significant retest gains ($p < .10$) in the Ryan et al. (1985) study employing the WAIS-R. The other subtests were Similarities, Picture Arrangement, and Object Assembly. Previous studies employing only the WAIS also indicated Comprehension and Picture Completion as two subtests which demonstrated considerable retest gains which were thought to be attributable to test-retest practice

effects (Becker, 1977; Catron, 1978; Drudge et al., 1984; Matarazzo et al., 1980).

Possible explanations for test-retest practice effects in Picture Completion and Comprehension include the subjects honing or altering their responses to conform to a new level of expectancy, easily recalling their original response, or refining their earlier responses on subsequent testing (Catron, 1978). In addition, improvements in Picture Completion could also have arisen due to the recognition format of the task for which only one answer is possible or in which a quicker solution was made resulting in a larger score (Steisel, 1951).

In summary, the results of this study suggest that test-retest practice effects may be present in samples retested with the WAIS-R, however, these effects may only be evident when the WAIS-R is used several times. Unfortunately, individual variability may obscure many of these effects from one person to another. Nevertheless, when test-retest practice effects occur, they are likely to manifest themselves as improvements in the subtests Comprehension and Picture Completion when compared to similar groups of recovering TBI patients.

Hypothesis Five.

Hypothesis five, which tested whether an interaction existed between test-retest gains on the dependent variables over time and how often a test was administered, provided

mixed results. Earlier in this discussion it was reported that both the control and experimental groups demonstrated significant retest changes in a positive (gain) direction ($p < .05$).

When the differences between the two groups' test-retest changes were compared, a corresponding analysis of the interaction between the amount of treatment and time was made. Twelve of the fourteen test-retest gains in the experimental group appeared larger than the control group. The only exceptions were the subtests Vocabulary and Arithmetic. Of the twelve measures noted above, six were significantly larger in the experimental than the control group ($p < .05$) including Performance IQ. The remaining five measures demonstrating significant test-retest changes listed in order of decreasing effect sizes, included the subtests Picture Arrangement, Picture Completion, Comprehension, Object Assembly, and Similarities. Recall that these same subtests were also identified by Ryan et al. (1985) as demonstrating statistically significant test-retest gains on a single retest. This suggested the possibility that test-retest effects on the WAIS-R do exist with repeated test administrations.

In short, support for the existence of test-retest practice effects stems from several sources; most of which are largely conceptual in nature and require a logical rather than a statistical approach. Unfortunately, randomization of the groups was not possible, and the

effects of potentially confounding variables may not be evenly distributed across the two groups. While it is logical to assume that the retest gains in the experimental group which exceed the control groups gains (thought to reflect recovery) are mostly due to the sources of test-retest practice effects discussed below, other effects may be influencing the test-retest results that have not been controlled for either statistically and/or through randomization. Since it was not possible to unambiguously statistically model test-retest practice effects in this study, teasing out practice effects remains a primarily logical as opposed to a statistical procedure.

In line with this thinking, various authors (Catron, 1978; Lezak, 1983; Matarazzo, 1972; Steisel, 1951) have suggested that some tests that have a speed component may be particularly susceptible to test-retest practice effects. Catron (1978) and Steisel (1951) theorized that in a test-retest situation, the reaction times of a subject would tend to be faster in a subsequent test than at the initial testing, and therefore the subjects would earn bonus points for speed assuming those tasks were accurately completed. According to Catron (1978) this speed element contributed to increased retest scores in his study and may be one reason why all five Performance subtests in the experimental group had larger retest gains; three of which demonstrated statistically significant differences when compared to control group retest gains. While speed is not a consideration in

improvements noted in Comprehension or Similarities, other factors may contribute to increased scores as discussed below.

Karson et al. (1957) reported that tasks involving manipulation of test items might increase transfer from test to retest. In addition, Catron (1978) and Quereshi (1968) stated that development of a strategy or insight into a problem to be solved, such as the Gestalt needed for Object Assembly is another important facet. In this regard, a solution to a problem (which may have been discovered by trial and error); once grasped, or conceptualized, may result in putting the puzzle together as quickly as possible (Catron, 1978; Lezak, 1981; Steisel, 1951). Thus, from this line of logic it is not surprising that the significant gains reported in the Performance section occurred as they did.

Insight into the problem or development of a Gestalt may also have contributed to the significant differences between the two groups on the subtests Comprehension and Similarities. In addition, Lezak (1983) theorized that tests which require an unfamiliar mode of response may also show significant test-retest practice effects. In this manner, items on the subtest Similarities which require categorization of seemingly dissimilar concepts may result in the subject altering their responses to conform to a new set or level of expectancy which the subject may later understand more completely (Catron, 1978). Comprehension may also lend itself to test-retest practice effects through the recall of

originally correct responses, or the refinement of others at the second test which, in turn, yields additional points and increases scores on outcome measures. This may also be a result of improved recognition memory in the subjects as well as the repetitive nature of the tasks which may have assisted in the acquisition of that information. Additionally, some questions may be recalled at a later time because of their uniqueness or emotional impact which might further facilitate a search for a better answer or solution between test administrations.

Further support for the existence of test-retest practice effects stem from Warner's (1983) allegations that subtests revealing the greatest susceptibility to test-retest practice effects were those that were also the least reliable. It should not be surprising to note, then, that the subtests Object Assembly, Picture Arrangement, and Picture Completion had the lowest reliabilities in the Performance section, and Comprehension and Similarities had the next to the lowest reliabilities in the Verbal section using WAIS-R (Wechsler, 1981) data. This argument also seems to hold for the retest studies reviewed in Chapter Two.

A review of Chapter Two also indicates that the test-retest gains in the Performance section are also characteristically the largest for the subtests Picture Arrangement, Picture Completion, and Object Assembly in both the WAIS and the WAIS-R. The results are less consistent for

Comprehension and Similarities but they do reveal, overall, the largest test-retest gains in the Verbal section for both the WAIS and WAIS-R.

In summary, a significant interaction appears to exist which suggests that multiple testing over time does have an effect on outcome measures. Logically, these influences would appear to be largely attributable to test-retest practice effects. While the sample was not randomized and potentially confounding variables may be influencing these results, the use of a reasonable matched control group allows for the separation of test-retest practice effects from recovery on a logical basis. Using the hypotheses proposed by others above (Catron, 1978); Lezak, 1983; Steisel, 1951; Warner, 1983), the trend toward greater test-retest gains due to test-retest practice effects as a result of multiple test administrations appears evident and consistent with published data.

Secondary Hypothesis One.

Secondary hypothesis one had two purposes. First, to describe the test-retest reliabilities of the WAIS-R and second, to determine whether the test-retest reliabilities from pretest to posttest in the experimental group differed from those of the control group.

In general, Performance IQ test-retest reliabilities were smaller than the Verbal IQ test-retest reliabilities in both the experimental and control groups. In addition,

Verbal, Performance, and Full Scale IQ test-retest reliabilities in both groups appeared smaller than the data reported by Wechsler (1981) or others (Ryan et al., 1985; Snow et al., 1989; Warner, 1983) on the WAIS-R in normal and clinical (non-TBI) populations. This difference may be due to the TBI population used in this study, and since the WAIS-R was not normed on this population there is some uncertainty whether these values are expected. This same pattern of lower reliability coefficients is also true for reports of test-retest IQ reliabilities employing the WAIS, whom with the exception of Kangas and Bradway (1971), Klonoff et al. (1970), and Wagner and Caldwell (1979), had test-retest intervals of 676, 416, and 229 weeks apart. Each of these test-retest intervals is considerably longer than the test-retest interval of 44 to 45 weeks in this study, and supports the assertion by Cronbach (1960) and Freeman (1962) who stated that a longer test-retest interval will result in a lowering of the reliability coefficient.. A similar trend was also evident in the individual subtest test-retest reliabilities in both the experimental and control groups where the reliabilities appeared, overall, smaller than the reliabilities in the retest studies employing the WAIS-R noted above.

The weakest test-retest reliabilities in the experimental group, Picture Arrangement (.48) and Picture Completion (.51) also demonstrated the greatest test-retest gains. The subtests with the highest retest reliability in the

experimental group, Arithmetic (.84) and Vocabulary (.80), demonstrated the least amount of test-retest gain. This is consonant with Warner's (1983) earlier assertion regarding the inverse relationship between test-retest reliability and stability. That is, subtests that are the least susceptible to practice effects are those which are the most reliable and vice versa. Therefore, the more reliable the test, the greater the consistency of scores obtained by the same persons when re-examined with the same test on different occasions or under different examining conditions.

The Vocabulary test-retest reliability (.82) was the control group's largest and the subtest also revealed the least amount of test-retest gain. Picture Completion (.51) and Object Assembly (.53) demonstrated the lowest test-retest reliabilities in the control group but demonstrated the second and fourth largest retest gains on retest.

Test-retest reliabilities appeared larger than the experimental group for nine of the fourteen WAIS-R measures with only Arithmetic being significantly larger than the control group retest reliabilities. These apparently larger test-retest reliabilities in the experimental group may possibly reflect that group's more recent testing experience. Recall that the reported test-retest reliabilities were calculated between the second and twelfth month post-injury test dates, but that the experimental group had had two more exposures to the WAIS-R than the control group at 4 and 8 months post-injury, respectively. It may be possible,

that the higher test-retest reliability in the experimental group reflects the more recent influences of these additional tests. Since the subjects in the experimental group may remember some of their answers, this in turn could have resulted in a carryover, or test-retest practice effect, which would yield a higher retest correlation (Anastasi, 1976; Campbell, 1983; Cronbach, 1960; Derner et al., 1950; and Freeman, 1962). The additional testing experienced by the experimental group may have also diminished the effect that the much longer 10 1/2 month test-retest interval may have had in possibly lowering the reliability coefficients in the control group. It is also possible that the small sample sizes may be under or over estimating test-retest reliabilities.

Of the six WAIS-R measures for which significant interactions were found three measures (Similarities, Object Assembly and Performance IQ) had larger test-retest reliabilities in the experimental group. One measure (Picture Completion) had the same retest reliability coefficient for both groups, while the remaining two measures (Comprehension and Picture Arrangement) had test-retest reliability coefficients that were lower in the experimental group. None of these six reliability coefficients, however, were statistically different ($p > .05$) between the two groups.

While the increased test-retest reliability in the former three measures may be due to the increased

test-retest familiarity discussed above, it is not clear why the test-retest reliabilities in the latter three were the same or weaker than the control group. One possible explanation, at least for Picture Completion, is that it is a highly unreliable measure regardless of the presence or lack of influences which are exerted upon it. As a result, when long test-retest intervals apply, it remains unreliable regardless of the level of familiarity a subject may have with it. Similar arguments could also be made for Comprehension and Picture Arrangement.

Another possible explanation, however, for the lower retest reliabilities may be due to the differential influences that practice effects have on individuals within a specific population as reported by Campbell (1983), Catron (1978), and Meer and Baker (1967). Accordingly, if test-retest effects differentially influence test scores due to certain factors (e.g., differences in intelligence level, cognitive tempo, ceiling effects, fatigue, boredom, memory or similarity of response), rather than to increase each individual score to the same degree, then the overall test-retest reliability coefficient will decline. This is because the original rank order of individuals on the first test will change. Another possible explanation is that some individuals (i.e., brighter subjects) may tend to gain less on retesting due to homogenization effects which, in turn, may also affect their ranking from the first test. Thus,

the reliability coefficient would decrease due to the restricted variability (Catron, 1978; Stanley, 1971).

In summary, test-retest reliabilities in TBI populations employing the WAIS-R demonstrate moderate to very high reliability coefficients (range .51 to .82) when tested only twice with a test-retest interval of 10 1/2 months. Test-retest reliabilities in a similar group tested twice as often also demonstrated moderate to very good reliability coefficients (range .48 to .84), however, the group tested more often appeared to reveal higher reliabilities in nine of fourteen WAIS-R measures. While only the subtest Arithmetic demonstrated a statistically larger test-retest reliability between the two groups, the trend noted above was suggestive of the possible influences due to test-retest practice effects.

Secondary Hypothesis Two.

Secondary hypothesis two tested whether the cumulative effects of test-retest gains and recovery would be quadratic over a one year period of time post-injury. In response to this question, both a linear and a quadratic trend were noted in the Verbal, Performance, and Full Scale IQ means.

The changes noted from pretest to posttest generally follow the characteristics discussed previously in this section. Specifically, all IQ means were impaired upon initial testing with the Performance IQ mean appearing weaker than the Verbal IQ mean. Subsequently, all IQ means

increased to average levels of functioning with the largest gains or improvements noted in the Performance section. Additionally, the Performance IQ improvements or retest gains exceeded those gains noted in Verbal IQ at posttest. The rationale for these observations have also been reviewed previously, however, it was of particular interest to determine the trend and, therefore, the pattern of recovery in the IQ means when subjects were tested more than twice with the WAIS-R.

Consistent with the recovery of intellectual deficits reported in studies of TBI survivors, a rather characteristic pattern was noted; namely, that the return of intellectual abilities was rapid and the final level appeared to be reached earlier for the Verbal section than the Performance section (Bond & Brooks, 1975; Diller & Ben-Yishay, 1983). In addition, this period of rapid improvement was followed by a period of decelerating rate of improvement until an asymptote was reached suggesting a quadratic trend in the recovery pattern as noted by others (Dodrill & Troupin, 1975; Eson et al., 1978).

This description of the pattern of recovery appeared to be more congruent for verbal intellectual skills measured by the Verbal IQ than for the nonverbal intellectual skills as measured by Performance IQ. Verbal IQ means over the four test times demonstrated a linear trend effect followed by a smaller quadratic trend effect. Data and graphs also demonstrated a slow-down in retest gains with actual

decreases noted in test scores between the eighth and the twelfth month post-injury. This pattern was generally consistent with the study of Mandleberg and Brooks (1975) who reported a period of approximately six months for Verbal IQ with the WAIS to maximize recovery. Differences in the tests used, severity of injury, subject selection, and test-retest intervals interfere with further comparisons.

The pattern of recovery in Performance IQ means appeared to vary from that of the Verbal IQ means inasmuch as Performance IQ did not demonstrate an obvious plateau effect prior to the last test administration. Although Performance IQ means revealed a linear and quadratic trend effect over time, a graph of the Performance IQ means demonstrated continued increases in mean values through the twelfth month post-injury. The Performance IQ recovery curve appeared to be consistent with Mandleberg and Brooks (1975) who reported that WAIS Performance IQ recovery took longer than Verbal IQ to occur. However, this study was not extended long enough to determine whether Performance IQ recovery plateaus at approximately thirteen months as suggested by the Mandleberg and Brooks study (1975). Unfortunately, without further testing it will not be known at which point Performance IQ means in this study would have actually plateaued with the WAIS-R.

Noting the Performance IQ recovery curve and the inflection upward from the eighth month to the twelfth month test administration, lends itself to speculation. Since

both Performance and Verbal IQ means were relatively equal and within the normal range of intelligence at eight months post-injury, it is unclear why the Performance IQ mean would continue to increase while Verbal IQ means remained generally static.

One possible explanation maybe the fact that the majority of experimental subjects had greater Performance versus Verbal IQ abilities premorbidly and the inflection of Performance IQ means after eight months simply demonstrated a return to those higher premorbid levels. Unfortunately, this study lacked the information that may have lent support to this argument. Another explanation might be the existence of greater left versus right hemisphere damage in survivors who, because of the laterality of their injuries, would have demonstrated limited recovery in Verbal rather than Performance IQ abilities. In fact, just the opposite of this was true, suggesting that the greater number of patients in this study with right or bilateral hemisphere damage should have demonstrated greater limitations in Performance IQ abilities relative to Verbal IQ skills. The most plausible explanation, however, may be the fact that because of the increased familiarity of the test due to multiple test administrations, and hence, practice effects, the tendency to anticipate strategies, and improved times on speed items, subjects earned more points and subsequently higher scores on the Performance section of the WAIS-R. This is in contrast to the Verbal section of the WAIS-R

which, with the exception of Comprehension and Similarities discussed previously, may have demonstrated a homogenization effect since there may not have been additional allowances to take advantage of (i.e., bonus points on timed items) on later test administrations.

These observations raise a concern in interpreting IQ scores, particularly Performance IQ, since the existence of test-retest practice effects may result in judging more people to be normal on successive administrations of a test particularly when cut-off points are used as the sole basis for making judgments (Dodrill and Troupin, 1975; Matarazzo et al., 1976). Additional caution in interpretation was reported by Dikman et al. (1983) who suggested that test-retest practice effects may possibly mask the true slowing of recovery over time, since test-retest practice effects would have the effect of emulating continued cognitive improvement by influencing scores to increase. For these reasons, the possible existence of test-retest practice effects requires that considerable caution be exercised before concluding that subjects have returned to normal, or are continuing to demonstrate substantial recovery, without corroborative, behavioral evidence even if IQs and subtest scaled scores are back to average and/or premorbid levels (Becker, 1977; Dodrill and Troupin, 1975; Matarazzo et al., 1980; Matarazzo and Herman, 1984).

Secondary Hypothesis Three.

Secondary hypothesis three tested the internal consistency of the WAIS-R in the experimental group and whether the internal consistency changed over time. As noted in Chapter Four, the estimates of internal consistency demonstrated only modest reliability among the subtest performances. In addition, it was noted that these median correlations demonstrated a decrease over time. The possible explanations for this trend may be similar to those discussed previously. That is, these correlation coefficients may have decreased due to a restriction of range as a result of ceiling or homogenization effects that were created from a combination of recovery and test-retest practice effects (Catron, 1978; Stanley, 1971). Freeman (1962) also reported that the test-retest method is more likely to underestimate the internal consistency of a test, because factors extraneous to it may effect the scores dissimilarly. These same factors may continue to effect the internal consistency of a test through increasing dissimilarity in scores. As noted previously, these extraneous factors might include subjects' cognitive tempo, emotional experiences, mental growth, and boredom, etc. (Catron, 1978).

Implications for Future Research

This study, despite its imperfections, is one of only a handful of studies which have attempted to measure the

test-retest reliability and stability of the WAIS-R. In addition, this study attempted to provide a basis upon which descriptive information could be utilized by other clinicians regarding the intellectual outcome and recovery that takes place over time following TBI (Adams and Putnam, 1989). In addition, this study was original in its attempt to partition test-retest practice effects from recovery in repeat test administrations using impaired controls.

Clearly though, this study had many limitations not restricted simply to a lack of randomization. Future research in this area should therefore, continue to employ matched control groups of similarly impaired populations but with randomized groups. Concurrently, attempts should be made to assess the degree to which other factors, not accounted for here, influence practice effects. Examples would be pre- and posttest evaluations of memory and learning as well as affective-emotive conditions. Greater control over the effects of situational and environmental variables such as the time of testing, attitudes toward testing, emotional sequelae and the testing environment, etc. should also be attempted and explored. A prospective study of this nature may be prohibitive due to costs, however, such a study would allow for greater opportunities to collect demographic data on a host of variables that might also influence outcome and recovery.

Additional studies should also be completed in individuals with other levels of severity (e.g., minor or

profound TBI) to assess the degree of recovery and possible effects of practice on outcome measures. To this end, subsequent testing to account for cognitive changes beyond one year post-injury also appears warranted.

In addition, practice effects and their possible influences may be demonstrated through varying the number of times tests are administered within the same time periods. For example, a third group having been tested three times in addition to the two or four times in this study may have shed additional light on the proportion of test-retest gains that exist based on scaled degrees of multiple testing.

Correlational studies also appear warranted in order to look more specifically at those variables that influence both outcome and possible test-retest practice effects. Multiple regression methods utilizing such data may also provide a basis on which to make judgments regarding outcome, and the data derived from these formulas may help the clinician to differentiate recovery from other influences such as test-retest practice effects in later stages of recovery.

Conclusions

Based upon the above noted results, the following conclusions are suggested:

1. Intellectual measures of TBI survivors in the initial stage of recovery (two months post-injury) are significantly impaired, with

Performance IQ measures usually weaker than Verbal IQ measures on the WAIS-R.

2. Intellectual recovery as measured by the WAIS-R demonstrates significant gains between the initial test in the early phase of recovery (two months post-injury) and testing in the later phase of recovery (one year post-injury).
3. Improvements in WAIS-R Verbal, Performance, and Full Scale IQ values range, on average, from 8 points in Verbal IQ to 14 points in Performance IQ on a single retest using the WAIS-R with a test-retest interval of 10 1/2 months. Test-retest improvements appear to be larger, on average, for patients who were tested more frequently within this same interval of time.
4. WAIS-R measures tend to return to average levels of intelligence, with Performance measures exceeding Verbal measures upon a single retest at one year post-injury. The Performance-Verbal IQ discrepancy may, however, become somewhat larger with additional testings with the WAIS-R within the same test-retest interval.
5. Comparisons on the WAIS-R between similarly matched groups of TBI survivors tested within the same period of time post-injury, but differing on the frequency of testing, may not

demonstrate significant differences on either IQ or subtest scaled scores due to the large variability in scores at one year post-injury. Possible exceptions may be the subtests Comprehension and Picture Completion whose scaled scores may increase due to the influences of test-retest practice as a result of multiple testings.

6. Significant interactions are noted between select WAIS-R IQ and subtest outcome measures and how often the test is re-administered within a specified period of time. Specifically, and in decreasing order, Picture Arrangement, Picture Completion, Performance IQ, Comprehension, Object Assembly and Similarities are particularly susceptible to increased test-retest gains with multiple administrations of the WAIS-R. These increases appear to be attributable to test-retest practice effects.
7. Test-retest gains on the WAIS-R in recovering TBI populations are likely to reflect a combination of: a) test-retest practice effects; b) general recovery and; c) a combination of other influences heretofore unexplained.
8. Test-retest reliabilities of the WAIS-R subtests on a single retest range from .51 to

.82 suggesting moderate to good reliability within a 10 1/2 month period of time.

9. Verbal retest reliabilities appeared larger than Performance retest reliabilities. Measures having the weakest reliabilities also demonstrated, in general, larger retest gains. Test-retest reliabilities tend, on average, to be slightly larger when the WAIS-R is administered twice as often within the same retest interval. However, the range of reliabilities remains generally the same (i.e., .48 to .84)
10. The recovery of intellectual function on the WAIS-R as measured by IQ means appears to follow established patterns observed on the WAIS with rapid, initial recovery resulting in a subsequent, decelerating rate of improvement. While this generally holds for Verbal IQ, Performance IQ may not demonstrate the same amount of slow-up as a result of multiple testing and the influences of test-retest practice effects.
11. Interpretation of WAIS-R scores upon its readministration and the subsequent gains from test to retest must be made with caution, particularly when the WAIS-R has been administered several times within a one year

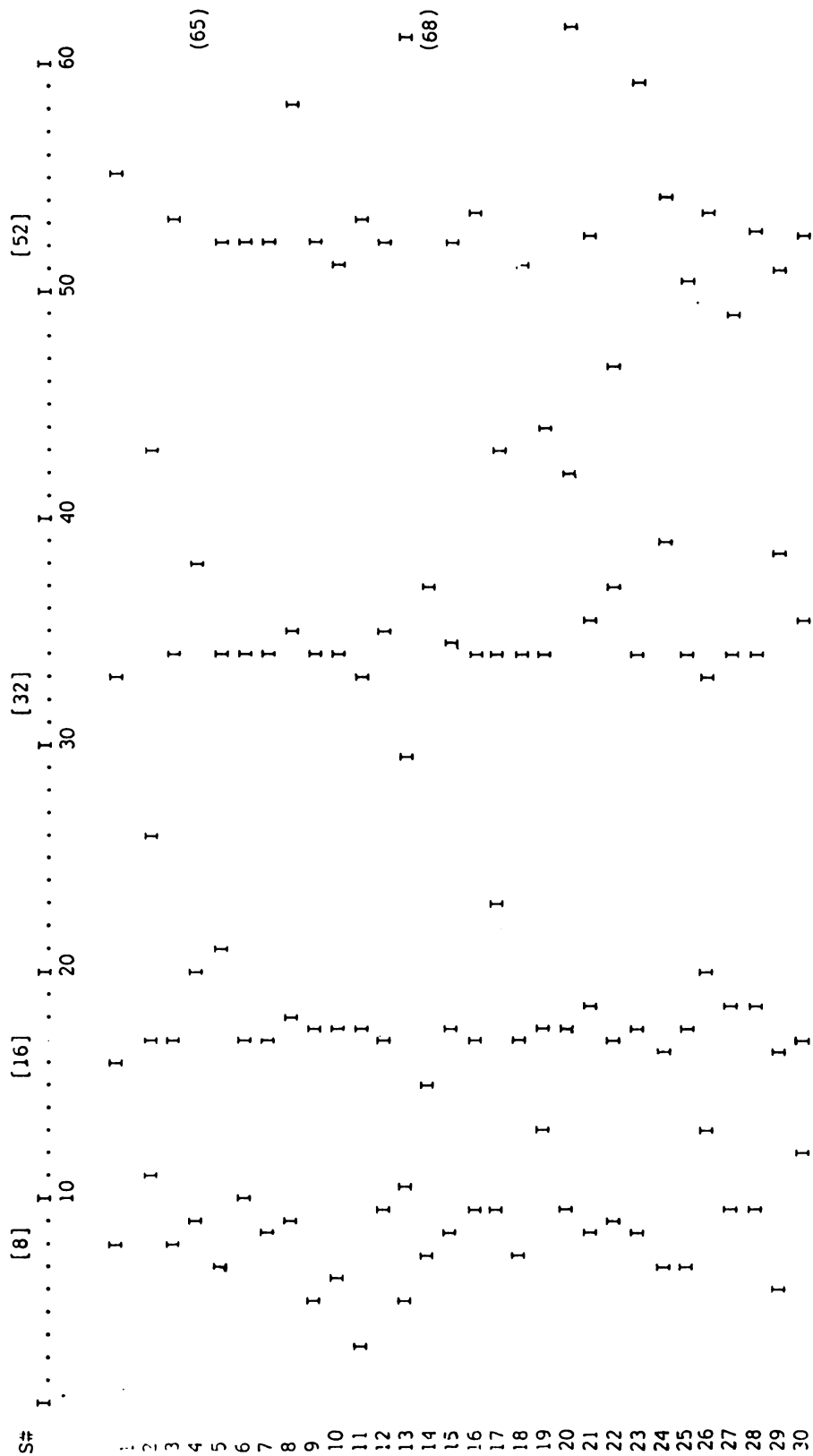
period of time. Without collaborative evidence, substantial gains may be due, in part, to retest practice effects which inflate scores, and mask the true slowing of recovery.

12. The internal consistency of the WAIS-R using global estimates demonstrate modest consistency among the subtest performances.

APPENDIX A

APPROXIMATE TEST TIMES POST-INJURY

Experimental Group Test Times Post-injury
(in weeks)



Experimental Group (continued)

S#	[8]	[16]	[32]	[52]	[60]
31	I	I	I	I	I
32	I	I	I	I	I
33	I	I	I	I	I
34	I	I	I	I	I
35	I	I	I	I	I
36	I	I	I	I	I
37	I	I	I	I	I
38	I	I	I	I	I
39	I	I	I	I	I
40	I	I	I	I	I
41	I	I	I	I	I
42	I	I	I	I	I
43	I	I	I	I	I
44	I	I	I	I	I
45	I	I	I	I	I
46	I	I	I	I	I
47	I	I	I	I	I

211

(77)

APPENDIX B

DATA SUMMARY FORM

Anoxia:	Yes	_____
	No	_____
Meningitis:	Yes	_____
	No	_____
Medication induced coma:	Yes	_____
	No	_____
Anticonvulsants:	Yes	_____
	No	_____
Psychotropic medications:	Yes	_____
	No	_____

DATA SUMMARY FORM

Type of Injury:		Surgical Intervention:	Yes	
Nonpenetrating	_____	(Cranial)	No	_____
No Skull Fracture	_____			
Skull Fracture	_____	CT scan (initial):	None	_____
Penetrating	_____		Right	_____
			Left	_____
Seizures:			Bilateral	_____
None	_____		Normal	_____
Pre-injury	_____		Intraventric	_____
(Epilepsy)	_____			
Post-Injury	_____	Handedness:	Right	_____
Both	_____		Left	_____
Unknown	_____		Ambidextrous	_____
Sig. Previous Head Injury:				
Yes:	_____			
No:	_____			
(PTA 24 hours, or GCS of 8				
or less, or hosp. observation				
24 hours)				

EVALUATION SCHEDULE

Date: _____ (Initial/2 Month)
 _____ (4 months)
 _____ (8 months)
 _____ (12 months)

Test Results:	Raw				
	Scores/Scaled				
Date:	_____	_____	_____	_____	_____
Time Postinjury:	_____	_____	_____	_____	_____
Information	____/____	____/____	____/____	____/____	____/____
D. Span	____/____	____/____	____/____	____/____	____/____
Vocabulary	____/____	____/____	____/____	____/____	____/____
Arithmetic	____/____	____/____	____/____	____/____	____/____
Comprehension	____/____	____/____	____/____	____/____	____/____
Similarities	____/____	____/____	____/____	____/____	____/____
P.C.	____/____	____/____	____/____	____/____	____/____
P.A.	____/____	____/____	____/____	____/____	____/____
B.D.	____/____	____/____	____/____	____/____	____/____
O.A.	____/____	____/____	____/____	____/____	____/____
D. Sym	____/____	____/____	____/____	____/____	____/____
Verbal IQ	____/____	____/____	____/____	____/____	____/____
Performance IQ	____/____	____/____	____/____	____/____	____/____
Full Scale IQ	____/____	____/____	____/____	____/____	____/____

APPENDIX C

MICHIGAN STATE UNIVERSITY COMMITTEE ON RESEARCH
INVOLVING HUMAN SUBJECTS (UCRIHS)
LETTER OF APPROVAL

MICHIGAN STATE UNIVERSITY

UNIVERSITY COMMITTEE ON RESEARCH INVOLVING
HUMAN SUBJECTS (UCRIHS)
206 BERKEY HALL
(517) 353-9738

EAST LANSING • MICHIGAN • 48824-1111

October 18, 1988

IRB# 88-386

David B. Rawlings, M.A.
319 Kingswood Dr.
East Grand Rapids, MI 49506

Dear Mr. Rawlings:

Subject: "TEST-RETEST PRACTICE EFFECTS, RELIABILITY, AND
STABILITY OF THE WAIS-R IN RECOVERING TRAUMATICALLY
BRAIN INJURED SURVIVORS IRB# 88-386"

The above project is exempt from full UCRIHS review. I have reviewed the proposed research protocol and find that the rights and welfare of human subjects appear to be protected. You have approval to conduct the research.

You are reminded that UCRIHS approval is valid for one calendar year. If you plan to continue this project beyond one year, please make provisions for obtaining appropriate UCRIHS approval one month prior to October 18, 1989.

Any changes in procedures involving human subjects must be reviewed by the UCRIHS prior to initiation of the change. UCRIHS must also be notified promptly of any problems (unexpected side effects, complaints, etc.) involving human subjects during the course of the work.

Thank you for bringing this project to our attention. If we can be of any future help, please do not hesitate to let us know.

Sincerely,



John K. Hudzik, Ph.D.
Chair, UCRIHS

JKH/sar

cc: N. Crewe

APPENDIX D

INFORMED CONSENT LETTER

Dear

I am a doctoral candidate at Michigan State University completing my Ph.D. in Counseling Psychology. As part of the requirements for my degree, I am working on a research project which will investigate how test scores change as individuals recover from their traumatic brain injuries.

The project would not require any further testing or interviews, but rather, a simple review of how you progressed from your injury based on existing records. You would not be identified in any way, and the information contained in your records would be combined with information from others in order to determine the pattern of change that occurs over time during one's recovery.

Your involvement only requires your signed approval to obtain records for review. Your decision to participate in this study will not effect in any your treatment now or in the future. Your participation is voluntary, you may choose not to participate. However, the results of this study may be extremely helpful to professionals who help newly head injured patients in the future.

I hope you will take a minute to read over the attached materials and agree to help me by participating in this study. Please indicate your decision by signing and returning both the consent and authorization to release information forms in the enclosed stamped envelope.

Please allow me to thank you in advance for your consideration and prompt response.

Sincerely,

David B. Rawlings, M.A.
Psychologist, Limited License

Enclosures: Consent Form
Authorization to Release Information
Addressed Stamped Envelope

APPENDIX E

CONSENT FORM

CONSENT FORM

Information About: "Test-retest practice effects, reliability, and stability of the Wechsler Adult Intelligence Scale - Revised (WAIS-R) in recovering traumatically brain-injured survivors."

1. This study, upon its completion, will be submitted in partial fulfillment of requirements for the degree of Doctor of Philosophy at Michigan State University.
2. This study is designed to improve the accuracy of tests which are used in the diagnosis and treatment of individuals suffering from traumatic brain injury. Information from this research project will enable professionals to determine the influence of repeated testing on obtained test scores. This, in turn, will improve our ability to monitor recovery from head-injury as well as the effects of rehabilitation, drugs, or surgery on brain functioning.
3. This study does not require your active participation.
4. In order to monitor the recovery from your brain-injury you have been given the same test (WAIS-R) on at least two occasions. This test may have been given by different psychologists in various settings as ordered by your physician or rehabilitation case manager. In this regard, you are being asked to sign a form authorizing release of that test information for analysis.
5. If you agree to participate, the release form as well as the consent form must be signed by you (or where appropriate, your guardian) and returned in the enclosed envelope. The signed release form will then be sent to the respective psychologist(s) or agency(s).
6. Test results which are received from these sources noted above will be kept strictly confidential. Your identity as a member of this study will not be revealed in any way; including any published or oral presentation of the results of this study.
7. The benefits of participating in this research project are that the results obtained from the test information you provide will help make the tests more accurate and useful assessment instruments. These improvements will, in turn, improve our evaluation of the effectiveness of rehabilitation treatment and rehabilitation programs, and increase our accuracy in determining the rate and extent of recovery from brain injury.
8. Participation in this study is entirely voluntary. You may choose not to participate at all or you may withdraw from the study at any time without prejudice or affect on the treatment you receive now or in the future.

Consent Form
page two

9. If you have any questions about the procedures of the study please ask me for more information. The principal investigator, David B. Rawlings, may be reached by calling (517)349-5471 or by writing him c/o Grand River Psychological Services, 2176 Hamilton Road, Okemos, MI 48864. In addition, Nancy M. Crewe, Ph.D., Professor, Michigan State University, will be able to answer any questions you may have about this research project. She may be reached by calling (517)355-1824.

I, _____, certify that above written statements were read and understood fully by me, and therefore I consent to participate in this study.

Date

Signature (indicate if signed by guardian)

David B. Rawlings, M.A.
Psychologist - Limited License

Please indicate if you would like to receive project results. _____ Yes

_____ No

APPENDIX F

CLIENT INFORMATION RELEASE AUTHORIZATION

BIBLIOGRAPHY

BIBLIOGRAPHY

- Adams, K.M., & Putnam, S.H. (1989). The efficiency of multidisciplinary rehabilitation programs for the traumatically brain-injured: a case study retrospective pilot study. (A study completed for the Michigan Catastrophic Claims Association) Livonia, MI: Adams and Putnam.
- Akerlund, E. (1959). The late prognosis in severe head injuries. Acta. chir. Scandinav., 117, 275-277.
- Anastasi, A. (1968). Psychological Testing (3rd ed.). New York: Macmillan Company.
- Anastasi, A. (1976). Psychological Testing (4th ed.). New York: Macmillan Publishing Co.
- Bayley, N. (1957). Data on the growth of intelligence between 16 and 21 years as measured by the Wechsler-Bellevue Scale. Journal of Genetic Psychology, 90, 3-15.
- Becker, B. (1977). Intellectual changes after closed head injury. Journal of Clinical Psychology, 31,(2) 307-309.
- Bell, A. & Zubek, J.P. (1960). The effect of age on the intellectual performance of mental defectives. Journal of Gerontology, 15(3), 285-295.
- Benton, A.L. (1979). Behavioral Consequences of Closed Head Injury (Central Nervous System Trauma Research Status Report). Washington, D.C.: National Institute of Neurological and Communicative Disorders and Stroke.
- Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C.W. Harris (Ed.), Problems in Measuring Change (pp 3-20). Madison: University of Wisconsin Press.
- Black, F.W. (1974). Cognitive effects of unilateral brain lesions secondary to penetrating missile wounds. Perceptual and Motor Skills, 38, 387-391.
- Bond, M.R. (1979). The stages of recovery from severe head injury with special reference to late outcome. International Rehabilitation Medicine, 1, (4), 155-159.

- Bond, M.R., & Brooks, D.N. (1976). Understanding the process of recovery as a basis for the investigation of rehabilitation for the brain injured. Scandinavian Journal of Rehabilitation Medicine, 8, 127-133.
- Bornstein, R.A. (1983). Verbal IQ-Performance IQ discrepancies on the Wechsler Adult Intelligence Scale-Revised in patients with unilateral or bilateral cerebral dysfunction. Journal of Consulting & Clinical Psychology, 51(5), 779-780.
- Bricolo, A., Turazzi, S., & Feriotti, G. (1980). Prolonged post-traumatic unconsciousness. Journal of Neurosurgery, 52, 625-634.
- Brooks, D.N. (1974). Recognition memory, and head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 37, 794-801.
- Brooks, D.N. (1975). Long and short term memory in head injured patients. Cortex, 11, 329-340.
- Brooks, D.N. (1976). Wechsler Memory Scale performance and its relationship to brain damage after severe closed head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 39, 593-601.
- Brooks, D.N., Aughton, M.E., Bond, M.R., Jones, P., & Rizvi, S. (1980). Cognitive sequelae in relationship to early indices of severity of brain damage after severe blunt head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 43, 529-534.
- Brooks, D.N., & Aughton, M.E. (1979). Cognitive recovery during the first year after severe blunt head injury. International Rehabilitation Medicine, 1, (4), 166-172.
- Brooks, D.N., & Aughton, M.E. (1979). Psychological consequences of blunt head injury. International Rehabilitation Medicine, 1, (4) 160-165.
- Brooks, D.N., Deelman, B.G., van Zomeran, A.H., van Dongen, H., van Harskamp, F., & Aughton, M.E. (1984). Problems in measuring cognitive recovery after acute brain injury. Journal of Clinical Neuropsychology, 6(1), 71-85.
- Brooks, D.N. Measuring neuropsychological and functional recovery. In H.S. Levin, J. Grafman, & H.M. Eisenberg (Eds.). Neurobehavioral Recovery From Head Injury, (pp. 57-72). New York: Oxford University Press.

- Brooks, N. (1986). Cognitive Outcome. Paper presented at the 10th Annual Post-graduate Course on Rehabilitation of the Brain-injured Adult and Child, Williamsburg, VA.
- Brooks, N. (1986). Recovery and prediction of cognitive outcome. Research Forum, 1(2), 1-3.
- Brooks, N. (1986). Methodological problems in research. Paper presented at the 10th Annual Post-graduate Course on Rehabilitation of the Brain-injured Adult and Child, Williamsburg, VA.
- Brooks, N. (Ed.). (1984). Closed head injury: psychological, social, and family consequences. Oxford: Oxford University Press.
- Brown, H.S.R., & May, A.E. (1979). A test-retest reliability study of the Wechsler Adult Intelligence Scale. Journal of Consulting and Clinical Psychology, 47(3), 601-602.
- Brown, J.C. (1975). Late recovery from head injury: case report and review. Psychological Medicine, 5, 239-248.
- Campbell, B.R. (1983). Reliability and practice effects on the Luria-Nebraska Neuropsychological Battery and the Revised Wechsler Memory Scale. Dissertation Abstracts International, 44(4-B), 1230.
- Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally College Publishing Company.
- Carlsson, C., von Essen, C., & Lofgren, J. (1968). Factors affecting the clinical course of patients with severe head injuries. Journal of Neurosurgery, 29 (September), 242-251.
- Catron, D.W. (1978). Immediate test-retest changes in WAIS Scores Among College Males. Psychological Reports, 43, 279-290.
- Catron, D.W., & Thompson, C.C. (1979). Test-retest gains in WAIS scores after four retest intervals. Journal of Clinical Psychology, 35(2), 352-357.
- Clarke, J., & Haughton, H. (1975). A study of intellectual impairment and recovery rates in heavy drinkers in Ireland. British Journal of Psychiatry, 126, 178-84.
- Conkey, R.C. (1938). Psychological changes associated with head injuries. Archives of Psychology, 232, 1-62.

- Coons, W.H. & Peacock, E.P. (1959). Inter-examiner reliability of the Wechsler Adult Intelligence Scale with mental hospital patients. Ontario Psychological Association Quarterly, 12, 33-37.
- Cope, D.H. (1985). Traumatic closed head injury: status of rehabilitation treatment. Seminars in Neurology, 5(3), 212-220.
- Crockett, D., Clark, C., & Klonoff, H. (1981). Introduction-An overview of neuropsychology. In S.D. Filskov & T.J. Boll (Eds.), Handbook of clinical neuropsychology. (pp. 1-37). New York: J. Wiley.
- Cronbach, L.J. (1960). Essentials of Psychological Testing (2nd ed.). New York: Harper and Brothers.
- Cullum, C.M., & Bigler, E.D. (1986). Ventricle size, cortical atrophy and the relationship with neuropsychological status in closed head injury: a quantitative analysis. Journal of Clinical and Experimental Neuropsychology, 8(4), 437-452.
- Derner, G.F., Aborn, M., & Canter, A.H. (1950). The reliability of the Wechsler-Bellevue subtests and scales. Journal of Consulting Psychology, 14, 172-179.
- Dikmen, S., Reitan, R.M., & Temkin, N.R. (1983). Neuropsychological recovery in head injury. Archives of Neurology, 40, 333-338.
- Diller, L., & Ben-Yishay, Y. (1983). Severe head trauma: a comprehensive medical approach to rehabilitation. (Report - IVIHR Grant No. 13-P-59082). Washington, D.C.: U.S. Department of Education.
- Dinning, W.D., Andert, J.N., & Hustak, T.L. (1977). Reliability and stability of WAIS IQs for institutionalized adult retardates. Psychological Reports, 40, 929-930.
- Dodrill, C.B., & Troupin, A.S. (1975). Effects of repeated administrations of a comprehensive neuropsychological battery among chronic epileptics. The Journal of Nervous and Mental Disease, 161(3), 185-190.
- Drudge, O.W., Williams, J.M., Kessler, M., & Gomes, F.B. (1984). Recovery from severe closed head injuries: repeat testings with the Halstead-Reitan neuropsychological battery. Journal of Clinical Psychology, 40(1), 259-265.

- Duke, R.B., Bloor, B.M., Nugent, G.R., & Majzoub, H.S. (1968). Changes in performance on WAIS, Trail Making Test and Finger Tapping Test associated with carotid artery surgery. Perceptual and Motor Skills, 26, 399-404.
- Dye, O.A., Milby, J.B., & Saxon, S.A. (1979). Effects of early neurological problems following head trauma on subsequent neuropsychological performance. Acta Neurologica Scandinavica, 59, 10-14.
- Dye, O.A., Saxon, S.A., & Milby, J.B. (1981). Long-term neuropsychological deficits after traumatic head injury with comatosis. Journal of Clinical Psychology, 37 (3), 472-477.
- Eisdorfer, C. (1963). The WAIS Performance of the aged: a retest evaluation. Journal of Gerontology, 1963, 18, 169-172.
- Eliason, M.R., & Topp, B.W. (1984). Predictive validity of Rappaport's Disability Rating Scale in subjects with acute brain dysfunction. Archives of Physical Medicine and Rehabilitation, 64 (2), 1357-1360.
- Eson, M.E., Yen, J.K., & Bourke, R.S. (1978). Assessment of recovery from serious head injury. Journal of Neurology, Neurosurgery and Psychiatry, 41, 1036-1042.
- Finlayson, M., & Block, R. (1982). Cognitive and emotional changes in head injury: a follow-up study. Paper presented at the annual meeting of the International Neuropsychological Society, Italy.
- Finlayson, M.A., Johnson, K.A., Reitan, R.M. (1977). Relationship of level of education to neuropsychological measures in brain-damaged and non-brain-damaged adults. Journal of Consulting and Clinical Psychology, 45(4), 536-542.
- Fisher, J.M. (1985). Cognitive and behavioral consequences of closed head injury. Seminars in Neurology, 5(3), 197-204.
- Forer, S. (1985) Rehabilitation outcomes and evaluation systems for traumatic brain injury. Paper presented at the Eighth Annual Head Trauma Conference: Coma to Community, San Jose, CA.
- Freeman, F. (1962). Theory and Practice of Psychological Testing (3rd ed.). New York: Holt, Rinehart, Winston.

- Fuld, P.A. (1984). Neuropsychological testing playing an increasing role in diagnosis. Generations, Winter, 52-53.
- Furst, C. (1984). The neuropsychological evaluation. National Head Injury Foundation (B0484).
- Glass, G.V., & Hopkins, K.D. (1984). Statistical methods in education and psychology (2nd ed.). New Jersey: Prentice-Hall, Inc.
- Golden, C.J. (1976). The value of neuropsychological testing to the physician. South Dakota Journal of Medicine, 29(9).
- Goldstein, S.G., Kleinknecht, R.A., & Gallo, A.E. (1970). Neuropsychological changes associated with carotid endarterectomy. Cortex, 6, 308-322.
- Grafman, J., Salazar, A., Weingartner, H., Vance, S., & Amin, D. (1986). The relationship of brain-tissue loss volume and lesion location to cognitive deficit. The Journal of Neuroscience, 6(2), 301-307.
- Groswasser, Z., Mendelson, L., Stern, M.J., Schecter, I., & Najenson, T. (1977). Re-evaluation of prognostic factors in rehabilitation after severe head injury. Scandinavian Journal of Rehabilitation Medicine, 9, 147-149.
- Hartlage, L.C., & Telzrow, C.F. (1980). The practice of clinical neuropsychology in the U.S. Clinical Neuropsychology, 2(4), 200-202.
- Heaton, R.K., Baade, L.E., & Johnson, K.L. (1978). Neuropsychological test results associated with psychiatric disorders in adults. Psychological Bulletin, 85(1), 141-162.
- Heiskanen, O. & Sipponen, P. (1970). Prognosis of severe brain injury. Acta Neurologica Scandinavia, 46, 343-348.
- Himelstein, P. (1957). A comparison of two methods of estimating Full Scale IQ from an abbreviated WAIS. Journal of Consulting Psychology, 21(3), 246.
- Ivnik, R.J. (1978). Neuropsychological stability in multiple sclerosis. Journal of Consulting and Clinical Psychology, 46(5), 913-923.
- Jarvik, L.F. & Falek, A. (1963). Intellectual stability and survival in the aged. Journal of Gerontology, 18, 173-176.

- Jennett, B. (1972). Some aspects of prognosis after severe head injury. Scandinavian Journal of Rehabilitation Medicine, 4, 16-20.
- Kangas, J., & Bradway, K. (1971). Intelligence at middle age: a thirty-eight year follow-up. Developmental Psychology, 5(2), 333-337.
- Kaplan, E. (1988). Neuropsychological Assessment. Paper presented at the Ninth Cape Cod Institute, Cape Cod, MA.
- Karson, S., Pool, K.B., Freud, S.L. (1957). The effects of scale and practice on WAIS and W-BI Test Scores. Journal of Consulting Psychology, 21 (3), 241-245.
- Karzmark, P., Heaton, R.K., Grant, I., and Matthews, C.G. (1985). Use of demographic variables to predict full scale IQ: a replication and extension. Journal of Clinical and Experimental Neuropsychology, 7, (4), 412-420.
- Kelly, M.P., Montgomery, M.L., Felleman, E.S., & Webb, W.W. (1984). Wechsler Adult Intelligence Scale and Wechsler Adult Intelligence Scale-Revised in a neurologically impaired population. Journal of Clinical Psychology, 40(3), 788-791.
- Kendrick, D.C., & Post, F. (1967). Differences in cognitive status between healthy, psychiatrically ill, and diffusely brain-damaged elderly subjects. British Journal of Psychiatry, 113, 75-81.
- King, G.D., Gideon, D.A., Haynes, C.D., Dempsey, R.L., & Jenkins, C.W. (1977). Intellectual and personality changes associated with carotid endarterectomy. Journal of Clinical Psychology, 33(1), 215-220.
- Kirk, R.E. (1982). Experimental Design (2nd Ed.). Belmont, CA: Brooks/Cole.
- Klonoff, H., Fibiger, C.H., & Hutton, G.H. (1970). Neuropsychological patterns in chronic schizophrenia. The Journal of Nervous and Mental Disease, 150(4), 291-300.
- Klove, H. & Cleeland, C.S. (1972). The relationship of neuropsychological impairment to other indices of severity of head injury. Scandinavian Journal of Rehabilitation Medicine, 4, 55-60.
- Knight, R.G. (1983). On interpreting the several standard errors of the WAIS-R: Some further tables. Journal of Consulting and Clinical Psychology, 51(5), 671-673.

- Knight, R.G., & Shelton, E.J. (1983). Tables for evaluating predicted retest changes in Wechsler Adult Intelligence Scale scores. British Journal of Clinical Psychology, 22, 77-81.
- Ladd, C.E. (1964). WAIS performances of brain damaged and neurotic patients. Journal of Clinical Psychology, 20, 115-117.
- Lawson, J.S., & Inglis, J. (1983). A laterality index of cognitive impairment after hemispheric damage: a measure from a principal-components analysis of the Wechsler Adult Intelligence Scale. Journal of Consulting and Clinical Psychology, 51(6), 832-840.
- Lawson, J.S., Inglis, J., & Stroud, T.W. (1983). A laterality index of cognitive impairment derived from a principal-components analysis of the WAIS-R. Journal of Consulting and Clinical Psychology, 51(6), 841-847.
- Leckliter, I.N., Matarazzo, J.D., & Silverstein, A.B. (1986). A literature review of factor analytic studies of the WAIS-R. Journal of Clinical Psychology, 42(2), 332-342.
- Lemke, E., & Wiersma, W. (1976). Principles of Psychological Measurement. Chicago: Rand McNally, 1976.
- Levin, H.S. (1985). Part II: Neurobehavioral Recovery (Central Nervous System Trauma Report). Washington, D.C.: National Institute of Neurological and Communicative Disorders and Stroke.
- Levin, H.S., Benton, A.L., & Grossman, R.G. (1982). Neurobehavioral Consequences of Closed Head Injury. New York: Oxford University Press.
- Levin, H.S., Grossman, R.G., & Kelly, P.J. (1976). Aphasic disorder in patients with closed head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 39, 1062-1070.
- Levin, H.S., Grossman, R.G., & Kelly, P.J. (1976). Short-term recognition memory in relation to severity of head injury. Cortex, 12, 175-182.
- Levin, H.S., Grossman, R.G., Rose, J.E., & Teasdale, G. (1979). Long-term neuropsychological outcome of closed head injury. Journal of Neurosurgery, 50, 412-422.
- Lewin, W., Marshall, T.F., & Roberts, A.H. (1979). Long-term outcome after severe head injury. British Medical Journal, 15 (Dec), 1533-1538.

- Lezak, M.D. (1979). Recovery of memory and learning functions following traumatic brain injury. Cortex, 15, 63-70.
- Lezak, M.D. (1983). Neuropsychological Assessment (2nd ed.). New York: Oxford Press.
- Lezak, M.D. & Gray, D.K. (1984). Sampling problems and nonparametric solutions in clinical neuropsychological research. Journal of Clinical Neuropsychology, 6, (1), 101-109.
- Lippold, S., & Claiborn, J.M. (1983). Comparison of the Wechsler Adult Intelligence Scale and the Wechsler Adult Intelligence Scale-Revised. Journal of Consulting and Clinical Psychology, 51(2), 315.
- Lundholm, J., Jepsen, B.N., & Thornval, G. (1975). The late neurological, psychological, and social aspects of severe traumatic coma. Scandinavian Journal of Rehabilitation Medicine, 7, 97-100.
- Malkumus, D., & Stenderup, K. (1974). Levels of Cognitive Functioning. (available from Rancho Los Amigos Hospital, Division of Neurological Services, Downey, CA).
- Mandleberg, I.A. (1975). Cognitive recovery after head injury 2. Wechsler Adult Intelligence Scale during post-traumatic amnesia. Journal of Neurology, Neurosurgery, and Psychiatry, 38, 1127-1132.
- Mandleberg, I.A. (1976). Cognitive recovery after severe head injury 3. WAIS verbal and performance IQs as a function of post-traumatic amnesia duration and time from injury. Journal of Neurology, Neurosurgery, and Psychiatry, 39, 1001-1007.
- Mandleberg, I.A., & Brooks, D.N. (1975). Cognitive recovery after severe head injury 1. serial testing on the Wechsler Adult Intelligence Scale. Journal of Neurology, Neurosurgery, and Psychiatry, 38, 1121-1126.
- Marascuilo, L.A., & Serlin, R.C. (1988). Statistical methods for the social and behavioral sciences. New York: W. H. Freeman & Co.
- Matarazzo, J.D. (1972). Wechsler's measurement and appraisal of adult intelligence (5th ed.). Baltimore: Williams & Wilkins.

- Matarazzo, J.D., & Herman, D.O. (1984). Base rate data for the WAIS-R: test-retest stability and VIQ-PIQ differences. Journal of Clinical Neuropsychology, 6(4), 351-366.
- Matarazzo, J.D., & Herman, D.O. (1984). Relationship of education and IQ in the WAIS-R Standardization Sample. Journal of Consulting and Clinical Psychology, 52(4), 631-634.
- Matarazzo, J.D., Carmody, T.P., & Jacobs, L.D. (1980). Test-retest reliability and stability of the WAIS: a literature review with implications for clinical practice. Journal of Clinical Neuropsychology, 2(2), 89-105.
- Matarazzo, J.D., Matarazzo, R.G., Wiens, A.N., Gallo & Klonoff, H. (1976). Retest reliability of the Halstead Impairment Index in a normal, a schizophrenic, and two samples of organic patients. Journal of Clinical Psychology, 32, 338-349.
- Matarazzo, R.G., Matarazzo, J.D., Gallo, A.E., & Wiens, A.N. (1979). IQ and neuropsychological changes following carotid endarterectomy. Journal of Clinical Neuropsychology, 1(2), 97-116.
- Matarazzo, R.G., Wiens, A.N., Matarazzo, J.D., & Manaugh, T.S. (1973). Test-retest reliability of the WAIS in a normal population. Journal of Clinical Psychology, 29, 194-197.
- McNemar, Q. (1957). On WAIS difference scores. Journal of Consulting Psychology, 21(3), 239-240.
- Meer, B., & Baker, J.A. (1967). Reliability of measurements of intellectual functioning of geriatric patients. Journal of Gerontology, 20, 410-414.
- Mehrens, W.A., & Lehmann, I.J. (1978). Measurement and evaluation in education and psychology (2nd ed.). New York: Holt, Rinehart, & Winston.
- Meir, B., & Baker, J.A. (1967). Reliability of measurements of intellectual functioning of geriatric patients. Journal of Gerontology, 20, 410-414.
- Miller, E. (1979). The long-term consequences of head injury: a discussion of the evidence with special reference to the preparation of legal reports. British Journal of Social and Clinical Psychology, 18, 87-98.

- Mittenberg, W. & Ryan, J.J. (1984). Effects of omitting one to five subtests on the WAIS-R full scale reliability. Perceptual and Motor Skills, 58, 563-565.
- Morrow, R.S., & Mark, J.C. (1955). The correlation of intelligence and neurological findings on twenty-two patients autopsied for brain damage. Journal of Consulting Psychology, 19(4), 283-289.
- Nagele, D.A., & Levine, M.J. (1983). Patterns of test performance on neuropsychological composite variables in patients with head injury. Paper presented at the 7th Annual Post-graduate Course on the Rehabilitation of the Brain Injured Adult, Williamsburg, VA.
- Naglieri, J.A. (1982). Two types of tables for use with the WAIS-R. Journal of Consulting and Clinical Psychology, 50(2), 319-321.
- Najenson, T., Groswasser, Z., Stern, M., Schechter, I., Daviv, C., Berghaus, N., & Mendelson, L. (1975). Prognostic factors in rehabilitation after severe head injury. Scandinavian Journal of Rehabilitation Medicine, 7, 101-105.
- Najenson, T., Mendelson, L., Schechter, I., David, C., Mintz, N., & Groswasser, Z. (1974). Rehabilitation after severe head injury. Scandinavian Journal of Rehabilitation Medicine, 5, 5-14.
- Newcombe, F. (1982). The psychological consequences of closed head injury: assessment and rehabilitation. Injury, 14, 111-136.
- Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. (1975). Statistical package for the social sciences. New York: McGraw-Hill.
- Norrman, B., & Svahn, K. (1961). A follow-up study of severe brain injuries. Acta Psychiatrica Scandinavia, 37, 236-264.
- O'Shaughnessy, E.J., Fowler, R.S., & Reid, V. (1984). Sequelae of mild closed head injuries. The Journal of Family Practice, 18(3), 391-394.
- Ommaya, A.K., & Gennarelli, T.A. (1974). Cerebral concussion and traumatic unconsciousness. Brain, 97, 633-654.
- Panikoff, L.B. (1983). Recovery trends of functional skills in the head-injured adults. The American Journal of Occupational Therapy, 37(11), 735-743.

- Parsons, O.A. & Prigatano, G.P. (1978). Methodological considerations in clinical neuropsychological research. Journal of Consulting and Clinical Psychology, 46(4), 608-619.
- Prifitera, A., & Ryan, J.J. (1983). WAIS-R/WAIS comparisons in a clinical sample. Clinical Neuropsychology, 5(3), 97-99.
- Quereshi, M.Y. (1968). Practice effects on the WISC subtest scores and IQ measurements. Journal of Clinical Psychology, 4, 73-81.
- Quereshi, M.Y. (1968). The comparability of WAIS and WISC subtest scores and IQ estimates. The Journal of Psychology, 68, 73-82.
- Rabourn, R.E. (1983). The Wechsler Adult Intelligence Scale (WAIS) and the WAIS-Revised: a comparison and a caution. Professional Psychology: Research and Practice, 14(3), 357-361.
- Rappaport, M., Hall, K.M., Hopkins, K., Belleza, T., & Cope, D.N. (1982). Disability Rating Scale for Severe Head Trauma: Coma to Community. Archives of Physical Medicine and Rehabilitation, 63, 118-123.
- Reitan, R.M. & Davison, L.A. (1974). Clinical neuropsychology: current status and applications. Washington, D.C.: Winston.
- Reynolds, C.R., Willson, V.L., & Clark, P.L. (1983). A four-test short form of the WAIS-R for clinical screening. Clinical Neuropsychology, 5(3), 111-116.
- Rhudrick, P.J. & Gordon, C. The age center of New England study. In L. Jarvik, E. Eisendorfer, & J. Blum (Eds.), Intellectual functions in adults. New York: Springer, 1973.
- Rimel, R.W., & Jane, J.A. (1983). Characteristics of the head-injured patient. In M. Rosenthal, E.R. Griffith, M.R. Bond, & J.D. Miller (Eds.), Rehabilitation of the Head Injured Adult. Philadelphia: F.A. Davis.
- Robb, Bernardoni, & Johnson, (1972). Assessment of Individual Mental Ability. London: Intext Educational Publishers.
- Rosen, M., Stallings, L., Floor, L., & Nowakiwska, M. (1968). Reliability and stability of Wechsler IQ scores for institutionalized mental subnormals. American Journal of Mental Deficiency, 73, 218-225.

- Ruesch, J. (1944). Intellectual impairment in head injuries. American Journal of Psychiatry, 100, 480-496.
- Ruesch, J., & Moore, B.E. (1943). Measurement of intellectual functions in the acute stage of head injury. Archives of Neurology and Psychiatry, 50, 165-170.
- Rutter, M., Chadwick, O., Shaffer, D., & Brown, G. (1980). A prospective study of children with head injuries: I. design and methods. Psychological Medicine, 10, 633-645.
- Ryan, J.J., Georgemiller, R.J., Geisser, M.E., & Randall, D.M. (1985). Test-retest stability of the WAIS-R in a clinical sample. Journal of Clinical Psychology, 41(4), 552-556.
- Ryan, J.J., Prifitera, A., & Larsen, J. (1982). Reliability of the WAIS-R with a mixed patient sample. Perceptual and Motor Skills, 55, 1277-1278.
- Sattler, J.M. (1988). Assessment of Children (3rd ed.). San Diego: Jerome M. Sattler, Publisher.
- Schacter, D.L. & Crovitz, H.F. (1977). Memory function after closed head injury: a review of the quantitative research. Cortex, 13, 150-176.
- Schau, E.J., O'Leary, M.R., & Chaney, E.F. (1980). Reversibility of cognitive deficit in alcoholics. Journal of Studies of Alcohol, 41(7), 733-740.
- Seidenberg, M., O'Leary, D.S., Berent, S., & Boll, T. (1981). Changes in seizure frequency and test-retest scores on the Wechsler Adult Intelligence Scale. Epilepsia, 22, 75-83.
- Seidenberg, M., O'Leary, D.S., Giordani, B., Berent, S., & Boll, T. (1981). Test-retest IQ changes of epilepsy patients: assessing the influence of practice effects. Journal of Clinical Neuropsychology, 3(3), 237-255.
- Seretny, M.L., Dean, R.S., Gray, J.W., & Hartlage, L.C. (1986). The practice of clinical neuropsychology in the United States. Archives of Clinical Neuropsychology, 1, 5-12.
- Shatz, M.W. (1981). WAIS practice effects in clinical neuropsychology. Journal of Clinical Neuropsychology, 3(2), 171-179.

- Smith, A. (1961). Duration of impaired consciousness as an index of severity in closed head injuries: a review. Disease of the Nervous System, 22(2), 69-74.
- Smith, A. (1975). Neuropsychological Testing in Neurological Disorders. In W.J. Friedlander (Ed.). Advances in Neurology. Vol. 7.: Current Reviews of Higher Nervous System Dysfunction. New York: Raven Press.
- Smith, E. (1974). Influence of site of impact on cognitive impairment persisting long after severe closed head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 37, 719-726.
- Smith, R.S. (1983). A comparison study of the Wechsler Adult Intelligence Scale and the Wechsler Adult Intelligence Scale-Revised in a college population. Journal of Consulting and Clinical Psychology, 51(3), 414-419.
- Snow, W.G., Tierney, M.C., Zorzitto, M.L., Fisher, R.H., Reid, D.W. (1989). WAIS-R test-retest reliability in a normal elderly sample. Journal of Clinical and Experimental Neuropsychology, 11(4), 423-428.
- Spitz, H.H. (1983). Intratest and intertest reliability and stability of the WISC, WISC-R, and the WAIS full scale IQs in a mentally retarded populations. The Journal of Special Education, 17(1), 69-80.
- Stanley, J.C. (1971). Reliability. In R.L. Thorndike (Ed.), Educational Measurement (pp 356-417). Washington, D.C.: American Council on Education.
- Steisel, I.M. (1951). The relation between test and retest scores on the Wechsler-Bellevue Scale (Form 1) for selected college students. The Journal of Genetic Psychology, 79, 155-162.
- Tabaddor, K., Mattis, S., & Zazula, T. (1984). Cognitive sequelae and recovery course after moderate and severe head injury. Neurosurgery, 14(6), 701-708.
- Teasdale, G., Knill-Jones, R., & VanDerSande, J. (1978). Observer variability in assessing impaired consciousness and coma. Journal of Neurology, Neurosurgery, and Psychiatry, 41, 603-610.
- Teasdale, G., Skene, A., Parker, L., & Jennett, B. (1979). Age and outcome of severe head injury. Acta Neurochirurgica, Supplement, 28, 140-143.

- Timming, R., Orrison, W.W., & Mikula, J.A. (1982). Computerized tomography and rehabilitation outcome after severe head trauma. Archives of Physical Medicine and Rehabilitation, 63, 154-159.
- Tooth, G. (1946). On the use of mental tests for the measurement of disability after head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 10, 1-11.
- Troll, L.E., Saltz, R., & Dunin-Markiewicz, A. (1976). A seven-year follow-up of intelligence scores of foster grandparents. Journal of Gerontology, 31(5), 583-385.
- Tuma, J.M., & Appelbaum, A.S. (1980). Reliability and practice effects of WISC-R IQ estimates in a normal population. Educational and Psychological Measurement, 40, 671-678.
- Turner, D.F., & Dunn, K.P. (1980). Neuropsychological assessment in public psychiatric hospitals. Clinical Neuropsychology, 2(3), 129.
- Urbina, S.P., Golden, C.J., & Ariel, R.N. (1982). WAIS/WAIS-R: initial comparisons. Clinical Neuropsychology, 4(4), 145-146.
- VanZomeran, A.H., & Deelman, B.G. (1978). Long-term recovery of visual reaction time after closed head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 41, 452-457.
- Wagner, E.E., & Caldwell, M.S. (1979). WAIS test-retest reliability for a clinical out-patient sample. Perceptual and Motor Skills, 48, 131-137.
- Warner, M.H. (1983). Practice effects, test-retest reliability and comparability of WAIS and WAIS-R: Issues in the assessment of cognitive recovery in detoxified alcoholics. Dissertation Abstracts International, 44(8-B), 2572.
- Wechsler, D. (1939). The measurement of adult intelligence. Baltimore: Williams & Wilkins.
- Wechsler, D. (1955). Wechsler Adult Intelligence Scale Manual New York: The Psychological Corporation.
- Wechsler, D. (1981). Wechsler Adult Intelligence Scale - Revised Manual. New York: The Psychological Corporation.

- Williams, J.M., Gomes, F., Drudge, O.W., & Kessler, M. (1984). Predicting outcome from closed head injury by early assessment of trauma severity. Journal of Neurosurgery, 61, 581-585.
- Willson, V.L., & Reynolds, C.R. (1985). Normative data on the WAIS-R for Selz and Reitan's Index of Scatter. Journal of Clinical Psychology, 41 (2), 254-258.
- Winne, J.F. (1974). Test-retest reliability of the Wechsler battery as a measure of PAS dimensions. Journal of Clinical Psychology, 30(3), 335-340.
- Wolfe, G.R. (1987). Clinical neuropsychology and assessment of brain impairment: an overview. Cognitive Rehabilitation, Sept./Oct., 20-25.
- Zimmerman, I. L., & Woo-Sam, J.M. (1973). Clinical interpretation of the Wechsler Adult Intelligence Scale. New York: Grune & Stratton.

MICHIGAN STATE UNIV. LIBRARIES



31293008918413