5-1992

A Cognitive Process Approach to Interpreting Performance on the Booklet Category Test and the Wisconsin Card Sorting Test

Phillip R. Wolfe
Utah State University

A COGNITIVE PROCESS APPROACH TO INTERPRETING PERFORMANCE ON THE BOOKLET CATEGORY TEST AND THE WISCONSIN CARD SORTING TEST

by

Phillip R. Wolfe

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1992

ACKNOWLEDGMENTS

I am indebted to the many brain-injured patients whose successes and frustrations in performing problem-solving tasks have taught me much about brain function and human behavior. Without those patients, this study would not have been possible.

I am indebted also to my friend and former wife whose support and encouragement throughout my graduate education has helped me complete this study, and who listened patiently for many hours to my "thinking aloud" about this project.

Phillip R. Wolfe

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

ABSTRACT

A Cognitive Process Approach to Interpreting Performance on the Booklet Category Test and the Wisconsin Card Sorting Test

by

Phillip R. Wolfe, Doctor of Philosophy

Utah State University, 1992

Major Professor: Michael R. Bertoch, Ed.D.
Department: Psychology

Modified administration techniques that relied on patient verbalization of reasoning on each item were devised. For the WCST, verbalized scores correlated highly with conventional scores. However, patterns of age, education, and IQ covariates for each scoring condition were very different, raising questions concerning what such verbalized scores measured. Further research based upon a prospective research design was suggested to address this question. Factor analysis of WCST scores for each scoring condition resulted in almost identical three-factor solutions in each case: (a) ineffective, perseverative responding; (b) nonperseverative number errors; and (c) Maintaining Set. A three-part hierarchy of response determinants for the CT was utilized, consisting of (a) concrete perceptual attributes; (b) cognitive organization of perceptual attributes into abstract patterns; and (c) relating abstract patterns to the corresponding number responses. Decision trees were devised to prescribe a set of rules for coding each score. Utilization of this approach yielded adequate test-retest reliability for recoding responses. Sets of variables for each subtest were factor analyzed, with second order factor analysis of all factors from each subtest in order to determine if common cognitive process scores on each subtest described cognitive process scores on other subtests. Results revealed similar factor solutions for each subtest, but subtest-specific factors were not predictive of similar factor scores on other subtests, except for Subtests V and VI, which are based upon the same principle. Factors related to Maintaining Set predicted most of the variance in subtest error scores. Factor scores related to Determinant Shifting were predictive of error scores to a much lesser degree than Maintaining Set factor scores. Determinant Shifting factor scores appeared to be independent of Maintaining Set factor scores, and also showed much more independence from age, education, and IQ covariates. The relationship between CT and WCST factor scores was slightly lower than the relationship between CT error scores and WCST summary scores. Suggestions for further research were discussed.

(259 pages)

CHAPTER I

INTRODUCTION AND STATEMENT OF THE PROBLEM

Clinical neuropsychology is that branch of psychology which is concerned with the assessment of behavior changes resulting from brain dysfunction (Lezak, 1983). Commonly, clinical neuropsychologists are asked to evaluate whether an individual patient displays evidence of cortical dysfunction, and if so what course of recovery or debilitation might be expected. Questions posed to neuropsychologists by referral sources often include requests for predicting performance of everyday activities such as self-care, driving an automobile, living independently, managing finances, returning to work, or any of a number of other behaviors. Increasingly, the results of neuropsychological evaluation are used to guide cognitive rehabilitation planning. Not surprisingly, these disparate goals of neuropsychological evaluation call for different strategies.

Neuropsychology in this country evolved from the psychometric tradition in psychology of measuring individual differences. The first full-time neuropsychology laboratory was established at the University of Chicago in 1935 by Ward Campbell Halstead (Kløve, 1974) for the purpose of investigating the biological basis of intelligence, particularly with respect to the functioning of the frontal lobes of the human brain, which was thought to have particular significance for higher cognitive functions (Chapman & Wolff, 1959). Toward this goal, Halstead meticulously mapped location and extent of brain damage in patients referred by cooperating neurologists, and quantified the performance of these patients on a series of tasks which he developed or adapted for that purpose, in comparison with the performance of nonbrain-damaged subjects on those same tasks (Halstead, 1947a). Although other researchers had long attempted to specify cognitive impairments characteristic of particular lesion locations, Halstead was the first to apply standardized measurement techniques to cognitive tasks sufficiently sensitive to demonstrate clear differences in adaptive behavior as a consequence of specific brain damage. This series of tasks evolved into the Halstead-Reitan Neuropsychological Battery (HRNB), which remains among the most commonly used standardized batteries of neuropsychological tests in this country (Guilmette, Faust, Hart, & Arkes, 1990). The focus of interest in such an assessment strategy is on the quantifiable features of the patient's test behavior that allow each individual's test performance to be compared to group test results--a statistical or quantitative norm--which permits inferences to be made concerning the relative adequacy of performance (Goldstein, 1987).

These data may be viewed from a variety of perspectives, and Ralph Reitan, Halstead's first graduate student at the University of Chicago (Boll, 1981), who has refined and extensively validated this battery of tests over the past 40 years, has strongly recommended a four-pronged approach to evaluating test performance: (a) level of performance; (b) patterns and relationships among individual test results; (c) specific deficits and pathognomonic signs; and (d) patterns of performance that may be attributed to differential functioning of the two sides of the brain (Reitan & Wolfson, 1985). This strategy may be referred to as an *actuarial* strategy, which, when applied to neuropsychological functioning, is well suited to addressing questions of whether cerebral pathology is present, and to a lesser degree the course of the disease process, and what part of the brain is preferentially affected. Predictions concerning the behavioral capacity of individual patients may also be inferred from such normative data. Such conclusions are consistent with Halstead's original research goals of identifying the behavioral correlates of brain function (Reitan, 1986).

These endeavors have long been identified as comprising the *raison d'être* of clinical neuropsychology (Diller, 1987). However, the use of neuropsychological evaluation for cognitive rehabilitation planning has become increasingly important in recent years. This goal necessitates a shift in focus and purpose from the structure and function of cerebral systems to cognitive processes and behavioral adaptation (Sohlberg & Mateer, 1989). Diller (1987) conceptualized that relationship between neuropsychology and rehabilitation in the following terms:

> The primary language of neuropsychology is one of impairment. Impairments are deficits which are dysfunctions in underlying mental or physical structures. Deficits are identified by responses to standardized tests. The primary language of rehabilitation is one of disability. Disabilities are limitations in actual functioning in daily activities. . . The neuropsychologist has a problem in translating the language of impairment into a language of disability without a complex chain of assumptions. While the neuropsychologist can categorize the complaint, the data are too nonspecific to illuminate management issues at more refined levels. (p. 4)

Clearly, a different methodology is necessary to meet this challenge. Rehabilitation of cognitive disabilities requires careful specification of skill parameters which are impaired as well as those which are not, and a conceptual model which relates cerebral dysfunctions to these skill domains (Sohlberg & Mateer, 1989). The data of actuarial assessment--quantifiable scores--represent "the final orchestrated result of many different cognitive functions" (McKenna & Warrington, 1986, p. 32), and fail to specify how those scores are related to the cognitive processes which must be targeted in a systematic treatment approach.

That goal may be better accomplished by process approach. This methodology retains the standard administration of psychometric instruments but data collection techniques are systematically modified to permit testing the limits of patients' abilities in order to more specifically describe cognitive strengths and weaknesses. An hypothesis testing approach is employed on many tasks in order to better specify the cognitive processes which contributed to patients' performances. This approach has been applied with success to the assessment of diverse cognitive processes, but ironically while more complex cognitive functions tend to be most sensitive to the effects of brain-damage, these abilities lend themselves less readily to process analyses (Milberg, Hebren, & Kaplan, 1986). Included among these so-called higher cognitive functions are reasoning and problem-solving ability, concept formation, cognitive flexibility, abstract thought, and a variety of executive skills. These are the abilities which are most called upon in accomplishing the tasks of daily living and therefore tend to be the focus of treatment goals (Goldstein & Levin, 1987).

Two tests which are most often employed to assess these skills are the Halstead Category Test (CT), one measure of the HRNB, and the Wisconsin Card Sorting Test (WCST). Many similarities in the constructs measured by these tasks have been noted in the clinical literature, but in the few studies in which error scores on both tests have been compared, only approximately 30% shared variance has been found. Consequently, the CT and WCST have each been considered to contribute unique information concerning higher cognitive functioning to the neuropsychological evaluation. Yet little research effort has been directed toward delineating precisely what unique information each test contributes (Anderson, Damasio, Hones, & Tranel, 1991) or the *nature* of the problem-solving strategies which are reflected in the scores reported for these tests. Consequently, although validation research has shown that performance on these tasks is among the most sensitive indicators of brain damage (Kløve, 1974), and differential performance on these tasks may be of some value in localizing the site of damage (Boll, 1981; Lezak, 1983), that information is of limited value in planning and implementing intervention strategies. Patients may show deficits in complex, nonverbal problem-solving skills (Reitan & Wolfson, 1985) due to cognitive perseveration, assuming a very stimulus-bound approach to the task, inability to generate alternative hypotheses, inability to maintain cognitive set, or a variety of other reasons. All of these approaches may yield similar quantitative (error) scores, leading to similar interpretive conclusions in the neuropsychological evaluation, yet these scores are of little value in directing attention to the specific cognitive skills which are deficient in producing this result or in suggesting specific remediation strategies.

The problem which this study addressed was that conventional administration of both the CT and the WCST yields only quantitative scores which do not support inferences concerning the cognitive strategies used on these tasks by individual patients. It has been suggested that asking patients to verbalize their reasoning on these tasks may be one method for assessing the cognitive processes employed by patients as they seek to solve the problems presented to them by these tasks (Bond & Buchtel, 1984). Others have also recommended this methodology as a means of clarifying the reasoning underlying patients' responses (Reitan & Wolfson, 1985), but use of this approach has not previously been reported in the clinical or research literature.

In view of the need for a more accurate description of the cognitive deficits that characterize the problem-solving behavior of patients with brain injuries, the purpose of this study was to devise a coding system to classify CT and WCST responses using a "think aloud" methodology, in order to achieve the following goals: (a) describe the cognitive strategies employed by individual patients in performing these tasks, and (b) differentiate the cognitive strategies assessed by the CT versus the WCST.

CHAPTER II

REVIEW OF THE LITERATURE

Both the Halstead Category Test and the Wisconsin Card Sorting Test have been extensively validated, but the focus of much of this research activity has been concerned with accurately discriminating brain-damaged from a variety of nonbrain-damaged groups, or differentiating a variety of brain-damaged groups from each other. While this approach is useful from a diagnostic perspective, it is entirely "conceptually opaque" (Bertram, 1984), and adds little to understanding what is being measured. This literature, therefore, is generally peripheral to the purposes of the proposed research, and will be only selectively reviewed here within the context of providing background history. The theme of the review of the literature which follows will be to describe research to date which elucidates the nature of the cognitive processes required for successful performance on the CT and WCST. Toward this end, literature relating to the development and construct validity of the CT will be considered first, followed by a review of the literature relating to the development and validation of the WCST. Those few studies which have directly compared performance on both tests will also be described, with a view toward elucidating the basis of common and unique variance.

The Halstead Category Test

Historical Background

Until the beginning of this century, most theories of cerebral functioning sought to associate specific cognitive "faculties" with discrete areas of the brain, although this was by no means a universal view. Flourens (1794-1867), for example, anticipated the notion of equipotentiality, arguing that the extent of tissue damage is more critical than the specific site of damage (Walsh, 1987). However, the theory of functional localization gained preeminence with Paul Broca's demonstration that articulate speech was impaired in patients with lesions in the posterior portion of the left frontal lobe. That discovery precipitated intensive efforts to localize other faculties in specific regions of the brain (Filskov, Grimm, & Lewis, 1981). However, a combination of influences produced a reaction against the excesses of functional localization. In the interval between the two world wars, the experimental work of Karl Lashley supported a shift of emphasis to more global theories of cerebral functioning (Lezak, 1983). In studies of the effects of cortical ablations on maze learning in rats, Lashley demonstrated that the mass of intact brain tissue is more critical than location of lesions. Thus, different cortical regions were seen as equipotential in subserving specific functions (Filskov et al., 1981).

Against this background, Kurt Goldstein's work with brain-injured individuals prompted him to theorize, drawing upon Hughlings Jackson's work (Chapman & Wolff, 1959), that the brain is organized in hierarchical levels, with higher levels inhibiting and subsuming the functions of lower levels. Goldstein postulated that cerebral insults caused regression to more primitive levels of functioning due to disinhibition of lower levels. Such regression could be chiefly distinguished by manifestation of the "concrete attitude" and loss of capacity for the "abstract attitude" (Goldstein, 1940). That distinction was described as follows:

> The concrete attitude is realistic. It does not imply conscious activity in the sense of reasoning, awareness or a self-account of one's doing. We surrender to experiences of an unreflective character: we are confined to the immediate apprehension of the given thing or situation in its particular uniqueness. This apprehension may be by sense or percept, but is never mediated by

> discursive reasoning. Our thinking and acting are directed by the immediate claims which one particular aspect of the object or of the outerworld situation makes. . . . The abstract attitude embraces more than merely the "real" stimulus in its scope. It implies conscious activity in the sense of reasoning, awareness and self-account of one's doing. We transcend the immediately given situation, the specific aspect or sense impression: we abstract common from particular properties; we are oriented in our action by a rather conceptual viewpoint, be it a category, a class, or a general meaning under which the particular object before us falls. We detach ourselves from the given impression, and the individual thing represents to us an accidental example or representation of a category. (Goldstein & Scheerer, 1941, pp. 2-4)

Goldstein and his co-workers devised a series of sorting tasks to assess the abstract attitude, but evaluated patients' performances qualitatively as they viewed abstract vs. concrete thought as dichotomous. Goldstein argued:

> Even in its simplest form, . . . abstraction is separate in principle from concrete behavior. There is no gradual transition from the one to the other. The assumption of an attitude toward the abstract is not more complex merely through the addition of a new factor of determination; it is a totally different activity of the organism. (Goldstein, 1940, p. 258)

Within this historical context, the emergence of the assumption that organicity was characterized by "a central and therefore universal defect" (Lezak, 1983) appeared reasonable. It is not surprising, consequently, that unidimensional tests of "organicity" to distinguish brain-damaged from nonbrain-damaged patients proliferated (Boll, 1981). However, that goal was elusive. Goldstein's qualitative sorting tasks appeared to have some potential in this regard, but inadequate quantification procedures impeded further development (Berg, 1948).

Development of Halstead's Battery

Halstead (1947a) approached the problem from a somewhat different perspective. He was keenly aware that these attempts to diagnose organicity failed to address the essential question of how biological changes in brain damage influenced behavioral or psychological events. This, then, became his goal. Toward that end, Halstead developed or adapted a variety of tasks which tested various domains of behavior, carefully specifying location and extent of brain lesions in his patients who had been referred to him by cooperating neurologists in the University of Chicago Department of Medicine. He selected those tests which were readily subject to quantification, and submitted the results of 50 patients who had undergone lobectomies and 30 control subjects to both Holzinger (Halstead, 1945) and Thurstone for factor analysis. Both solutions produced four discrete factors, but Thurstone's analysis formed the basis for Halstead's Theory of Biological Intelligence (1947a), by which he sought to relate the adaptive coping of the individual to the intact functioning of the central nervous system. These four factors included:

1. A central integrative field factor C. This factor represents the organized experience of the individual. It is the ground function of the "familiar" in terms of which the psychologically "new" is tested and incorporated. It is a region of coalescence of learning and adaptive intelligence. Some of its parameters are probably reflected in measurements of psychometric intelligence which yield an intelligence quotient.

2. A factor of abstraction A. This factor concerns a basic capacity to group to a criterion, as in the elaboration of categories, and involves the comprehension of essential similarities and differences. It is the fundamental growth principle of the ego.

3. A power factor P. This factor reflects the undistorted power factor of the brain. It operates to counterbalance or regulate the affective forces and thus frees the growth principle of the ego for further ego differentiation.

4. A directional factor D. This vector constitutes the medium through which the process factors, noted here, are exteriorized at any given moment. On the motor side it specifies the "final common pathway," while on the sensory side it specifies the avenue or modality of experience. (p. 147)

Included in the group of 10 tests (see Table 1) validated by Halstead was the Category Test, which had evolved from stimuli similar to Goldstein's sorting

Table 1

Description of Original Halstead-Reitan Neuropsychological Battery

Test Description	Cut-Off Score
1. Category Test: Subject is presented with 208 stimulus figures on a milk glass screen projection apparatus. Below the screen are four numbered levers, and the subject is instructed to pull the numbered lever which corresponds to the number which is suggested by the figure presented. If the response is correct, a bell sounds; if not, a harsh buzzer sounds.	51 or more errors
2. Critical Flicker Test Frequency: An electronic instrument with a short flash duration provides light of variable frequency. The subject's task is to adjust a knob until the variable frequency light appears to fuse into apparently steady light. Score is light frequency, in cycles per second.	20.9 and below
3. Critical Flicker Test Deviation: Deviation from mean values of Critical Flicker Frequency for five successive trials.	15.7 and above
4. Tactual Performance Test - Total Time: A form board with 10 geometric figures cut into the board is placed on a stand in front of the subject, along with wooden blocks which fit into each space. The subject is blindfolded and must fit each block into the correct space, first using only the preferred hand, then only the nonpreferred hand, then both hands together. Score is total time for the three trials.	15.7 minutes and above
5. Tactual Performance Test - Memory: After completion of the third trial (described above), the subject is asked to draw a picture of the board, which he/she has never seen, with as many of the geometric figures as possible in the correct placement on the board. Score is expressed in terms of number of geometric shapes correctly recalled.	5 figures and below
6. Tactual Performance Test - Localization: Score is number of geometric shapes correctly localized in drawing described above.	4 figures and below
7. Rhythm Test: This is a subtest of the Seashore Test of Musical Talent. The subject is required to differentiate which of 30 pairs of rhythmic sounds are the same and which are different. Number of correct responses are transformed into a rank score, ranging from 1-10.	Rank score 6 and above
8. Speech Sounds Perception Test: 60 oral nonsense words, which are variants of the "ee" sound, are presented. Subject must determine which of four alternatives were presented. Score is number of errors.	8 errors and above
9. Finger Oscillation (Tapping) Test: Subject must tap a finger tapping device with the index finger of each hand for five trials of 10" each.	Mean score of 50 and below
10. Time Sense Test - Memory Component: Subject presses a key which permits a sweep hand to rotate on a clock face. The task is to stop the sweep hand after 10 rotations in the same position in which it started. After 20 trials, the clock face is removed, and the subject must duplicate the visually controlled performance as closely as possible. The score is error made for memory trials.	261 and above

Adapted from Reitan (1966b)

tasks. These tasks required subjects to sort objects into hierarchical categories (Weigl, 1941). Halstead's initial quantitative analysis of the results clearly differentiated brain-lesioned patients from normal subjects, though qualitative analysis did not support Goldstein's prediction that the strategies employed by the two groups would be qualitatively distinct (Bertram, 1984). In Thurstone's factor analysis, this test loaded most highly on the abstraction factor (A), with secondary loading on the central integrative factor (C) (Halstead, 1947a). These results suggested that the Category Test demands abstraction of salient features of the stimuli presented and the capacity to integrate new with previously learned information.

Further development and validation of Halstead's battery of tests was undertaken by his former student, Ralph Reitan, who initiated a series of studies with the goal of validating the usefulness of Halstead's 10 measures (derived from seven tests) in differentiating brain-damaged from nonbrain-damaged patients. Although he stated that he employed the "same tests administered in the same way" (Reitan, 1955a, p. 29), in fact, in consultation with Halstead (Wetzel, 1983), he shortened the Category Test to 208 items and deleted two of the nine original subtests, proportionally decreasing the criterion error score from 80 to 50 errors.

The current version of the CT retains that format. As presently constituted, the CT consists of photographic slides of geometric figures of systematically varied complexity, organized into seven subtests presented in sets of similar items ranging in number from 3 to 17, though a number of items are presented singly. Test subjects are informed as follows (after completing the initial practice items which comprise Subtest I):

> . . . This test is divided into seven subtests. In each subtest there is one idea or principle that runs throughout the subtest. Once you have figured out

> what the idea or principle in the subtest is, by using this idea you will get the right answer each time. Now we are going to begin the second subtest and the idea in it may be the same as the last one or it may be different. (Reitan, 1979, p. 33)

These instructions are reiterated throughout the test in abbreviated form, in order to clarify the task as much as possible. Further, Reitan emphasized that the subject's task is discovering the organizing principle in each subtest, not discovering what is required by the task. Therefore, rephrasing and repeating instructions as necessary is encouraged.

The slides are displayed on a screen with four numbered response keys below the screen, and the subject is instructed, "Something about the pattern on the screen will remind you of a number between one and four . . . " (Reitan, 1979, p. 33). The subject presses the key corresponding to that number for each slide to obtain feedback concerning response accuracy. A correct response sounds a bell, while an incorrect response causes a harsh buzzer to sound.

The first two subtests are essentially practice items, and few subjects, even those who are severely impaired, miss more than one or two items. Subtest I consists of 8 items and requires the simple matching of Roman Numerals to the appropriate response key. Subtest II has 20 items and requires simple counting of the number of designs on each slide. Subtest III contains 40 items and requires identification of the linear position of the form which is most dissimilar from the other three. Subtest IV also has 40 items and requires identification of the quadrant of the design, numbered clockwise, which is discrepant. Each quadrant on the first six items of this subtest contains a Roman Numeral corresponding to the number of the quadrant, thereby providing practice items to familiarize the subject with the correct principle. Subtests V and VI each consist of 40 items and are based upon the same principle. The subject is required to select the lever that corresponds to the numerator of the fraction of the figure which is pictured (1/4, 2/4, 3/4, 4/4). The final subtest contains 20 items which review the principles contained in the first six subtests, and is commonly regarded as a memory task (Reitan & Wolfson, 1985).

Predictive Validity

Reitan's (1955a) original cross-validation of Halstead's battery established the methodology for much subsequent validation research in neuropsychology (Goldstein, 1969). In this initial demonstration of that model, Reitan compared 50 pairs of subjects, matched for age, sex, education, and ethnic origin, with one member of each pair with confirmed brain damage, and the other member of the pair serving as a control subject[1]. Results of these pairwise comparisons indicated that all tests in the battery, with the exception of two measures based on Critical Flicker Fusion, discriminated between the two groups of subjects at levels equal to or greater than those in Halstead's original study.

In this analysis, the Category Test proved to be the most accurate discriminator of all the tests administered, second only to Halstead's Impairment Index, a summary measure computed as the proportion of measures which exceed the criterion for discriminating impaired vs. normal performance. In a subsequent study, Reitan (1956) demonstrated that the CT error score correlated substantially with the Impairment Index (r=.71 for brain-damaged subjects and r=.50 for nonbrain-damaged subjects), and he concluded that the CT is a multidimensional task in terms of cognitive and perceptual elements. That complexity is confirmed by findings (Reitan, 1955a) that the sensitivity of the CT

[1]Control subjects consisted of 13 paraplegics, 17 depressed inpatients, 6 patients with acute anxiety state, 2 obsessive-compulsive neurotics, and 12 "normals."

to cerebral impairment is not confined to specific lesion sites. These results have subsequently been confirmed by numerous other investigators. For example, in their cross-validation of Reitan's (1955a) results, Vega and Parsons (1967) compared the performance of a group of 50 persons with confirmed brain-damage representing a variety of etiologies and 50 control subjects without evidence of brain pathology on Halstead's measures, which by this time were referred to as the Halstead-Reitan Neuropsychological Battery. Again, all measures discriminated between the two groups at a level far exceeding chance ($p <.001$) for all but the Time Sense Test. However, using the recommended cut-off score on the Impairment Index (.6), 54% of the control subjects were misclassified as brain-damaged, but upward adjustment of the cutoff score to .7 yielded a correct classification rate of 73%. The less than expected efficiency of this battery, including the CT, in accurately discriminating groups was attributed to the fact that the control subjects were "malfunctioning individuals" (p. 623) resembling "pseudoneurologic" patients and were older than subjects in other studies reported in the literature.

Heilbrun (1962) suggested that it is precisely this form of discrimination which is most useful in clinical neuropsychological evaluations. Most often, neuropsychologists are asked to assist in ruling out organic pathology when patients present neurologic-like symptoms. Addressing this problem, Matthews, Shaw, and Kløve (1966) compared the performance of 32 brain-damaged patients with 32 "pseudoneurologic" patients, hospitalized with complaints of headaches, paraesthesias, motor weakness and/or incoordination, gait disturbance, visual difficulties, or ictal episodes simulating epilepsy, but for whom extensive neurological testing yielded negative results. Both groups were matched on age, sex, and years of education, thus maximizing the similarities between the groups.

Again, all measures, including the CT, significantly discriminated between the groups. The authors reported that no single measure by itself was able to obtain accurate agreement with neurological findings for members of both groups on this difficult task, but the example they provided for the composite Impairment Index, comprised of 10 separate measures, indicated a rate of correct classification of 83%, which compared very favorably with the accuracy of other techniques.

Shaw (1966) found a significant correlation (r=.64) between magnitude of total error score on the CT and Severity Continuum scores, computed as an expanded Impairment Index utilizing nine HRNB variables, for a heterogenous group of 674 subjects referred for neuropsychological evaluation, indicating 41% common variance between these measures. Numerous other studies have confirmed the usefulness of the CT in discriminating brain-damaged from nonbrain-damaged groups (Hevern, 1980; Kløve, 1974), although it has been much less successful in discriminating between brain-damaged and psychiatric patients, possibly because of the organic basis of many psychiatric disorders (Goldstein, 1969).

Localization Studies

Despite Halstead's initial claims that the CT effectively discriminated between patients with frontal lobe lesions and posterior lesions (Halstead, 1939; Halstead, 1940; Halstead, 1947a; Halstead & Settlage, 1943; Shure & Halstead, 1958), subsequent researchers have failed to confirm consistent localizing effects for either caudality or laterality. Walsh (1987) noted that the original claims were based upon small group comparisons and levels of statistical significance were accepted which were far short of those which are conventional today. Chapman and Wolff (1959) reanalyzed Shure and Halstead's (1958) data, controlling for size of cerebral lesions, laterality, and locus (frontal vs. nonfrontal), and found no significant main effects for side or site of lesion, but degree of impairment on the CT was directly related to the mass of cerebral tissue affected. Those findings were confirmed by Matthews and Booker (1972), who correlated pneumoencephalographic measurements of ventricular size with neuropsychological test performance. The CT was one of only 3 of 24 neuropsychological measures that showed significant differences on all three ventricular size group comparisons (largest vs. smallest planimeter and linear measurements of the lateral ventricles and largest vs. smallest third ventricle measurements).

McFie and Piercy (1952a) found that left brain-damaged patients performed significantly worse on a task (Weigl, 1941) with which the CT shares common ancestry, in comparison with right brain-damaged patients, appearing to suggest a role for language mediation on this task, but no difference in performance was found for aphasics vs. nonaphasics within the left brain-damaged group, as might be expected if this task were facilitated by linguistic analysis. De Renzi, Fagliori, Savoiardo, and Vignolo (1966), employing a modified version of Weigl's Test, also found poorer performance by left-brain-damaged patients, but only those with aphasic symptoms, in comparison with right brain-damaged and nonbrain-damaged hospitalized control patients. Left brain-damaged nonaphasic patients performed more like right brain-damaged patients and control patients. De Renzi et al. postulated that these discrepant results reflected different proportions of patients with prefrontal lesions in their left brain-damaged group in comparison with those studied by McFie and Piercy, due to differing etiologies, and frontal dysfunction has been specifically associated with impaired abstracting ability (Milner, 1963)

Although this task is similar in some respects with the CT, similar results have been inconsistently obtained with the Category Test. Reitan (1960) investigated the relationship between thinking and language by comparing performance on neuropsychological testing, including the CT, of 32 dysphasic patients and 32 nondysphasic patients, matched, insofar as possible, on sex, color, chronological age, years of education, and type of brain lesion. Verbal subtests of the Wechsler-Bellevue Scale of Adult Intelligence significantly discriminated between groups, but both groups were approximately equally impaired on the CT, prompting Reitan to conclude that language skills are less important in complex thinking than was generally presumed.

Similarly, Doehring and Reitan (1962) found no significant difference in CT total error scores for 50 left brain-damaged patients in comparison with 50 right brain-damaged patients, although both groups were significantly impaired in comparison with a nonbrain-damaged comparison group, and both brain-damaged groups displayed a pattern of performance on the Wechsler-Bellevue Adult Intelligence Scale consistent with site of lesion. Subtest analysis of all three groups revealed a similar pattern of relative performance across all CT subtests, suggesting that differential patterns of brain damage did not result in qualitatively different patterns of cognitive processing.

Parsons, Jones, and Vega (1971) obtained similar results in comparing the CT performance of left and right brain-damaged patients matched for general level of psychological impairment. Both groups were impaired in comparison with a nonbrain-damaged comparison group, but did not differ from each other with respect to total error score. However, the left brain-damaged group made significantly more errors on Subtest IV in comparison with Subtest III than the other groups. It was suggested that these results may have resulted from the verbal content (Roman Numerals) of the sample items on this subtest, with which left brain-damaged patients might be expected to have difficulty. However, that pattern is discrepant with results obtained by Doehring and Reitan (1962).

Schreiber, Goldman, Kleinman, Goldfader, and Snow (1976), using neurological diagnosis as a criterion, evaluated the accuracy of the HRNB in discriminating brain-damaged from nonbrain-damaged patients and localizing lesion site. They found the CT to be the single most sensitive test of brain damage, but it was not useful in identifying lesion site. Similar results have been reported by Matthews and Booker (1972), Goldstein and Shelly (1973), Russell (1974), and Bornstein (1986).

Thus, somewhat paradoxically, although CT performance is regarded as a reflection of perceptual organization or visuospatial skills, which are generally attributed to predominantly right hemisphere functions (Lansdell & Donnelly, 1977), it has repeatedly been found to be insensitive to lesion site. Reitan attributed this paradox (Reitan & Wolfson, 1985) to the complexity of skills required by this task which necessarily involves, in addition to visuospatial elements, adequate visual memory, abstract reasoning skills, and the ability to translate visual stimuli into verbal concepts (Rothke, 1986), reflecting the overall integrity of the cerebral cortex.

On the other hand, Doehring and Reitan, in an earlier study (Doehring & Reitan, 1961), compared the performance of brain-damaged patients with right or left homonymous visual field defects, brain-damaged patients without visual field defects, and nonbrain-damaged patients on neuropsychological tests, including the CT. All tests administered significantly discriminated the nonbrain-damaged comparison group from the three brain-damaged groups and visual field defects did not produce significantly more impaired performance than brain-damage not associated with visual field defects. Surprisingly, patients with left visual field defects were significantly more impaired on the CT than patients with right visual field defects, implicating right hemisphere dysfunction with poorer CT performance. The authors postulated, given a similar pattern of results on Total Time on the Tactual Performance Test, that right hemisphere mediated deficits in shape recognition may account for these results, a conclusion which seems intuitively reasonable given the strong association of the CT with the Block Design subtest on the WAIS, which is generally viewed as a predominantly right hemisphere-mediated task.

However, the failure of subsequent studies to replicate those results has led to inferences that the CT lacks localizing significance due to the complexity of the task which calls upon multiple cortical systems (Bornstein, 1986) Nonetheless, a more recent study which has obtained results reminiscent of Doehring and Reitan's earlier findings again raises questions concerning possible lateralization of CT performance. Winkleman (1982), in seeking an optimal combination of measures not dependent upon intact motor/sensory skills that would accurately discriminate right hemisphere deficits, unexpectedly found that the CT alone discriminated lateralization of lesions nearly as well as a combination of measures identified through a review of the literature. The CT, in combination with the Picture Arrangement subtest of the WAIS, equaled the discriminating power of the literature-derived battery. Winkleman suggested that these discrepant results may have been due to differences in the rehabilitation population she studied in contrast to the acute care population referred for neuropsychological evaluation in most medical centers, from which most previous studies have drawn their subjects. The possibility that a systematic difference in severity between right and left hemisphere groups may have accounted for these results was considered, but rejected. Rather, Winkleman suggested that recovery of function by nonfocally affected cortical tissue may have produced systematic differences in her rehabilitation population in comparison with an acute care population. This suggestion is reminiscent of Hughlings Jackson's argument that the signs and symptoms produced by specific brain lesions can only be understood within the context of the manner in which those lesions affect the remaining intact cortical areas (Reitan, 1966b).

Moderator Variables

Discrepant results such as these point to the methodological problems faced by neuropsychological researchers who attempt to specify the cognitive processes tapped by a wide variety of behavioral measures. Since it is not feasible (nor desirable if it were) to randomly assign subjects to lesion groups in order to control for a host of organismic variables which may affect results in unknown ways, it is especially important to define moderator variables which may systematically influence performance on the psychological measures of interest. A number of such individual variables have been identified which correlate substantially with CT performance.

In Halstead's (1947a) original validation of his battery of tests, he carefully specified age, sex, education, socioeconomic status, and occupation of his subjects, but apparently did not treat these demographic variables as covariates of test performance. Nonetheless, in Reitan's (1955a) cross-validation, brain-damaged and nonbrain-damaged subjects were matched for race, sex, age, and years of formal education, to control for possible confounding effects produced by these demographic variables. This procedure yielded better discrimination between groups than in Halstead's original study (Parsons & Prigatano, 1978). Consequently, Reitan (1955b) investigated chronological age as an independent predictor of test performance for new groups of brain-damaged (N=194) and nonbrain-damaged (N=133) subjects on Halstead's neuropsychological battery of tests. Each group was divided into 5-year cohorts ranging in age from 15 to 65. For the brain-damaged group, age was only weakly correlated with the Impairment Index, a summary measure representing the proportion of tests performed within the brain-damaged range (r=.23), but the correlation was much more substantial for the nonbrain-damaged group (r=.54). Further analysis revealed that this relationship was primarily generated by subjects over 45 years of age. Under that age, no significant relationship between age and test performance was apparent. These results appeared to support evidence of organic brain changes in older adults, although it was noted that some individuals in each age group continued to score within the normal range.

Subsequently, Reed and Reitan (1963a) compared the relative performance of a young brain-damaged group and a young nonbrain-damaged group with the relative performance of an old nonbrain-damaged group (age 40-49) and an older nonbrain-damaged group (age 50 and above). The scores for each test on Halstead's battery were ranked from poorest to best performances for each group and the resulting distribution of ranks were converted to a normalized T-score distribution with a mean of 50 and standard deviation of 10. T ratios were then computed for each test. The results for the young groups indicated the sensitivity of each test to the presence of brain damage, while for the older groups the resulting T ratios indicated the sensitivity of each test to differences associated with age. Results of the rank-order distributions for the two sets of comparisons suggested that the decrements in performance associated with brain-damage for the young group were similar to the decrements in performance associated with age for the older group (r=.49). Analysis of covariance revealed that educational differences between the two older groups were not responsible for these results.

Vega and Parsons (1967) obtained similar results in their cross-validation of Reitan's (1955a) validation study, although they found a substantial correlation with education (r=.56) as well as age (r=-.57) for their control group, whereas both correlations were significantly attenuated in the brain-damaged group (r=-.33 and r=.20, respectively). Both correlations were significant at $p = .01$ for the nonbrain-damaged group, but for the brain-damaged group only age was a significant predictor, and only at the $p = .05$ level. The CT was the measure most highly correlated with age for both the brain-damaged and nonbrain-damaged groups, followed by other complex measures (Tactual Performance Test and Impairment Index). It was concluded that this battery of tests, especially including the CT, is sensitive to any factor which impairs the integrity of physical and psychological functioning.

The confounding role of nonbrain-damage variables, including age, education, and psychiatric status, was more specifically addressed in Prigatano and Parsons' (1976) cross-validation of Vega and Parsons' (1967) results. Whereas in the earlier study the brain-damaged group was severely impaired, in the cross-validation study degree of impairment was more moderate. Additionally, the nonbrain-damaged group in the cross-validation consisted of psychiatric patients referred for neuropsychological evaluation to differentiate neurological from pseudoneurological deficits. Age was again significantly correlated with all test variables for the nonbrain-damaged group, but education was not for either group, and the relationship between age and test performance was attenuated by the presence of brain damage. The CT and other complex tasks continued to be strongly associated with age in this study, although the relationship was attenuated somewhat in the psychiatric control group. The fact that partialing out the contribution of education to test performance in the nonbrain-damaged group in the earlier study attenuated the relationship between age and test performance, but had no such effect on the nonbrain-damaged psychiatric group in the cross-validation, appeared to suggest that these patients failed to bring prior learning experiences to bear on solving problems, which the authors suggested may have prompted their referral for neuropsychological evaluation.

This series of studies appears to suggest rather convincingly that chronological age is associated with predictable organic brain changes in most, but not all, individuals which impact problem-solving skills on neuropsychological tests in a manner similar to brain damage in younger subjects. However, in the presence of confirmed brain damage, the impact of chronological age is much less predictable. The contribution of educational level appears to be much more moderate.

The specific pattern of cognitive changes associated with aging was investigated by Meyerink (1982). He studied 125 subjects with no evidence of neurological injury or disease, divided into five age groups: (a) 20-29 years of age; (b) 30-39 years of age; (c) 40-49 years of age; (d) 50-59 years of age; and (e) 60-70 years of age. Multivariate analysis of variance, based upon HRNB performance, suggested that tests related to prior learning showed little change across age groups, and sensory input functions and simple motor functions also showed little change, but as tasks became more complex, greater deterioration of functions across age groups became evident. Thus, substantial decrement of complex visual-spatial abilities occurred with advancing age, and tasks involving reasoning, abstraction, and logical analysis showed pronounced deterioration.

Among the tasks most representative of these complex cognitive skills is the CT. Reed and Reitan (1963b) asked three experts to rate 29 tests, including the HRNB, on a continuum of those most highly dependent upon prior experience and learning to those most dependent upon immediate adaptive ability, with consideration given to the complexity of problem-solving involved. The CT was rated most dependent upon immediate adaptive ability and complex problem-solving skills, and was also the task which best discriminated the performance of an older group (mean age = 52.96; S.D. = 6.27) and a younger group (mean age = 28.05, S.D. = 5.07). The differences between the groups on tasks rated most dependent upon previous experience and learning were minimal. These results were subsequently replicated on a group of 283 subjects with long-standing confirmed cerebral dysfunction, divided into old (mean age = 49.66, S.D. = 10.72) and young (mean age = 24.21, S.D. = 5.98) groups (Fitzhugh, Fitzhugh, & Reitan, 1964).

Mack and Carlson (1978) compared the performance of 41 bright, aged normal subjects (mean age = 69.76, S.D. = 4.87), 40 young normal subjects (mean age = 25.03, S.D. = 3.70), and 43 neurologically impaired patients (mean age = 41.70, S.D. = 16.47) on the CT. Although the older group scored significantly higher on IQ testing (mean WAIS FSIQ 119.90) than either of the other two groups, they scored significantly worse than the young normal group and similar to the neurologically impaired group on all CT subtests and on total error score, although the young normal group barely exceeded the cutoff for impaired performance. Both the aged and neurologically impaired groups experienced significantly greater difficulty on the most complex subtests (III and IV) than those which were somewhat less complex (V and VI).

While the contribution of age to CT performance has been fairly consistent across studies, the contribution of education has been more variable. As noted previously, Vega and Parsons (1967) found significant education effects among their nonbrain-damaged group, although less substantial than age effects, but Prigatano and Parsons (1976) cross-validation did not find significant correlations of education with CT performance for either their nonbrain-damaged psychiatric group or their brain-damaged group. However, Finlayson, Johnson, and Reitan (1977) did find that education substantially influenced CT performance, as did Ernst (1987), utilizing the Booklet Category Test (BCT), a portable form of the CT which yields comparable results (DeFilippis, McCampbell, & Rogers, 1979).

In view of this variability, Heaton, Grant, and Matthews (1986) performed a large normative study (N=553), drawing normal subjects from three diverse geographical areas, and investigating possible interaction effects among age, education, and sex. Subjects were divided into three age categories (less than 40 years of age, 40-59 years of age, and over 59 years of age) and three education categories (less than 12 years of education, 12-15 years of education, and over 15 years of education). Males and females were matched on age and education variables. Results indicated that age accounted for approximately 32% of the variance in CT total error score, while education accounted for about 13% of the variance. However, age by education interaction effects were evident. These effects suggested that the least educated individuals show the least adequate performance initially and display the greatest decline from the young to middle age periods, but the best educated persons tend to "catch up" in old age, showing little superiority at that point over the least educated individuals. Heaton et al. postulated that these results may reflect accumulated insults to the central nervous system that occur during the life span of the individual.

This hypothesis receives some indirect support by the results of Reitan and Shipley's (1963) study concerning the relationship between changes in serum cholesterol levels and neuropsychological test performance. Serum cholesterol levels of 156 medically healthy individuals between the ages of 25 and 65 were measured at the beginning of the study, 6 months later, and again 12 months later. Differences in neuropsychological test performances at the beginning of the study and at the end were compared for younger (under 40 years of age) vs. older (40 years of age or more) subjects, who had decreased serum cholesterol blood levels by 10% or more vs. those who had not decreased serum cholesterol levels. Results indicated that test-retest changes could not be attributed to decreased serum cholesterol levels for the younger group, but individuals in the older group who had not decreased serum cholesterol levels showed progressive decline in neuropsychological test performance in each 5-year interval between age 40 and 65, while those who had lowered serum cholesterol showed no such decline. These results appear to suggest increased vulnerability to cerebral insults among older persons, which may be reflected in relative impairment on tasks designed to assess cerebral integrity.

Construct Validity

Goldstein (1940) argued that the loss of the abstract attitude accompanying brain damage produced a qualitative difference in the cognitive behavior of brain-damaged persons in comparison with nonbrain-damaged persons. If so, quantitative scores on a given task would be measures of different abilities, precluding group comparison upon which research concerning the behavioral consequences of brain damage was based. Reitan (1958) addressed this issue by reasoning that if performance on Halstead's battery assesses different cognitive processes in the two groups, that performance would be revealed by significantly different interrelationships between scores comprising the battery. Using results obtained in his (Reitan, 1955a) cross-validation study, Reitan found that the correlations between CT subtest scores for each group did not substantially differ. Subsequently, these results were cross-validated (Reitan, 1959a). Each CT subtest is based upon a different principle, except for Subtests V and VI, which require a response based upon the proportion of the design that is pictured. Consequently, Reitan reasoned that if brain-damaged patients' performance is based upon qualitatively distinct processing characterized by an incapacity for assuming the abstract attitude, they will be unable to demonstrate improvement from Subtest V to Subtest VI, unlike nonbrain-damaged subjects. Results revealed that while absolute level of performance on CT subtests significantly discriminated the groups, the relationships among subtests were not significantly different for the two groups. Reitan concluded, therefore, that similar cognitive processes were utilized by both groups in these studies, although the efficiency with which they did so was significantly different.

Reitan assumed that the abstract organizing principle of proportionality was the governing principle utilized by both groups, with differing levels of efficiency, but inspection of the stimuli reveals that these are the only two subtests (other than one of the introductory subtests on which few patients make any errors) where a simple counting response would be reasonably successful. Thus, utilizing a concrete approach to this task may achieve a level of success statistically comparable to that achieved by deducing the abstract principle.

Perhaps as a result of the quantitative tradition out of which the CT developed, only one study to date has examined such item-by-item response determinants. Simmel and Counts (1957) conducted extensive analysis of response determinants of CT performance for 21 patients who participated in a study of the effects of lobectomy on psychomotor epilepsy, 14 patients with psychomotor seizures who did not undergo surgery, and 26 student nurses. Results revealed that both patients' and nurses' responses to CT stimuli are stable and systematic, whether responses were correct or incorrect.

Simmel and Counts (1957) reported a large number of "correct" responses are "incidentally" rather than "essentially" correct; that is, correct responses were given for the wrong reasons. The most persistent response tendency identified was some form of counting conspicuous parts of the stimulus configuration -- number of stimuli presented, number of stimuli that were identical or similar, number of stimuli that were unlike the others, number of component parts in the stimulus configuration, number of similar or dissimilar stimuli in spatial contiguity, et cetera. This was observed not only initially for each subtest, when subjects were attempting to discover the organizing principle, which they had been informed would remain the same throughout each subtest, but also often when the stimulus configuration within a subtest changed, even though they had previously demonstrated having learned the organizing principle. Simmel and Counts remarked, " . . . It appeared almost as if our subjects said to themselves: 'If in doubt, count'" (p. 137). This response tendency was attributed to several factors. First, the combined effect of the apparatus, the instructions, and the stimulus material encourage counting (see Appendix D). For example, the subject is instructed that the numbered response keys are to be used to communicate "a number suggested . . . by the pattern on the screen" (p. 137).

Second, the organization of the CT reinforces counting responses. One of the introductory subtests, which serves as a training set, is based upon the principle of counting stimuli, a principle which requires apparently no learning trials, since few subjects make any errors on this subtest. Consequently, counting as a response tendency is well established prior to the presentation of the test stimuli. Simmel and Counts described counting as a "necessary but not sufficient condition for attaining essentially correct responses" (p. 137). Throughout the remainder of the test, counting continues to be rewarded, and the correct answer can be obtained by counting some feature on 75% of the 180 items in subtests III-VII (excluding only the introductory subtests). This would be sufficient to obtain a score within the normal range, if utilized consistently[2].

Simmel and Counts (1957) reported that approximately one half of the items on the CT in this study showed nonrandom error distributions, indicating a systematic basis for subjects' incorrect responses. Further, it was noted that the same items which were difficult for patients were difficult for the student nurses, and both groups tended to make the same errors on those items. Those findings lend support to Reitan's conclusions concerning the comparability of cognitive processing for brain-damaged and nonbrain-damaged groups, although the reasoning on which those conclusions were based is not supported by this study.

Slightly under one half of the items did show random error distributions, but Simmel and Counts observed that statistical random error distributions do not necessarily suggest that individual subjects are responding randomly to those items. For any given stimulus configuration, several alternative hypotheses are

[2]Only 35 % of the items on subtest IV will be correct using counting responses, but correct responses can be obtained on 85 % of the items on subtests III, V-VII using counting responses, although counting rules may need to change.

available as response determinants, Hypothesis A, Hypothesis B, Hypothesis C, et cetera. On those items where the stimulus characteristics do not "favor" a given alternative, which would bias the distribution of errors, each subject may systematically problem-solve among those alternatives, but choose different hypotheses for a given item. Thus, the group error distribution may appear random, when in fact it was individually systematic. Simmel and Counts (1957) commented:

> This assumption [the a priori assumption of the random nature of errors] is frequently stated explicitly, and even more often quietly implied in the treatment of correct and incorrect responses of a test, i.e., the "scoring" of the test. Specifically, it stated that, for any given test item, errors are made only because the subject does not know the correct answer, and that each of the various error alternatives is therefore equally likely to be selected. In other words, this assumes that the knowledge of the right answer is the one and only determinant of the subjects' responses, just as it is in fact the only determinant for the person who scores the test. Everything else is irrelevant for the tester, *ergo* it is thought to be for the subject. (p. 152)

Frequently, both correct and incorrect responses may be the product of partial insight, or what Simmel and Counts called "working rules" (p. 147). This results from incidentally correct responses that are reinforced, at least on certain configurations, and therefore the hypothesis on which those responses were based is not rejected, but is not adequate to produce correct responses on all configurations. For example, on Subtests V and VI, on which Reitan's (1959a) study was based, the abstract principle is "proportion of the figure which is pictured within the solid lines," but "counting parts" of the figure can result in incidentally correct responses to as many as 73 of 80 items. This characteristic of the CT prompted Simmel and Counts to question the degree to which this may actually be considered a test of abstraction. They noted:

> Abstraction does demand the rejection of conscious and unconscious mental sets induced by the immediate perceptual characteristics of the stimuli, by the explicit and implicit features of the various aspects of the surroundings, by

> learned content, and everyday common procedures and manipulations. However, while this holds for the principle of correct responses, it does not necessarily -- or even typically -- hold for the individual correct responses which are actually given by the subjects. . . Therefore, *a high number of correct responses does not necessarily reflect the attainment of the principle, nor does it indicate the level of abstraction at which the subject actually operated, nor the degree of abstraction of which he might be capable* [italics theirs]. (p. 155)

Simmel and Counts concluded, however, that the stimuli comprising the CT might serve as the basis for a test of abstraction and permit the response tendencies which they identified to be explored systematically. They suggested that individual differences manifesting in different response tendencies might have clinical implications. This suggestion has not been implemented in subsequent research, however.

CT/IQ Covariation

Simmel and Counts noted a substantial connection between IQ and CT performance for the patient population, although not for the student nurses, and suggested that errors may reflect abstraction difficulties or may be related to low intelligence. Other researchers have also found level of intellectual functioning to exert a substantial influence on CT performance for both brain-damaged and nonbrain-damaged subjects. In fact, it is the nature of that relationship which has served as the focus of investigations concerning the construct validity of the CT.

The failure of psychometric tests of intelligence to reflect the biological condition of the brain is the factor which prompted Halstead (1947a) to devise measures of what he referred to as "biological intelligence," which he described as follows:

> [W]hat I mean by biological intelligence raises a difficult problem in brief communication. Some years ago Cannon used the title The Wisdom of the Body for one of his books. I do not believe that Cannon intended to omit the nervous system in his concept. In biological intelligence I am trying to direct attention to the "wisdom of a healthy nervous system." It is my belief that

> psychometric intelligence, as reflected by the I.Q., does not adequately indicate this 'wisdom' of the healthy or of its alterations in the pathological nervous system. We have repeatedly found normal or superior I.Q.'s in neurosurgical patients lacking up to one-fourth of the total cerebrum following frontal lobectomy. Yet our measurements of biological intelligence indicate that these are not normal individuals. . . Biological intelligence, we believe, with high adaptability, is the normal outcome of the functioning of a healthy nervous system. (Halstead, 1947b, cited in Reitan, 1956, p. 537)

The insensitivity of psychometric intelligence testing to cerebral dysfunction was the subject of some controversy in the 1930's and 1940's when Halstead devised his tests of "biological intelligence." This controversy focused on the function of the frontal lobes, which some researchers, including Halstead, believed had some special significance in subserving human intelligence. Some researchers reported decrements in intellectual functions subsequent to frontal lobe injury, while others found none (Walsh, 1987). In part, these discrepant findings stemmed from the nature of tasks employed. Intelligence tests at that time were highly verbal in content (Lansdell & Donnelly, 1977) and presumably heavily influenced by prior learning. This is the type of task which Heaton et al. (1986) found to be least sensitive to cerebral dysfunction. Halstead was the first researcher to devise a series of quantifiable tasks which reliably reflected brain-behavior relationships (Goldstein, 1969). As noted previously, the CT is among the most sensitive of Halstead's measures of biological intelligence (Reitan, 1955a), but somewhat paradoxically has been criticized for failing to assess cognitive abilities separate from psychometric intelligence (Rattan, 1986). This discrepancy is central to understanding what constructs are, in fact, measured by the CT.

In Halstead's (1947a) factor analysis, the CT loaded predominantly on the abstraction (A) factor (.63), but also showed a secondary loading on the central integrative (C) factor (.49). Two other measures which also loaded substantially on those same factors included the Henmon-Nelson Test of Mental Ability (.27, .58, respectively), a test of verbal intelligence, and the Carl Hollow-Square Test (.45, .25), a test of nonverbal intelligence. Although these results have not been subsequently replicated and the latter tests have since disappeared from common usage, these factors appear to correspond to Horn and Cattell's (1966) theory of fluid and crystallized intelligence. Crystallized abilities referred to the influence of "experiential-educative-acculturation" variables, reminiscent of Halstead's conceptualization of psychometric intelligence, while fluid abilities were described as "the measurable outcome of the influence of biological factors" (p. 254), which Halstead called "biological intelligence." These results, then, suggested substantial overlap between psychometric and biological intelligence, but indicated also that these were not measures of identical constructs. Subsequent research has generally supported these conclusions, but discrepant findings have also been reported.

Presently, the most commonly employed individual measure of adult intelligence in the United States is the Wechsler Adult Intelligence Scale (WAIS) or its successor, the Wechsler Adult Intelligence Scale - Revised (WAIS-R) (Lezak, 1983). These measures, like their predecessor, the Wechsler-Bellevue Adult Intelligence Scale (W-B), consist of 11 subtests, each of which is converted to a standard score, allowing inter-subtest comparisons. Six of these subtest scores are summed to yield a Verbal Scale IQ score and five are summed to yield a Performance IQ score. The sum of the Verbal Scale scores and the Performance Scale scores yield an age-corrected Full-Scale IQ score. A variety of factor analytic studies have been conducted to ascertain the relationships among these subtests. These studies have consistently yielded three factor solutions, which have been referred to as Verbal Comprehension, Perceptual Organization, and

Memory/Freedom from Distractibility (Cohen, 1957; Leckliter, Matarazzo, & Silverstein, 1986; Matarazzo, 1972). While recent studies have found two stable factor patterns, Verbal Comprehension and Perceptual Organization, that have appeared little affected by neurological or psychiatric dysfunction (Fowler, Zillmer, & Newman, 1987; Zillmer, Fowler, Newman, & Archer, 1986), others have found that factor structure varies according to type of brain damage (Zimmerman, Whitmyre, & Fields, 1970).

Reitan (1956) correlated Wechsler-Bellevue IQ scores with performance on all HRNB variables, including the CT, for 50 brain-damaged subjects matched on the basis of sex, race, chronological age, and years of formal education with 50 nonbrain-damaged subjects. Correlations between Verbal, Performance, and Full-Scale IQ scores and the CT ranged from .58 to .72 for both the brain-damaged and nonbrain-damaged groups. These results suggest that approximately 34-52% of the variance on one measure can be accounted for by the other measure. All Wechsler-Bellevue subtest scores also showed substantial correlation with the CT. Unlike the age and CT relationship, the presence of brain damage did not significantly attenuate the IQ/CT relationships.

Reitan interpreted these results as supportive of Halstead's concept of the factorial structure of biological intelligence. This interpretation appears warranted not only on the basis of the shared variance between biological and psychometric measures of intelligence, but also in view of findings which demonstrate that IQ measures also are moderately successful in differentiating groups of brain-damaged from nonbrain-damaged patients.

Goldstein and Shelly (1984) compared the discriminative validity of the HRNB, the Luria-Nebraska Neuropsychological Battery (LNNB), and the WAIS. Results revealed that the WAIS correctly classified 65.5% of subjects, in comparison with 77.4% correct classifications for the HRNB and 79.8% for the LNNB. Misclassifications for the WAIS included significantly more false negatives.

A variety of approaches for differentiating brain-damaged from nonbrain-damaged subjects on the basis of W-B or WAIS performance have been presented, with variable results. Wechsler (1944) devised a Mental Deterioration Index (M.D.I.) which was computed as a ratio of subtest scores which did not vary significantly as a function of age, "Hold" subtests, and those subtest scores which show marked deterioration with increasing age, "Don't Hold" subtests. Wechsler proposed that Hold subtests on the Wechsler-Bellevue included Vocabulary, Comprehension, Picture Completion, Object Assembly, and Information. Don't Hold subtests included Arithmetic, Digit Span, Digit Symbol, and Block Design. Subsequent validity studies applying this procedure reported some success was achieved in discriminating nonbrain-damaged from brain-damaged groups (Rogers, 1950a), but application to questions of differential diagnosis found disappointing results (Allen, 1948). Allen sought to improve discriminant validity by comparing performance on the two subtests seen as most resistant to cerebral dysfunction, Information and Comprehension, with performance on the two subtests seen as most sensitive to cerebral dysfunction, Digit Span and Digit Symbol, but these efforts were largely ineffectual (Blake & McCarty, 1948; Rogers, 1950b). Similar ratios were also proposed (Hewson, 1949; Reynell, 1944), but these were not substantially more successful in clinical application (Gutman, 1950).

These approaches were characterized by an assumption that brain-damage is a diagnostic entity that will be manifest unidimensionally, while, in fact, as Reitan (1966a) noted, the effects of brain-damage are enormously complex and multidimensional. Somewhat more success has been achieved relating various patterns of performance on IQ tests to more specific patterns of brain damage.

Kløve (1959a) compared groups of patients manifesting a variety of differential EEG patterns, including a group with EEG abnormalities maximized over the right hemisphere and a group with EEG abnormalities maximized over the left hemisphere. Results of comparisons of these groups revealed that the left hemisphere group achieved Verbal IQ scores which were significantly lower than those achieved by the right hemisphere group, and vice versa with respect to Performance IQ scores, although in the latter case the difference was less pronounced, perhaps because three of five Performance Scale subtest comparisons were not significantly different. Two of these nonsignificant findings, for Picture Completion and Digit Symbol, might have been predicted as these are not viewed as lateralizing subtests.

Doehring, Reitan, and Kløve (1961) compared the performance of patients with right homonymous field defects, indicating left hemisphere dysfunction, with the performance of patients with left homonymous field defects, indicating right hemisphere dysfunction, and two control groups. They found a pattern of performance similar to that reported by Kløve (1959a), with less consistent lateralization effects for Performance IQ scores, but subtest comparisons between the two groups revealed only 4 of 11 comparisons were significantly different and only one, Comprehension, reached the $p<.01$ level of confidence.

Russell (1979) compared the performance of four groups of subjects manifesting differing patterns of cerebral functioning: (a) nonbrain-damaged control subjects for whom neurological examination had ruled out neurological problems; (b) neurology patients with slowly progressive diffuse cortical degenerative disease; (c) patients with acute right hemisphere lesions; and (d) patients with acute left hemisphere lesions. Results revealed three different patterns of performance on the WAIS. The first pattern was demonstrated by the control group and was characterized by normal performance or above on all subtests. The second pattern was characterized by relatively high verbal scale subtests, with Arithmetic somewhat more impaired, and relatively low Performance Scale subtest performance, except for Picture Completion. This pattern was obtained by both diffusely impaired patients and right hemisphere impaired patients. The third pattern consisted of slightly better Performance Scale scores, except for Digit Symbol, than Verbal Scale scores, but the difference did not obtain statistical difference. This pattern was achieved by left hemisphere impaired patients.

Russell suggested these results appear to be produced by interactions of three effects of brain damage on IQ test performance. First, a general effect of brain damage appears to impair performance on all subtests, as evidenced by the significantly better performance of the control group across all tasks. However, some tasks are more sensitive to the effects of general cerebral impairment than others, presenting a pattern of deterioration somewhat similar to that observed by Wechsler and others, although the pattern of subtest deterioration appears to correspond roughly to the verbal/performance dichotomy, except for Picture Completion, rather than in terms of formulas previously proposed. Second, in right hemisphere damage, which preferentially affects visual-spatial skills, Block Design, Picture Arrangement, and Object Assembly are significantly more adversely affected than verbal subtests. Third, in left hemisphere damage, the opposite pattern obtains, but is attenuated by the fact that these Performance subtests are also Don't Hold tests which are differentially sensitive to the general effects of brain-damage in comparison with Verbal Scale subtests, which are

Hold tests. Thus, the decrement in verbal skills associated with left hemisphere dysfunction tends to be offset by a similar decrement in nonverbal skills, as represented by Block Design, Picture Arrangement, and Object Assembly, because these are the tasks on the WAIS which are most representative of Halstead's (1947a) concept of biological intelligence or Horn and Cattell's (1966) concept of fluid intelligence, and are differentially sensitive to general cerebral dysfunction. The marked difference between Verbal and Performance Scales for the right hemisphere impaired group reflected lateralization effects reinforced by deterioration effects. The pattern for the diffusely impaired patients was not differentiable from the right hemisphere group due to the greater sensitivity of Performance Scale subtests, except for Picture Completion, to general cerebral impairment, while for the left hemisphere patients lateralization effects on verbal subtests were offset by deterioration effects on performance subtests.

While these individual effects of brain damage on IQ test performance have been previously well established (Reitan, 1955c), this pattern of interactions between lateralization effects and differential deterioration effects has not been identified. Russell noted that previous studies may have failed to discern the pattern of interactions due to the use of heterogenous samples of brain-damaged subjects and use of T-score means, which in normalizing subtest variability tends to minimize differentiable subtest sensitivity. Moreover, as Reitan (1959b) noted, no two groups of brain-damaged subjects are exactly comparable. This is especially true in studies comparing the performance of lateralization groups on psychological test variables, since not only may one group of subjects with lateralized brain damage not be comparable with a similarly defined group in another study, lateralization groups within the same study may not be comparable with respect to variables other than locus of lesion. For example, severity of impairment may be systematically different between right and left hemisphere lesion groups since severely impaired left hemisphere lesion subjects may be expected to display receptive and/or expressive aphasia to the extent of being unable to comply with test instructions, whereas equivalent severity with respect to right hemisphere impairment may not preclude testing. In such a case, the relationship between test variables may vary as a function of severity rather than lateralization effects or deterioration effects per se, or some interaction between them. Nonetheless, some support for Russell's hypothesis may be found in Matthews and Reitan's (1964) meta-analysis comparing the Wechsler-Bellevue performances of 20 groups of subjects (six right hemisphere, six left hemisphere, and eight nonlateralized brain-damaged groups) derived from six separate published studies (Doehring et al., 1961; Fitzhugh, Fitzhugh, & Reitan, 1962; Kløve, 1959a; Kløve, 1959b; Kløve & Reitan, 1958; Reitan, 1955a). A correlation matrix for the 20 groups based upon rank orders of Wechsler-Bellevue subtest means was computed. Somewhat greater consistency of correlations among rank-ordered subtest means was found for right hemisphere groups than for left hemisphere groups (median coefficients of .80 vs. .67, respectively), and somewhat more consistency was also found among right hemisphere vs. nonlateralized group comparisons than among left hemisphere vs. nonlateralized group comparisons. These relationships are in the direction predicted by Russell, despite marked differences among the groups on a variety of other variables.

Taken together, these studies provide suggestive support for the hypothesis that the factor structure of the WAIS and WAIS-R is invariant for a variety of diagnostic groups, but the relationship among subtests may be attenuated for at least some brain-damaged groups due to the greater sensitivity of some Performance Scale subtests to the general effects of cerebral dysfunction. The nature of that relationship has not been identified, but the variability of the correlation between the CT and intellectual performances among different groups may provide some insight.

Shore, Shore, and Pihl (1971) reported that WAIS Full Scale scores were the best predictor of CT total errors, correlating at =.87, for 29 nonbrain-damaged subjects. Correlations between CT total errors and Cohen's factors on the WAIS were similarly high: .84 with Verbal Comprehension, .72 with Perceptual Organization, 1.00 with Memory, and .76 with General Intellectual Functioning. Shore et al. concluded that these results were not consistent with the use of the CT to diagnose "organicity" independently of intelligence.

However, Lin and Rennick (1974) obtained inconsistent results when they compared performance on the CT and the WAIS for two larger samples (N=177 and 62) of epileptic subjects. In the first sample, correlations between CT total errors and Verbal Scale subtests were relatively moderate, ranging from .35 to .47, while correlations with Performance Scale subtests were somewhat higher, ranging from .37 to .59. The highest correlation was with Cohen's Perceptual Factor (.61), consisting of the sum of scaled scores on Block Design, Picture Arrangement, and Object Assembly. However, on the second sample, the relationships were reversed, and correlations with Verbal Scale subtests were now higher than for Performance Scale subtests. The highest correlations (r=.68) were with Verbal IQ and Cohen's Verbal Factor (sum of Information, Comprehension, Similarity, and Vocabulary) and Information (r=.69). Insufficient information is available to explain this variability, but Lin and Rennick suggested that the greater severity of the first sample may have influenced results. Additionally, many subjects in the second sample had previously taken the WAIS and a practice effect may have elevated their scores, producing greater heterogeneity which tends to increase intercorrelation patterns.

Lansdell and Donnelly (1977) factor analyzed the WAIS, CT, and Finger Tapping Tests of 94 patients, including 59 with psychiatric diagnoses (primarily depression) and 24 with neurological disorders (principally epilepsy). The mean CT total error score of the psychiatric subjects was 70.6, considerably greater than the mean error score of the neurological patients, 50.0, which was barely in the normal range. Factor analysis of scores for the combined group produced a four factor solution, including the usual three factors common the other factor analytic studies of the WAIS: Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility (which Lansdell and Donnelly called numerical). The CT loaded predominantly on the Perceptual Organization factor, as did most Performance Scale subtests. Lansdell and Donnelly concluded that the CT does not measure an ability separate from nonverbal intelligence.

Cullum, Steinman, and Bigler (1984) found higher correlations between Performance IQ and CT total error scores (r=.52) than between CT error and Verbal IQ scores (r=.31) or Full Scale IQ (r=.47) for 92 cerebral trauma patients with confirmed brain damage, although the authors did not report whether these correlations were significantly different statistically. They computed prediction formulas for estimating expected CT performance on the basis of given Performance IQ scores, although the normative value of predictions based upon a brain-injured population is of questionable use clinically. For example, predicted CT scores do not reach the normal range until Performance IQ exceeds 120. In contrast, Logue and Allen (1971) computed a prediction formula using WAIS Full Scale IQ scores, based upon the relationship between CT total error score and Wechsler-Bellevue Full Scale IQ reported by Reitan (1959b) for 50 normal subjects (r=.65), which did not predict impaired performance on the CT until Full Scale IQ scores fell below 70.

Corrigan, Agresti, and Hinkeldey (1987) also found a stronger relationship between Performance IQ and CT total errors (r=-.644) than between Verbal IQ and CT performance (r=-.111) for 102 patients in a rehabilitation hospital with diagnosis of either closed head injury (CHI) or cerebrovascular accident (CVA). Multiple regression analysis revealed that Performance IQ accounted for approximately 36% of the variance in CT total errors, and age, which entered the equation next, accounted for an additional 26% of the variance. This combination of variables was viewed as somewhat redundant since Performance IQ is already age-corrected, and the addition of the age variable might be expected to contribute little additional variance. Consequently, the analysis was recomputed without an age correction, by summing the scaled scores for Block Design, Picture Arrangement, and Object Assembly, the triad comprising Cohen's Perceptual Organization factor. This nonage corrected variable was the only one that accounted for substantial variance (59%). The authors concluded these results suggest age affected both CT performance and this combination of WAIS subtest scores.

Wiens and Matarazzo (1977) investigated the relationship between WAIS and CT performance for two groups of 24 young normal subjects with above average intelligence (mean age 23.6 and 24.8; mean FSIQ 117.5 and 118.3). Group II was used to cross-validate the results for Group I. Results revealed that only performance on the Block Design subtest significantly predicted CT performance for both groups, although when only Group I was considered, both Performance IQ and Full Scale IQ were significantly associated with CT performance (r=.68

and r=.56, respectively). These results suggest that for normal subjects with above average intelligence, individual differences in intellectual abilities may be of only minor significance with respect to problem-solving skills that are measured by the CT.

Moses (1985) investigated the contribution of level of performance variability on HRNB measures, including the CT, for a heterogenous group of 480. He computed a series of multiple regression analyses, using each HRNB measure as the dependent variable and WAIS subtest scores as the predictor variables. With respect to the CT, the Block Design subtest accounted for approximately 34% of the variance in CT performance, followed by Arithmetic, Picture Arrangement, Vocabulary, Picture Completion, and Object Assembly, which together accounted for an additional 9% of the variance.

When level of performance was estimated by the Impairment Index (not including the contribution of the CT in the ratio), which was entered into the equation first, results were virtually identical, with the same subtests entering into the same order, except for Object Assembly. This result was predictable in view of previous findings that the CT is essentially equivalent with the Impairment Index in discriminating brain damage (Reitan, 1955a). However, when level of performance was estimated employing a T-score mean (TMean) of all HRNB variables, derived from published score distributions, and again entered into the equation first, a substantially greater proportion of the variance in CT performance was explained, 56%. The TMean alone contributed 31% of the variance, followed by Block Design (an additional 20%). Digit Symbol, Picture Completion, Vocabulary, Arithmetic, and Picture Arrangement entered the equation next, accounting for an additional 2% of the variance.

Moses subsequently employed Russell, Neuringer, and Goldstein's (1970) Average Impairment Rating (AIR) as a level of performance estimator. The AIR rates performance on each measure of the HRNB on a four-point scale, to permit finer gradations in ratings of severity of impairment than the Impairment Index allows for. Entering AIR into the regression equation first accounted for 39% of the variance in CT scores, followed by Block Design, Vocabulary, Arithmetic, Picture Arrangement, and Digit Symbol, which together contributed another 7% of the variance.

These findings have several implications. First, despite the substantial degree of relationship between CT and WAIS subtest performances, these were clearly not measures of identical constructs, and a greater proportion of the variance remained unexplained than was explained. Second, perceptual organization skills, especially as represented by the Block Design subtest, displayed a much stronger relationship with CT performance than verbal comprehension skills, yet CT performance does not vary with lateralization, as do perceptual organization skills. Third, a substantial proportion of that variance appeared to be represented by a factor related to intact cerebral functions, as represented by Russell et al.'s Average Impairment Rating. This factor appears similar to Halstead's biological intelligence or Cattell and Horn's fluid intelligence.

Factor Analytic Studies

The CT is the product of a factor analytic approach to identifying the cognitive sequelae of brain-damage. It is fitting, therefore, that factor analysis should be the methodology utilized for further investigations concerning the cognitive process variables which determine CT performance. A number of researchers have utilized this approach, usually assessing the CT in the context of all or part of the HRNB and a variety of other neuropsychological measures.

This technique offers the advantage that variance common to a number of tasks may be extracted and the cognitive processes contributing to such common variance may be more readily identified. Aftanas and Royce (1969) factor analyzed the performance of 100 normal persons ranging in age from 16 to 70 on a battery of 25 tests, chosen from an extensive literature review on the basis of validity, reliability, and objectivity of scoring procedures. The entire HRNB, except for Finger Oscillation, was included in this battery, but it did not include the Wechsler Scales. Factor analysis yielded 12 factors, but only three of these were readily interpretable. The CT achieved a modest loading (.443) on a factor interpreted as Perceptual Organization. Other variables which had high loadings on this factor included the TPT, Ravens Colored Progressive Matrices, Porteus Mazes, Memory-for-Designs, and Hooper Visual Organization Test. A common element required in all of these tests is the ability to integrate relevant aspects of nonverbal, perceptual stimuli for problem-solving. In that sense, this factor appears to relate to adaptive problem-solving behavior with limited reliance on previously learned skills, characteristics which are used to describe Halstead's biological intelligence. However, the relatively modest loading of the CT on this factor lends support to interpretations concerning the complexity of cognitive processes involved in CT performance -- at least for this normal population.

Few other factor analytic studies utilizing normal subjects have been performed, but Barnes and Lucas (1974) performed separate factor analyses on test results of 39 subjects referred for routine psychological and neuropsychological evaluation for whom no neurological deficits were subsequently discovered, but who were diagnosed with neurosis or character disorders, and 77 subjects who were found to display definite neurological dysfunctions. All subjects completed the HRNB and WAIS, although only the

WAIS Full Scale score was entered into the correlation matrix. Both analyses resulted in six factor solutions, with the CT loading on Factor I in both cases. The CT loaded on a factor which Barnes and Lucas named Basic Adaptive Ability or Biological Intelligence. A variety of measures that are especially sensitive to cerebral dysfunction loaded highly on this factor, including WAIS Full-Scale IQ. These results, therefore, do not support the distinction of psychometric versus biological intelligence for either the psychogenic group or the organic group, but the failure to enter WAIS subtest scores separately leaves that question open. Barnes and Lucas suggest that cerebral dysfunction may impair cognition less pervasively, if more severely, than nonpsychotic personality problems, but observe that Reitan (1956) has suggested that psychogenic-organic differences probably relate more to a pattern of impairment than a dichotomous characteristic.

Boyle (1988) also included normal subjects in his factor analytic study of the construct validity of a shortened form of the CT, following the suggestion of Chelune (1983) that analyses of data across diagnostic groups have the advantage of increasing variability so that relationships are more easily seen. Like Barnes and Lucas, Boyle did not enter WAIS subtest scores in the correlation matrix, but did enter Verbal and Performance IQ scores, as well as Full-Scale score. Results revealed a high loading for all IQ scores on the same factor on which the CT also loaded highly. Boyle referred to this factor as a general intelligence factor. However, the high loading of education (.76), combined with the low loading for age (.19), and the substantially higher loading for Verbal IQ (.98) than Performance IQ (.66) might suggest, rather, that this factor represents psychometric intelligence. That conclusion would appear to be supported by the fact that the shortened CT loads only moderately (-.42) on this factor. The principal portion of the remaining variance in the CT was unexplained.

Boyle also reanalyzed Halstead's original (1947a) correlation matrix, using modern factor analytic techniques, and reported results similar to his own factor analytic study, with both nonverbal (Carl Hollow Square) and verbal (Henmon-Nelson) intelligence measures loading to approximately the same degree on the same factor as the CT. Boyle again concluded that this factor represents a general cognitive-intellectual dimension, although again the equivalent loadings of the IQ measures on this factor (.56 and .59, respectively) might equally well support an interpretation similar to Halstead's Abstraction (A) factor.

Other factor analytic studies have yielded findings more nearly comparable to Halstead's interpretation. Russell (1982) found the CT loaded principally on a nonverbal or figural factor, in an investigation of the factor structure of a research version of the Revised Wechsler Memory Scale (WMS-R), and Goldstein and Shelly (1972) found the CT loaded on a factor which appeared to represent complex, primarily nonverbal problem-solving ability, in a large factor analytic study. All of the WAIS Performance Scale subtests, along with TPT and Trails B, loaded substantially on this factor, in addition to the CT, while no Verbal Scale subtest did. The Performance Scale subtests also loaded to a lesser degree, with the CT, on the same factor on which Verbal Scale subtests loaded highly. Goldstein and Shelly noted the similarity of this factor structure to that obtained by Halstead.

Russell (1982) also found the CT loaded principally on a nonverbal or figural factor, in an investigation of the factor structure of a research version of the Revised Wechsler Memory Scale (WMS-R). Similar findings were reported by Corrigan and Hinkeldey (1988), who concluded that this factor may represent more than one component, and suggested that it may be sensitive to nondominant cerebral hemisphere functioning and reflect active processing of nonverbal information. They suggested that unique patterns of performance may be related to particular clusters of patient abilities, and may be useful for clinical inference and rehabilitation.

Several researchers have moved in that direction by separately considering a variety of variables which relate to CT task performance. Royce, Yeudall, and Bock (1976), in a continuation of the research of Aftanas and Royce (1969), addressed the task of identifying the brain correlates of cognitive factors in a factor analytic study in which CT subtest performance was analyzed separately. CT subtests loaded on three different factors. Subtest IV, which requires identification of the clockwise numbered quadrant of the design that is discrepant, loaded moderately (.40) on a perceptual organization factor, which also included most elements of the TPT, the Trail Making Test, parts A and B, and WAIS Block Design and Object Assembly, among other tests. This factor was interpreted as comparable to the Perceptual Organization factor identified by Aftanas and Royce.

All CT subtests also loaded on two factors identified as Halstead Abstraction I and II. Subtests V and VI comprised Halstead Abstraction I, with a lesser loading for Subtest VII (-.45), which is regarded as a memory subtest involving items from the previous six subtests. Subtests III and IV comprised Halstead Abstraction II, again with a lesser loading for Subtest VII (-.39). Royce et al. (1976) suggested that this multifactorial loading of CT subtests reinforced previous suggestions regarding the multidimensionality of abstract, problem-solving behavior (Haynes & Sells, 1963). That interpretation is reinforced by different correlation patterns between each of these factors and brain-damage sites. Royce et al. reported that factor scores on Halstead Abstraction I correlated with damage throughout the right hemisphere, whereas Halstead Abstraction II correlated with damage restricted to the frontal and parietal regions of the right hemisphere. This distinction supported suggestions that Halstead Abstraction I represented more factorially complex abilities than Halstead Abstraction II. Assumptions regarding the multidimensionality of the CT were reinforced by Holland and Wadsworth's (1976) findings that while none of five concept formation scores, including CT total errors and CT subtest scores considered separately, could discriminate schizophrenic from brain-damaged subjects, subtest IV minus subtest V did significantly discriminate between groups (although at a level well short of that required for clinical usefulness).

Appraisal

The CT has been described as a measure of complex concept formation, which has been defined in the following terms:

> It requires sophisticated ability in noting similarities and differences in stimulus material, postulating hypotheses which appear reasonable with respect to recurring similarities and differences in the stimulus material, testing these hypotheses with respect to positive or negative reinforcement . . . and the ability to adapt hypotheses in accordance with the reinforcement accompanying each response. (Reitan, 1967)

Halstead defined this ability as "abstraction," and factor analysis revealed that the CT loaded more highly on this factor than any other test in his battery. Simmel and Counts' (1957) careful item-by-item analysis of patients' responses confirmed that this process appears relevant for *essentially* correct responses, but not necessarily for *incidentally* correct and incorrect responses. As a result, they noted, performance on the CT does not necessarily reflect abstraction or concept-formation skills of particular subjects.

While the discriminant validity of the CT has been well-established, there has been no follow-up on Simmel and Counts' suggestion that individual variations in response determinants may well be clinically significant. Reitan (1958) sought to establish that both brain-damaged and nonbrain-damaged groups utilized similar cognitive processes in approaching the task presented by the CT, but his results may well have been obtained if one or both groups predominantly utilized concrete counting responses, rather than abstract principles, as he supposed. Criterion-related validity studies, using brain lesion sites and age and intelligence as criteria, and factor-analytic studies have confirmed that CT performance is sensitive to the integrity of the cerebral cortex, but the question of the nature of cognitive processes called upon in performing the task presented by the CT has produced much conflicting information. Research findings have generally suggested that the CT taps nonverbal problem-solving skills and so varies with Performance IQ (or Perceptual Organization ability), though CT performance is more sensitive to cerebral dysfunction than is Performance IQ. Results of studies have indicated, however, that skills assessed by the CT are essentially equivalent to those measured by Performance IQ and so are redundant in the neuropsychological evaluation. Still others have found the CT to correlate more highly with Verbal IQ. Other unresolved questions surrounding the CT relate to whether the relationship with Performance IQ is a function of the Perceptual Organization skills tapped by both measures, or if it relates to the distinction between previously learned skills and adaptive problem-solving skills not calling upon previous learning, or some interaction between these two skill domains. Questions have been raised, too, concerning the "odd" relationship of the CT to Perceptual Organization skills, wherein the latter are clearly lateralized to the right hemisphere but CT performance is not.

These inconsistencies have typically been resolved in the literature by describing the CT as a measure of *complex* nonverbal problem-solving skills, or as *multidimensional*. While this is clearly the case for group results, it may be less true in individual cases. As Simmel and Counts (1957) noted, individual patients' performances may be characterized by different cognitive features:

> Perhaps the behavior of some subjects is determined more strongly by the perceptual characteristics of the stimulus than that of others. Some subjects might be more affected by what they have learned on an immediately preceding group of items and may have special difficulty in rejecting what was, but is no longer, rewarded. Still other subjects may find it particularly difficult to free themselves of specific procedures which are practiced in everyday life and which have in the course of their long history achieved a considerable degree of automaticity. Still others may be more victimized by the mental sets induced by the explicit or implicit characteristics of the surroundings, e.g., the apparatus, the response keys, etc. Some subjects may have special difficulty in rejecting a dichotomous response tendency. For some, the variation of basic stimulus figures may be disorganizing, while others may profit from the changes and thereby attain principles of greater generality. (p. 156)

The relevance of these individual differences for the cognitive behavior of individual subjects needs to be more fully explored, as suggested by Simmel and Counts.

The Wisconsin Card Sorting Test

The Wisconsin Card Sorting Test (WCST) is often regarded as similar to the CT with respect to the cognitive constructs measured by each. Both are purported to assess "abstract thinking," "concept formation," "conceptual flexibility," "problem-solving," and "new learning" (Strang, 1983). Both the CT and WCST were developed from common roots, but followed different paths of development (Perrine, 1984). Unlike the CT, much of the research related to development of the WCST focused on determination of the response determinants in contrast to clinical validation. A variety of quantitative process scores were developed to describe subjects' performance. Only after testing procedures and scoring protocols were established within the context of concept formation theory was the WCST applied to clinical uses.

WCST Development

The Wisconsin Card Sorting Test also combined many features from the Weigl (1941), Goldstein-Scheerer (1941), and Vigotsky Tests (Hanfmann & Kasanin, 1936) sorting tests employed by Goldstein and coworkers to assess abstract thinking, with the difference that it permitted objective, quantitative scoring (Grant & Berg, 1948). Berg (1948) observed that neurologically intact rhesus monkeys at the University of Wisconsin Primate Laboratory responded to shifts of positive and negative stimuli with no observable clues other than a change in stimulus object rewarded, but brain-lesioned monkeys lost their ability to follow shifting problems, although they were able to solve discrimination problems which did not require shifting. Consequently, she sought to apply this methodology to assess human abstraction and shift of problem-solving set. Unlike the CT, the WCST was developed from the context of concept formation theory using nonclinical subjects. Berg devised 60 response cards, which pictured one to four geometric figures (stars, crosses, triangles, or circles), of different colors (red, blue, green, or yellow). Thus, each card might be sorted or categorized in terms of three dimensions: shape of stimuli, number of stimuli, and color of stimuli. Four stimulus cards were also created: (a) one red triangle; (b) two green stars; (c) three yellow crosses; (d) four blue circles. Each of these stimulus cards were also exactly reproduced in the response cards. Thus, response cards might match stimulus cards on only one dimension, on two dimensions, or all three dimensions.

Fifty-one psychology students at the University of Wisconsin were used as subjects. They were instructed, "I want you to put these cards into four groups, underneath the ones lying on the table. I will tell you whether you are 'right' or 'wrong.' " If the subjects asked for more direction, instructions were repeated, "I will tell you whether you are 'right' or 'wrong.' " The initially correct category was chosen arbitrarily in advance, and after each subject sorted five successive cards correctly, the "correct" category shifted without explanation, so that now the previously correct percept received feedback of "incorrect" and a different category received feedback of "correct." When again the subject obtained five successive correct responses to the new percept, the category shifted to a third principle without warning. This continued until the subject had completed nine categories.

Results revealed three different patterns of behavior for test subjects. Slightly less than one third of the subjects, Group A, readily deduced the nature of the task and were able to verbalize the rules without difficulty. Approximately 40% of the subjects, Group B, became confused when previously correct sorting principles received "incorrect" feedback, and they showed perseveration to the former principle, even though it was no longer effective, and were unable to precisely verbalize the rules of the task. Nonetheless, these subjects were able to complete all nine categories. Slightly under one third of the subjects, Group C, showed extreme perseveration or extreme variability in their responses, sometimes continuing to sort to the previous principle for more than 100 trials. Testing was discontinued for subjects in this group due to time limitations. The difference in mean errors between Groups A and B was not significant until the fourth category, reflecting the greater learning curve of Group A. In contrast, the difference between Group C and Groups A and B was significant by the second category.

These suggestive findings were further explored by Grant and Berg (1948) in an investigation concerning the influence of reinforcement strength on shift of set. The card sorting task was modified somewhat in this study. Response cards were increased to two sets of 64 cards, representing all possible combinations of color (C), form (F), and number (N). Additionally, the initially correct sorting category was predetermined to be color, whereas previously it was randomly selected. This was followed by number, then shape, in the following sequence, C, F, N, C, F, N, and testing was discontinued after completion of six categories. Cards were presented in standardized order within sets as well, so that no like color, form, or number was presented consecutively.

Psychology students at the University of Wisconsin again served as subjects, and were divided into seven groups of approximately 20 students in each group. Number of consecutively reinforced trials, ranging from three through eight and ten, was varied for each group. It was hypothesized that increased number of reinforced trials would decrease likelihood of shifting set. Responses were classified as (a) correct; (b) errors; (c) perseverative responses; or (d) nonperseverative errors. A perseverative response was defined as one which would have been correct in the immediately preceding category[3]. These responses were usually, but not always, errors and represented failures to shift solution.

[3]If responses were ambiguous (matched the stimulus card on more than one dimension, one of which was correct in the immediately preceding category) and were both preceded and followed by a perseverative error, it was counted as a perseverative response. For example, if the following responses were obtained, C, C, C, CF, C, CNF, C, F, F, F, F, F, after shifting from color (C) to form (F), scoring would be psv error, psv error, psv error, psv correct, psv error, psv correct, psv error, correct, correct, correct, correct, correct.

Contrary to expectations, increased number of consecutively reinforced trials did not decrease likelihood of shifting set; rather, the opposite result was obtained: Increased reinforcement increased the likelihood of shifting set when the previous response was no longer reinforced. This difference reaches statistical significance for categories five and six, when groups were combined into less reinforced and more reinforced groups. This effect was even more pronounced in a subsequent study by Grant and Cost (1954), where the number of consecutively reinforced trials was increased to 5, 10, 20, and 40, presented in the following order: C, F, N, C, F, N. Grant and Cost noted, however, that the distributions were markedly skewed and recommended the use of the following score transformation for purposes of group comparisons: $t_x = R(x+.5)$, where t_x is the transformed score and x is the original frequency score.

Subsequently, Grant, Jones, and Tallantis (1949) investigated the comparative difficulty of the various dimensions -- color (C), form (F), and number (N) -- as an influence on what was called for the first time the Wisconsin Card Sorting Test. It was hypothesized, based on Heidbreder's (1948) and Heidbreder and Overstreet's (1948) research concerning transformations in concept formation, that form would be the easiest category attained, followed by color, then number. The procedures described above were followed, except that the order of correct concepts was varied, with all 24 sequences utilized in which no concept was consecutively repeated and all concepts were utilized prior to repetition. Shifts occurring without explanation following ten consecutive correct responses. Contrary to expectations, results revealed that number was the concept most easily obtained, followed by form, and color was most difficult. However, none of these differences reached statistical significance, although a significant relationship was demonstrated between ease of acquisition of the number concept and perseverative errors on the following category. Thus, the number concept was more easily acquired and tended to perseverate longer. Grant et al. (1949) suggested that these results may have been due to perceptual features of the cards, in which the configuration of figures rather than number per se was the salient characteristic[4].

Grant (1951) confirmed that suggestion by alternating the standard systematically arranged response cards and a set of specially created response cards in which the figures were arranged unsymmetrically, so that they did not perceptually match the stimulus cards. Order of administration was C, F, N, C, F, N. Scoring procedures were revised somewhat so that the initial perseverative response, following a shift, was not scored, nor were ambiguously correct responses. Additionally, correct responses did not include responses in the criterion run of 10 consecutive correct responses, so that the larger the number of correct responses, the lower the subject's efficiency in shifting from the previous percept to the new. These results revealed that unsystematic configurations were more difficult for subjects to acquire than systematically arranged figures which matched the configuration of the stimulus cards, and when asked to explain "what they were doing and had been doing" (p. 28), 38% of the subjects confirmed they had sorted to color, form, and *configuration*.

These results suggested that the conclusions of Grant et al. (1949) concerning relative difficulty of the three sorting categories were probably due to this artifact. Consequently, Grant and Curran (1952) replicated the design of that study, except that they used the unsystematic response cards and the scoring

[4]On cards with one figure, the design is centered on the card; for two figures, the designs are placed diagonally; for three figures, the designs form an equilateral triangle; and for four figures, the designs form a square shape.

conventions introduced by Grant (1951). Whereas in the initial study the number concept was the most easily acquired and subsequently the most likely "perseverate to" principle, in this study the number concept was the most difficult to acquire as the initial sorting principle, and form was acquired most readily. However, once acquired, number remained the most likely "perseverate-to" principle, and showed no appreciable decrease following the third category, as did perseveration to color and form. These results correspond more closely to Heidbreder's (1948) and Heidbreder and Overstreet's (1948) hypotheses than did the earlier results. The authors suggested that abstraction as measured by this test is a multidimensional construct which cannot be accounted for by simple unidimensional quantitative scores.

Unlike the clinical tradition of the CT, the research tradition out of which the WCST developed was concerned with delineating the response determinants which related to the acquisition of concepts. Thus, WCST researchers addressed the question of determining the process by which subjects decided, "This is [X]; this is not [X]." Of primary concern in this regard were those determinants which influenced perseveration of responses which were no longer reinforced. This behavior has long been defined as cognitive rigidity (Pishkin & Williams, 1977) as it is assumed that subjects will shift hypotheses following disconfirmation through absence of reinforcement (Levine, 1966; Restle, 1962; Trabasso & Bower, 1966). However, Matthews and Patton (1975) demonstrated with a group of normal college students that the degree of success experienced with a given hypothesis prior to disconfirmation is related to failing to shift subsequent to disconfirmation. Matthews and Patton suggested that an hypothesis may not be immediately abandoned simply because it leads to an error, but instead must rely on what is most probable.

Perseveration might be considered a function of the degree to which irrelevant dimensions are reinforced. Since the 64 response cards represent all possible combinations of color, form, and number, correct responses might match the stimulus card on only one dimension, an *unambiguous* match, or on two or three dimensions, an *ambiguous* match. The standard set of cards includes 24 cards which can match on only one dimension, 36 cards which, depending on placement, may match on one or two dimensions, and 4 cards which can only match on all three dimensions simultaneously. Thus, up to 40 responses for each set of cards may receive reinforcement of ambiguous dimensions, and the subject may perseverate due to difficulty in ruling out the irrelevant features.

Gormezano and Grant (1958) investigated the influence of intermittent reinforcement of irrelevant features on concept acquisition, by rearranging unsystematically configured response cards into sets of 48 in which a given dimension might be ambiguous 0%, 25%, 50%, or 75% of the time. Criterion performance was successful completion of the first two categories, color and number, within the 48-card set. Subjects were divided into four groups (N=40). The first group received a set of cards in which, when sorted correctly to color, none of the cards matched for number; the second group received a set of cards in which 25% when sorted to color also matched on number. In the third group, 50% of the cards were ambiguous as to number. In the fourth group, 75% were ambiguous as to number. Each of these groups was then divided into four groups (N=10), and the process was repeated with sets of cards which varied in ambiguity as to form, when cards were correctly sorted to number. Transformed frequency scores were used for comparison of groups. Results revealed that increased reinforcement of irrelevant dimensions made acquisition of the relevant concept more difficult, but increased reinforcement of the number dimension when it was irrelevant had no effect on acquisition of the number concept when it became relevant. These findings were interpreted in terms of concept formation theory (Bourne, 1965; Bourne, 1974; Brown, 1974; Buss, 1953; Buss, 1956; Dominowski, 1973; Dominowski, 1974; Kendler, Glucksberg, & Keston, 1961; Peterson & Colavita, 1964; Wetherick & Dominowski, 1976).

Up to this point, WCST research was related to investigating response determinants which influenced concept acquisition and cognitive rigidity. Methods of administration and scoring protocols which had developed were directed toward those research goals, but normative data had not been provided nor was any information available relative to discriminant validity with clinical groups. These concerns now became paramount in the further development of the WCST.

Localization of Brain Damage

Sorting tasks from which the WCST was developed have long been associated with cerebral dysfunction. As noted previously, McFie and Piercy (1952a; 1952b) found that left brain-damaged patients were more impaired on Weigl's (1941) Test than right brain-damaged patients, but no differences for aphasics vs. nonaphasics within the left hemisphere group. In contrast, De Renzi et al. (1966), employing a modified version of Weigl's Test, also found more impaired performance by left-brain damaged patients, but only those with aphasic symptoms. They postulated that these discrepant findings might be attributable to different proportions of patients with prefrontal lesions in their left brain-damaged group in comparison with those studied by McFie and Piercy, as Goldstein and co-workers had long contended.

While the Weigl Test was similar to the WCST in many respects, it could only be scored dichotomously as "pass" or "fail" and was often considered too easy to discriminate subjects effectively. Milner (1963) was the first to study the utility of the WCST in localization of brain lesions. She compared the performance of 33 patients with frontal lobe surgical excisions for control of focal epilepsy with the performance of 61 patients with surgical excisions at a variety of other sites. Most patients were tested both preoperatively and postoperatively (N=71), but the remainder (N=23) were tested only postoperatively, in some cases several years after the surgery. Surgeon's records were obtained in each case to obtain precise localizing information. The frontal lobe group was further subdivided into lesions of the dorsolateral aspect (superior) and the medial orbital aspect (inferior). All of the surgical excisions for the dorsolateral group were unilateral (14 nondominant right hemisphere and 11 dominant left hemisphere, although Broca's speech area was spared in all cases). The nonfrontal group consisted of 34 patients with excisions of the dominant left hemisphere, 30 patients with excisions of the nondominant right hemisphere, 3 patients with excisions of the nondominant left hemisphere, and 2 patients with bilateral surgical excisions. Groups were equivalent with respect to age and IQ, both preoperatively and postoperatively.

Milner employed the unsystematic response cards (Milner, 1963) developed by Grant (1951) and considered a more difficult task than the standard systematic response cards. Scoring categories included the usual scores, and she employed the Grant and Cost (1954) frequency score transformation for data analysis, but Milner redefined the scoring of perseveration. Grant and Berg (1948) defined perseveration as those responses which would have been correct in the previous category, preventing the scoring of perseveration in the first category, resulting in misleading scores for subjects who failed to complete the first category or who displayed significant difficulty in the first category. Consequently, Milner included those responses in the first category which represent repetition of the patient's initial concept, which was easily discerned since the first eight cards in her response deck could only be matched unambiguously, but she implied that she counted only perseverative *errors*, which would exclude those correct responses defined as perseverative by Grant and Berg. Additionally, she established the practice of terminating administration after completion of 128 cards, or completion of six categories, whichever came first.

Results revealed that preoperatively, WCST performance significantly discriminated the dorsolateral frontal group from all other localization groups, except the small parieto-temporo-occipital group (N=5), with respect to total errors (p<.05) and perseverative responses (p<.01). Postoperatively, the differentiation was even more marked. The dorsolateral frontal group made significantly more errors than all other groups (p<.001), achieved fewer categories (r<.001), and made more perseverative errors (p<.001), although there were no group differences with respect to nonperseverative errors. Significantly, while surgical excision further deteriorated the performance of the dorsolateral frontal group, the nonfrontal group displayed an opposite trend. The performance of patients who were tested only postoperatively displayed the same pattern. None of the dorsolateral frontal patients completed more than two categories, while 51 of the 69 control subjects did so.

Milner observed that while age and intelligence were controlled in this study, size and extent of lesions were not and may have accounted for these effects. Consequently, the performances of the five dorsolateral frontal patients with the smallest surgical excisions were compared with 12 nonfrontal patients with the largest lesions, including some with complete lobectomies. Preoperatively, these groups showed no significant differences, although the dorsolateral frontal group tended to perform more poorly. Postoperatively, the performance of the nonfrontal group was unchanged, despite the extent of tissue removed, but the performance of the dorsolateral frontal group further deteriorated, with significantly more total errors ($p <.05$), more perseverative responses ($p <.001$), and fewer categories sorted ($p <.05$) than the control subjects. Milner concluded, therefore, that locus of lesion was the relevant factor bearing upon impaired performance of the dorsolateral frontal group in comparison with other brain-damaged groups.

Laterality comparisons were nonsignificant, in contrast to the findings by McFie and Piercy (1952a) and De Renzi et al. (1966), but Milner noted that left hemisphere excisions were consistently smaller than those on the right, and suggested that those findings may indicate greater sensitivity of the left hemisphere. Subsequently, Milner (1971) reported findings that confirmed that suggestion.

An earlier study by Teuber, Battersby, and Bender (1951) found results discrepant from those reported by Milner on a modified sorting test, but the extent of modifications appears to have changed task demands significantly. Teuber et al. employed only 60 stimulus cards in which the figures were laid out linearly, and "shifts from one principle of sorting to another [were] enforced after every tenth trial" (p. 421). Subjects were also forewarned that a change may occur (Milner, 1963). The authors did not specify criteria of performance, nor indicate the order of presentation. Test subjects included 131 World War II combat veterans with penetrating missile wounds, four to seven years post-injury, who were divided into three groups on the basis of site of wounds of entrance. These groups were identified as Anterior, Intermediate, and Posterior. Forty control subjects with peripheral nerve injuries due to gunshot wounds were also tested. Results revealed that the posterior subjects performed more poorly than any other group, although only mean error scores were reported. In view of the modifications to the sorting task used in this study, and questions concerning extent of injuries not reflected in wounds of entrance, these results cannot be considered comparable to other studies, even though Teuber et al. is often cited in the literature as contradictory to more frequent findings regarding localization.

Milner (1963) suggested that her results implicating the dorsolateral surface of the frontal lobes in impaired set shifting behavior were consistent with findings with lower primates that dorsolateral frontal excisions but not ventromedial (orbital) excisions impair delayed alteration and delayed response behavior (Mishkin, 1957; Pribram, Mishkin, Rosvold, & Kaplan, 1952). However, Drewe (1974) contended that Milner's findings were surprising in view of increased perseveration by primates on nonspatial reversal tasks followed orbital but not dorsolateral excisions. Consequently, she sought to replicate Milner's findings with a more heterogenous group.

Drewe tested 91 patients with unilateral localized lesions of varying etiology. Patients were divided into four mutually exclusive localization groups, right frontal (RF), left frontal (LF), right nonfrontal (RC), and left nonfrontal (LC). Groups were equivalent with respect to age, sex, chronicity, and nature of lesion. She did not specify which card set was utilized, systematic or unsystematic, but since the Grant and Berg manual (undated) was followed with respect to scoring and administration, it is likely the systematic response cards were utilized. Drewe added the scoring category, Number in Maximum Classification, primarily to reflect perseverative responses in the initial category since these were not reflected in Grant and Berg's definition of perseverative responses. This scoring category was defined as the number of cards sorted into each category (color, form, number, unique), excluding criteria runs of 10.

Results reflected very complex relationships between type and number of errors made, which varies from group to group, but on the whole these findings are consistent with those of Milner, despite substantial differences in procedure and scoring. Patients with frontal lesions achieved fewer categories and showed greater perseveration than nonfrontals, and there was no significant difference with respect to nonperseverative errors. However, in contrast to Milner, they did not make more total errors, and a greater range of variability in categories achieved was evident. These patients performed more like Milner's preoperative patients than her postoperative patients; in fact, patients after frontal lobectomies performed better than patients with other frontal lesions.

In general, left frontal patients performed more poorly than right frontal patients, consistent with Milner's findings, and those of previous researchers, although they did not differ on perseverative errors. However, differences in definition may account for this finding. Drewe found that Number in Maximum Classification was a more sensitive measure of overall performance than perseverative responses, and on this measure left frontals were more impaired. Some discrepancy with respect to anatomical locus within the frontal lobe was evident, but Drewe concluded that her findings may be consistent with Milner's actual data, although the area of maximum sensitivity is the medial surface of the frontal lobes that are critical rather than the dorsolateral convexity.

The most significant findings which emerged from this study, however, relate to the complexity of relationships between scores on the WCST and the site of damage. Drewe concluded: "All scores on the WCST do not necessarily measure the same ability or disability, nor does the same 'impaired' score in different patients indicate a necessarily similar deficit" (1974, p. 168).

Robinson, Heaton, Lehman, and Stilson (1980) noted that few other neuropsychological measures effectively localize focal frontal impairment, and so sought to replicate findings reported by Milner (1963) and Drewe (1974) that perseverative responses on the WCST are useful in localizing brain damage. They contrasted performance of a normal control group of paid volunteers (N=123) on the WCST, with the performance of seven brain-damaged groups: (a) right frontal (N=13); (b) left frontal (N=10); (c) right frontal plus focal right nonfrontal (N=11); (d) left frontal plus focal left nonfrontal (N=12); (e) right nonfrontal (N=9); (f) left nonfrontal (N=14); and (g) diffuse (N=38), controlling for severity of impairment as reflected in the Average Impairment Rating (Russell et al., 1970) on the HRNB, as well as age and education. Like Drewe, Robinson et al. apparently used systematically configured response cards but followed Milner's definition of perseveration rather than Grant and Berg's, thus apparently counted only errors as perseverations, including those in the first category.

Results revealed significant differences among groups with respect to WCST perseverative responses, in which covariance due to age, education, and severity of impairment was controlled ($p <.02$). The performance of all brain-damaged groups was significantly impaired in comparison with normals, except that of the left nonfrontal group. Comparison of the four frontal groups with nonfrontal groups confirmed previous findings with respect to the sensitivity of the WCST to frontal lobe impairment, but no difference was found for frontal groups in comparison with diffusely impaired patients. This should not be surprising, Robinson et al. noted, since the frontal lobes comprise approximately 40% of the brain. These findings suggest that the perseveration score on the WCST is quite sensitive to general cerebral dysfunction, except possibly when focal impairment does not extend to the frontal areas.

These findings were further confirmed by linear discriminant analyses using perseverative responses, age, education, and Average Impairment Rating on the HRNB to classify subjects on the basis of lesion site. Perseverative responses entered the analyses first and AIR failed to enter at all, yielding a correct classification rate of 68%, suggesting that perseverative responses, with age and education, account for variability in the AIR due to severity of impairment. Consequently, a second discriminant analysis was computed to determine if perseverative responses on the WCST would enhance the discriminant validity of the AIR. The 12 individual variables that comprise the AIR were entered into the analysis, with age, education, and perseverative responses. Again, the perseverative response score entered the analysis first, and this combination yielded a correct classification rate of 85%, when the population base rate of frontal lobe impairment was set to .67. Discriminant efficiency deteriorates with more extreme population base rates, however.

With respect to lateralization, however, these findings are discrepant with those of Milner and Drewe. The combined right frontal groups were significantly more impaired than the left frontal groups, which could not be accounted for by differences in severity, as assessed by the AIR. However, Milner's (1963) caution concerning the multidimensionality of determinants on WCST performance should be noted in this regard, particularly since performance on this measure was assessed only on the basis of a single score.

Variability in task complexity and other procedural differences in administration and scoring, in addition to the usual differences in subject selection, has confounded interpretation of differing results across these studies, but surprisingly the differentiation of frontal vs. nonfrontal patients has proven to be very robust, although substantial variability has been evident with respect to individual scores. Nelson (1976) addressed the issue of the effects of task complexity directly by significantly simplifying the task, even in comparison with the systematic response cards, in creating what she referred to as the Modified Card Sorting Test. Nelson removed all ambiguous cards from the response set (apparently using systematic cards), thereby eliminating examiner uncertainty concerning what concepts patients were using to sort, and removing patient uncertainty concerning which concept was being reinforced by the examiner. This left 24 cards in each 64 card set, and Nelson combined two sets of cards for a complete response set of 48 cards. Administration instructions were also revised as follows:

> Here we have four key cards [examiner points to stimulus cards]. I want you to sort these cards [indicating the response cards] under the key cards according to certain rules: but the whole point of the test is that I shall not tell you what the rule is. I want you to find that out by trying out different rules and each time I shall tell you whether it's right or wrong. Now, go ahead and try to find out the rule. (p. 316)

Following these instructions, whichever category the patient chose was scored "correct," and subsequent responses were scored accordingly. In contrast, in standard administration procedures color is the initial category, followed by form and number in that order when the criterion run of 10 consecutive correct responses was achieved, with the cycle being repeated a second time for a total of six categories. Nelson also completed six categories, but after six consecutive correct responses, stated: "The rules have now changed; I want you to find another rule" (p. 316). This instruction was repeated each time the patient completed six consecutive sorts to whichever new category was selected. Testing was discontinued when six categories had been completed in this fashion, or all 48 cards had been placed. Nelson reported that she utilized Milner's definition of perseverative responses (PM), although her testing procedure would have excluded perseverative responses in the initial category. Additionally, Nelson devised a new scoring of perseveration (PN) which consisted of repetition of an incorrect response despite feedback that it was incorrect.

Despite these substantive differences in administration and scoring, comparison of the performance of 64 brain-damaged inpatients, divided into four groups on the basis of cerebral quadrant of lesion site, with the performance on 46 control subjects, including 32 inpatients with extra-cerebral lesions and 14 normal subjects, Nelson found that frontal patients made significantly more perseverative errors (both PM and PN) than nonfrontal groups, after the records of 13 patients (lesion group not specified) who successfully completed the Modified Card Sorting Test were excluded from analysis. However, the lesions groups barely missed showing significant differences with respect to total errors and categories completed (p =.056), and no difference was evident between right and left lesion groups, either frontal or nonfrontal. Milner (1963) observed that patients often were able to verbalize which responses were correct and which incorrect, although they seemed unable to conform their behavior with their observations, suggesting that task difficulty may be less salient in influencing outcome on the WCST than other cognitive process factors. Nelson's findings lend credence to that interpretation.

Nonetheless, Heaton (1981) noted that comparison of results across studies was handicapped by the variety of procedural variations employed in these studies. Consequently, he sought to standardize materials and procedures in order to better focus attention upon the neurological and cognitive variables bearing upon test performance, rather than upon methodological variables. He included the subjects used by Robinson et al. (1980) in a normative study which provided the basis for standardization of administration and scoring procedures and materials. He employed two sets of the systematic configuration response cards (64 cards in each set) devised by Grant and Berg (1948), but the specified order of cards within each set was that first used by Robinson et al. in which five of the initial eight cards are ambiguous, including the first card, unlike the Grant and Berg set in which the first eight cards are all unambiguous. Instructions were specified as follows:

> This test is a little unusual, because I am not allowed to tell you very much about how to do it. You will be asked to match each of the cards in these decks to one of the four key cards. You must always take the top card from the deck, and place it below the key card you think it matches. I can't tell you *how* to match the cards, but I will tell you each time whether you are right or wrong. If you are wrong, leave the card where you've placed it, and try to get the next card correct. Use this deck first, and then continue with the second deck. There is no time limit on this test. (p. 19)

A standard scoring sheet was also devised, and the following order of correct concept matching was specified: C F N C F N. Designated scoring categories included:

1) Total Errors
2) Total Correct (including criteria runs and initial correct responses following shifts)
3) Categories Completed (all criteria runs of 10 correct responses)
4) Perseverative Responses
    a) those which would have been correct in the previous category;
    b) those responses prior to completion of the first category which match the first incorrect unambiguous response;
    c) ambiguous responses are also scored as perseverative responses if they are both preceded and followed by unambiguous perseverative responses;

d) within any category, three consecutive incorrect unambiguous responses will establish a new "perseverate-to" principle for that category, although ambiguous perseverative responses may be interspersed among those three consecutive perseverative responses, which may be scored beginning with the second of the three consecutive perseverative responses.

5) Perseverative Errors (those perseverative responses which are also incorrect)
6) Nonperseverative Errors (errors which are not perseverative responses).

Several optional scoring categories were also defined:

1) Percent Perseverative Errors (Perseverative Errors ÷ Total Responses)
2) Trials to Complete First Category
3) Percent Conceptual Level Responses (runs of three or more consecutive correct responses ÷ Total Responses)
4) Failure to Maintain Set (number of runs of five to nine consecutive correct responses)
5) Learning to Learn (for patients who complete at least three categories, mean sum of the differences of percent error score for consecutive categories with at least 10 trials).

Heaton's normative study employing these procedures included comparisons of WCST performance of 208 subjects with structural cerebral lesions, confirmed by neuroradiological techniques, and 150 normal controls, including subjects studied by Robinson et al. (1980). Brain-damaged subjects were divided into Diffuse (N=94), Nonfrontal (N=35), Focal Frontal (N=43), and Frontal Plus (N=36). Normal subjects were paid volunteers without histories of neurological illness, head injury, or substance abuse. All subjects were administered the WAIS-R, HRNB, and WCST. WAIS-R Full Scale IQ and HRNB Average Impairment Rating (AIR) were used as measured of general neuropsychological impairment.

Results revealed that the total brain-damaged group were more impaired on all measures (including WAIS-R FSIQ and AIR), except Failure to Maintain Set, which was relatively rare in all groups. Since Focal Frontal and Frontal Plus groups were not significantly different on demographic or neuropsychological variables, these groups were combined for data analysis. This combined frontal group performed significantly more poorly than the nonfrontal group on WCST Total Errors, Perseverative Errors, Percent Perseverative Errors, Perseverative Responses, and Percent Conceptual Level Responses. Groups did not differ on Categories Completed, Nonperseverative Errors, Trials to First Category, Learning to Learn, or Failure to Maintain Set. Thus, the primary differentiation between groups was in terms of perseveration. The Diffuse group performed much like the Frontal group.

The most efficient cutoff score for predicting brain damage in this population is a perseverative response score of 19 or above, which correctly classified 74% of the brain-damaged group and 72% of the nonbrain-damaged group. In comparison, inferring a base rate for this population based on group composition reported, calling all subjects brain-damaged, would have resulted in 58% correct classification. A slightly higher cutoff score (20 or above perseverative responses) achieved 71.1% accuracy in discriminating focal frontal from focal nonfrontal brain-damaged subjects, but with better accuracy in classifying focal frontal patients (79.8% vs. 51.4%). The cutoff score for perseverative errors was 13 and 16 for discriminating brain damage and focal frontal lesions, respectively, yielding approximately equivalent efficiency to a cutoff based on perseverative responses. A cutoff score based on total errors above 24 was also approximately equivalent in classifying focal frontal impairment but was not useful in identifying general cerebral dysfunction. Other measures were not useful in predicting brain damage or focal frontal involvement in individual cases, although in most cases the correlation between AIR and these variables was substantial for the brain-damaged group (.47 to .61), except for Learning to Learn (-.18) and Failure to Maintain Set (-.05). Nonperseverative errors were most strongly associated with overall neuropsychological functioning for the normal group, but the strongest association for the brain-damaged group was with measures of perseveration.

Construct Validity

Strang (1983) devised eight new hypothesis testing scoring categories to investigate the normative expectation that intact persons will change concept strategies when they receive "incorrect" feedback, and persist with hypothesized concepts when the feedback is "correct." Four responses to examiner feedback regarding accuracy are possible: (a) Wrong-Shift, defined as shifting to a new concept when the examiner responds "incorrect"; (b) Wrong-Stay, defined as persisting with the same concept when the examiner responds "incorrect"; (c) Right-Shift, shifting concepts when the examiner responds "correct"; and (d) Right-Stay, persisting with the same concept when the examiner responds "correct." However, since up to 62.5% of all responses may be ambiguous, it is often not possible to ascertain whether a subject has shifted to a new concept. Consequently, each score was computed twice: once, assuming that the patient persisted with the previous concept when responses were ambiguous, and again assuming that the patient had shifted. In discriminant function analyses, with significance level set very conservatively, of the 32 most promising WCST variables, only the transformed hypothesis testing score, Stay-Behavior (assuming stay), entered the prediction equation, with population base rates set at values ranging from .10 to .90, yielding a correct classification rate of .81. Additional post hoc analyses also supported the usefulness of these hypothesis testing scores.

In order to better investigate the relationship among these scores, factor analysis was performed, employing 30 WCST variables. A three-factor solution was obtained, with the first factor representing cognitive perseveration, the second factor representing what appeared to be slow or partial learning, and the third factor representing successful strategies. Strang hypothesized that underlying this proliferation of scores reflecting problem-solving approaches to the WCST may lie a single simple dimension.

More recent results (Anderson et al., 1991), however, have suggested that WCST performance is multidimensional, consistent with findings by Robinson et al. (1980) that perseverative responses reflect general cerebral dysfunction. Anderson et al. tested 91 patients with single, focal brain lesions, as established by CT and MR images, in contrast to previous studies. All subjects were tested at least three months post CVA (N=71) or neurosurgical resection for tumor or seizures (N=20). Patients with a history of psychiatric disorders, head injuries, substance abuse, or neurological disease were excluded from the study. Patients were divided into frontal groups (including dorsolateral, mesial, and orbital frontal lesions), frontal plus (lesions extending posterior to rolandic areas), and nonfrontal. All patients were administered the WCST and WAIS-R. The WCST was administered and scored according to standard procedures (Heaton, 1981), yielding the usual scores: number correct, number of errors, perseverative responses, perseverative errors, nonperseverative errors, and categories completed. None of these measures revealed significant differences, or even a trend toward significant differences, between frontal and nonfrontal groups. The cutoff scores which best discriminated the groups were the same as those found by Heaton, but these correctly classified only 62% of the subjects (42% of nonfrontals). Anderson et al. considered the possibility that systematic differences in lesion size may have contributed to these findings, but found differences in lesion size did not result in systematic differences within the frontal groups. Differences between left and right hemisphere groups were also not found, nor were differences evident between frontal groups divided on the basis of anatomical divisions within the frontal lobes. Finally, no consistent anatomical correlates of specific patterns of WCST performance were found. Anderson et al. concluded that while both human and animal studies clearly demonstrate frontal lobe mediation of cognitive abstraction and set shifting tasks, "performance on a multifaceted cognitive task such as the WCST will necessarily involve the coordinated interaction of multiple and separate brain regions" (pp. 15-16). Regional blood flow studies (rCBF) confirm that while prefrontal lobe regions are activated by performance on the WCST, so are several other brain areas (Weinberger, Berman, & Zec, 1986).

Anderson et al. (1991) observed, with Heaton and others, that the WCST may be most useful in assessment of multiple cognitive constructs, including ability to abstract information from relevant stimuli, ability to shift cognitive set, and perseverative tendencies, regardless of anatomical correlates.

Moderator Variables

While individual studies have demonstrated significant effects in differentiating the WCST performance of various clinical groups, replication of these effects has often been impeded by a variety of methodological variables which makes comparisons across groups and settings difficult. Moderator variables, including age, education, and IQ differences, are among those variables which may influence test research outcome, but this issue has received less attention relative to the WCST than for the CT.

Age and Education Correlates

Developers of the WCST employed groups of college students in their research, who were uniformly young and well-educated, with only one exception. Berg (1948) administered what became known as the WCST to a group of 22 older students, ranging in age from 58 to 73 years, reporting that none of these students progressed beyond the first category shift. Loranger and Misiak (1960) also found that older normal subjects were very impaired on the WCST. Only two of her 50 aged female subjects between the ages of 74 and 80 were able to complete all six categories and the mean number of categories completed was only 1.86 (S.D. = 1.54). Correlation with age was not significant (r=-.16), however, due to the restricted age range of her sample. Years of education was also not significantly correlated with WCST performance (r=.21). Misiak and Loranger (1961) found somewhat higher correlation between age and WCST perseverative errors (.34) in a group of elderly subjects, combining the aged females tested by Loranger and Misiak with 36 males, age 68 to 80.

In studies comparing the performance of clinical subjects with other groups, most studies have used restricted age and education ranges, which may have modulated relationships that might have been discerned in more heterogenous groups. Exceptions to this general rule often have not provided data sufficient to analyze the relationship. For example, Drewe (1974) concluded that age was significantly associated with WCST performance only when other disabilities were not present, but she provided no data to support that conclusion. Nelson (1976) anecdotally observed that the standard WCST procedure was often too difficult for older subjects, but also provided no data concerning the effect of her modification on performance by older subjects. Similarly, Tarter and Parsons (1971) asserted that the significant differences observed in their study between college students and hospitalized controls was the result of age and education differences, but provided no support for that conclusion.

Robinson et al. (1980) provided more adequate data analysis of age and education variables. They reported significant correlations of age with perseverative responses (r=.23, r<.05) but nonsignificant correlations with education (r=-.19, r<.10) for brain-damaged subjects. For normal subjects, the relationships were reversed: education correlated significantly with perseverative responses (r=-.17, r<.10) but age did not (r=-.25, r<.01). In attempting to classify focal brain-damaged subjects into frontal vs. nonfrontal groups, both age and education entered all discriminant function analyses, as variables contributing to discrimination equations. The proportion of the variance accounted for by these variables was not reported, however.

Heaton's follow-up normative study, based in part on these same data, provides much more complete analyses of the influence of age and education correlates. With the exception of Learning to Learn and Failure to Maintain Set scores, which proved to have little predictive value in any of the analyses, age showed mild to moderate correlations with all other variables, ranging from .12 with nonperseverative errors to -.33 with categories achieved for the total brain-damaged group. Correlations for the normal group were comparable, except for Trials to First Category, which was essentially unrelated. The relationship with education was much more modest for the brain-damaged group, but roughly equivalent for the normal group.

The opposite pattern was obtained by King and Snow (1981). In their analysis of patients referred for neuropsychological evaluation, patients who were ultimately found to be brain-damaged by independent diagnoses showed roughly equivalent correlations for education and age (.23 and -.29, respectively), but these moderator variables were essentially unrelated for patients found not to be brain-damaged.

Analysis of these demographic variables for the normal group using four age groups (under 40 years of age, 40-59 years of age, 50-59 years of age, and over 59 years of age) and three levels of education (less than 12 years, 12-15 years, and over 15 years) revealed that age and education variables do not interact. Significant main age effects were obtained for six of ten WCST scores, but in each case this was due to the poor performance of the oldest group (over 59 years), who achieved mean scores on all variables within the impaired range. Significant main education effects were also obtained for six of the ten scores, but in this case these results were primarily due to the excellent performance of the best educated group. However, in several instances the performance of the least educated group fell just into the impaired range. These results suggest that interpretation of WCST performance should be amended somewhat for older persons, and performance of poorly educated subjects should perhaps be interpreted conservatively, despite the attenuation of education effects accompanying brain damage.

Somewhat different findings were obtained by Heinrichs (1990), who compared the performance of psychiatric patients and brain-damaged patients. Results reveal that his sample was relatively impaired on all measures, with education showing stronger relationships with WCST variables (.44 to .54) than age (.34 to .37). Heinrichs concluded that these relationships appear to vary with respect to populations under consideration.

IQ Correlates

A majority of studies have also attempted to equate IQ levels of research subjects and controls, but have seldom evaluated the relationship of IQ correlates to WCST for clinical vs. control groups. Fey (1951) was the first to report IQ as a covariate. She found no significant relationship between Full Scale IQ on the

Wechsler-Bellevue (Form I) and categories completed on the WCST for either the schizophrenic group or the normal control group. The correlations between categories completed and W-B subtests also were not significant. She concluded that intelligence is not a major factor influencing performance on the WCST, and noted parenthetically that the success of some "feebleminded" children on the WCST is supportive of that conclusion.

Milner (1963) reported group IQ scores for her brain-damaged subjects, but failed to specify what test was administered and did not report any IQ/WCST analyses, other than comparing IQ loss for frontal vs. nonfrontal groups preoperatively and postoperatively. Drewe (1974) specified that her subjects were administered a short WAIS and reported that groups did not differ on Verbal IQ, but patients with left hemisphere lesions obtained significantly higher Performance IQ scores than patients with right hemisphere lesions ($r<.01$) and patients with nonfrontal lesions also scored significantly higher on the WAIS Performance Scale than patients with frontal lesions ($r<.05$). Drewe suggested that this may have accounted for some of the results, but did not analyze this relationship further.

Loranger and Misiak (1960) examined the interrelationship among several tests of nonverbal intellectual function, including the WCST, for elderly females. Using concepts completed as a measure of success on the WCST, they found moderate correlations with all measures administered, including Porteus Mazes ($r=.59$), Primary Mental Abilities Reasoning (timed: $r=.46$; untimed: $r=.36$), and Ravens Progressive Mazes ($r=.65$). A follow-up of this study (Misiak & Loranger, 1961) employing an enlarged sample found these relationships to be somewhat weaker, however. Centroid factor analysis of these measures yielded only one factor, which they termed a general intellectual factor, on which WCST perseverative errors loaded .35.

Robinson et al. (1980) incorporated WAIS Full Scale IQ into the Average Impairment Rating on the HRNB, which was covaried as a measure of severity in their analyses. They did not analyze the relationship of IQ with WCST performance independently. However, in discriminant analyses intended to facilitate differentiation of focal frontal brain damage, the AIR score representing Full Scale IQ was the only variable which did not enter the prediction equation, suggesting that the variability represented by that measure was accounted for by other measures. The normative follow-up of this study by Heaton (1981), however, revealed moderate relationships for most WCST variables with Full Scale IQ, except for Failure to Maintain Set, Learning to Learn, and Nonperseverative Errors, for the brain-damaged group (.32 to .45). Interestingly, the correlations are more modest for the normal group. In contrast, Heinrichs (1990) found a much more substantial relationship between WCST performance and WAIS-R Full Scale IQ in his more impaired group of subjects (.69 to .77). Pendleton and Heaton (1982) found similar moderate correlations in both brain-damaged and normal control groups between Full Scale IQ and WCST perseverative responses.

Only one study has investigated the relationship of WAIS subtest variables and scale scores with WCST performance. Strang (1983) compared the performance of her brain-damaged group with her normal group on all WCST variables which have been reported in the literature and WAIS subtests and scale scores. Results reveal that for the pooled data from both samples, most of the major WCST scores commonly used yielded Pearson correlation coefficients ranging from r=.40 to r=.60. No difference was observed between the strength of relationships between Verbal IQ and Performance IQ scores, but a trend toward more frequent significant correlations was observed for Performance Scale subtests than for Verbal Scale subtests.

However, a very different pattern of relationships is observed when brain-damaged and nonbrain-damaged groups are considered separately. For the brain-damaged group, none of the WCST scores are correlated with WAIS summary scores, except for Categories Completed (r=.41 with FSIQ). The strongest subtest score relationships are with Comprehension, which was correlated with every major score above r=.40. Only slightly less frequent correlations were found for Digit Symbol and Block Design. For Digit Symbol, only Total Correct (excluding criterion runs and signal trials) and Total Trials were not significantly correlated; for Block Design, only Categories Completed and Total Correct were not significantly correlated. Similarities and Picture Arrangement demonstrated secondary relationships with major WCST variables. This pattern of relationships was not expected, particularly the very consistent correlations with Comprehension.

A somewhat different pattern of relationships was evident for the normal group. For this group, WAIS summary scores were consistently related to performance on major WCST variables, and less association was observed with Comprehension. However, the relationship with Block Design and Digit Symbol was even more evident. These results were interpreted as suggesting that among brain-damaged patients the use of practical problem-solving skills, most especially involving common sense, may be prerequisites for these measures of success on the task presented by the WCST, but with intact cerebral functioning, nonverbal reasoning skills take on greater importance.

Taken together, these results show marked variability in the association between moderator variables, particularly intelligence, and WCST performance. These results might be expected on the basis of variable constructs assessed by the WCST in various populations and emphasize the need for further clarification relating to the cognitive processes involved in WCST performance. This might best be accomplished by directly relating performance on the WCST to CT performance characteristics.

Comparison of the CT and WCST

Despite their obvious differences, the CT and WCST share many similarities. Both are complex measures which require concept formation based on identification of common elements in a visual array, sufficient flexibility to form and then reject hypotheses based on an ambiguous pattern of reinforcement, the ability to profit from feedback concerning the accuracy of responses, the ability to shift response sets, as well as the ability to maintain response sets when appropriate (Pendleton & Heaton, 1982; Strang, 1983). Only a few studies, however, have directly compared these measures.

King and Snow (1981) compared the performances of 150 subjects referred for neuropsychological evaluation on the WCST and CT. Patients were divided into two groups on the basis of discharge diagnoses not derived from neuropsychological test data. The WCST was administered according to Milner's (1963) procedures, except that all patients completed all 128 cards, rather than terminating testing after completion of six categories. Dependent measures compared were categories completed on the WCST and total errors on the CT. Partial correlation coefficients were computed, controlling for the effects of age and education on both variables. The correlation obtained for the brain-damaged group was $r=-.44$ ($p <.001$) and for the nonbrain-damaged group $r=-.42$ ($p <.001$). The correlation for the pooled data from both groups, again controlling for age and education effects, was $r=-.55$ ($p <.001$). At most, this degree of correspondence accounted for approximately 30% of the variance in both measures. King and Snow concluded that these are not measures of identical constructs.

Similar results were reported by Pendleton and Heaton (1982), who also found 30% shared variance in a correlational study comparing the CT and WCST. Subjects were 207 brain-damaged patients who had been referred for neuropsychological evaluation at the University of Colorado Health Sciences Center. Lesion site was confirmed and type was independently confirmed by neuroradiological and neurosurgical data. Brain-damaged patients were divided into three subgroups: (a) Frontal, including those with lesions extending into nonfrontal areas; (b) Nonfrontal; and (c) Diffuse. A fourth group included 150 normal paid volunteers. Performance of each group was compared with respect to CT total errors and WCST perseverative responses. Correlations for all groups were roughly similar, ranging from $r=.48$ to $r=.58$, and all were significant at $p <.001$, reflecting approximately 30% shared variance. Moreover, Full Scale IQ correlated more highly with the CT ($p <.01$) than for WCST perseverative responses ($p <.05$). Diagnostic accuracy of these measures was also compared. Results reveal that the WCST is slightly more sensitive to frontal than to other cerebral lesions, but the WCST was more accurate in classifying nonfrontal and diffuse groups. However, these differences did not reach statistical significance. This same pattern of results was obtained when these measures disagreed with respect to classification: The WCST was more accurate in classifying frontal lobe patients whereas the CT was more accurate in classifying nonfrontal patients, and these results did reach significance (p <05). Together these results suggested that these measures assess somewhat different cognitive abilities, despite their similarity. Consequently, Pendleton and Heaton recommend continued use of both tests.

Despite these differences, Brandon and Chavez (1985) found that when both tests are administered in the same battery, presenting the CT first facilitates performance on the WCST, as measured by the relationship between CT total errors and WCST perseverative response and total error scores. The authors postulated those results were due to the more diversified range of abstract concepts represented in the CT, but this effect suggests considerable similarity in the cognitive processes assessed by each task, at least for this population of normal functioning college students.

Rothke (1986) investigated the extent to which differences in set shifting cues given in the instructions on each test may have suppressed correlations between them. A heterogenous group of 52 inpatients in a VA medical center were randomly assigned to one of four experimental conditions. In the first condition, subjects were administered the WCST first and both tests were given with set shifting instructions. In the second condition, subjects were also given the WCST first, but neither test was given with set shifting instructions. In the third condition, subjects were given the CT first, and both tests were given with set shifting instructions. In the fourth condition, subjects were given the CT first, and neither test was given with set shifting instructions. Thus, in each condition, one test was given with standard instructions and the other with modified instructions. When set shifting cues were given, patients administered the WCST were informed that the sorting principle might change from time to time, while the CT was given in the standard manner, being informed at the beginning at each new subtest, "The principle might be the same as it was in the last one or it might be different." When no set shifting cues were given, the WCST was given in the standard manner, where the only information given was feedback concerning the accuracy of responses. Similarly, the CT was given without any mention of subtests or a change of principles. Unlike the results reported by Brandon and Chavez (1985), Rothke did not find a significant effect for order of presentation of these tasks. However, a significant effect for cueing conditions was demonstrated ($p < .001$). Eliminating set shifting cues for the CT did not substantially affect performance, but WCST perseverative responses were much reduced by providing set shifting cues. The author concluded that for this relatively impaired population, these tasks measure substantially different cognitive skills. Rothke postulated that mental set shifting ability is of paramount importance on the WCST, while the CT measures more complex skills.

Strang (1983) compared performance on all WCST variables reported in the literature, in addition to her new hypothesis-testing scores, and CT performance, including total errors and subtest scores. Since WCST variables are typically frequency scores which reflect the greater number of trials required with poorer performance, they may be expected to vary with CT error scores, which are also frequency scores. Partialing out this source of methodological variance resulted in much attenuated relationships between WCST variables and CT error scores. This resulted in significant relationships in the brain-damaged group only for WCST perseveration related scores, although total errors was also significantly correlated with performance on CT Subtest VII. The most consistent relationships were with CT Subtests V and VII and total errors, with less consistent correlations with Subtest IV, and no significant relationship with

Subtest I-III and VI. For the normal group, the only significant correlations using transformed WCST scores were with proportion of Matching to Form and Reinforcements to Form with CT total errors and Subtest IV. It should be remembered that form is the easiest concept for most subjects and the most frequent perseverate-to principle. These relationships for normals appear to reflect ability to shift flexibly from this intuitive matching principle.

The relationship between WCST variables and the CT is further attenuated when the effect of moderator variables was partialed out. For the normal group, partialing age or FSIQ, or any combination of age, FSIQ, and education reduced the significant correlations to none. For the brain-damaged group, significant correlations remained when any one moderator variable was partialed out, but partialing age and FSIQ simultaneously reduced all correlations to less than $r=.27$ (7.29% common variance).

When the hypothesis shifting scores were considered in relation to CT performance for the total group, it was evident, as might be expected, that Wrong-Shift behavior is associated with better CT performance, and Wrong-Stay is an unsuccessful strategy. However, controlling for demographic variables also reduces these relationships to nonsignificance in the brain-damaged group. In contrast, for the normal group, when controlling for all three demographic variables simultaneously, Shift-Behavior remains significantly correlated with performance on Subtest VI ($p < .01$).

Overall, this very complex study suggested that WCST performance may be viewed from a variety of perspectives, which may reflect multiple cognitive processes, but relationships between these measures and CT performance, when controlling for variance due to age and intellect, are nonsignificant. This may be due to the complexity of cognitive processes reflected by CT error scores, which may not vary consistently in predicted directions, or these measures may simply reflect different cognitive processes.

Conclusions

Although the CT and WCST derived from common sources, in some senses their development may be described as mirror images of each other. Reitan (1966b) referred to the CT as a continuing long-term experiment, and added:

> We are still administering the same tests in the same way we did more than a decade ago. Of course, standardization of testing procedure has been necessary in order to be in a position gradually to accumulate enough subjects for study of specific neurological variables. As a result, we have not been able to modify or manipulate the test battery in order to learn experimentally what the tests measure or the particular requirements of the tests which might be most sensitive to cerebral dysfunction. The approach has been oriented toward subdivision and analysis of the independent neurological variables while the dependent variables (psychological measurements) have been held constant. (p. 163)

In contrast, the WCST has been developed out of seemingly endless methodological variations. Heaton (1981) described 32 different scoring methods which have been utilized, and almost as many variations of materials and procedures. Only a few of these variations have been studied systematically; consequently, comparability across studies has been lacking. Unlike the CT, where a single score conceals complexity of constructs, with the WCST a proliferation of scores has impeded investigation of the constructs they each reflect and the interrelationships among them. Heaton's (1981) standardization appears to have brought some measure of consistency to research and clinical practice involving the WCST. What is now needed is a means of relating the performance profile obtained on WCST variables to the broader pattern of patients' neuropsychological functioning, particularly that represented by CT performance, in order to relate test findings to adaptive behavior in a natural environment.

CHAPTER III

METHODS AND PROCEDURES

Modified Administration Procedures

While previous research has firmly established both the CT and WCST as sensitive measures of complex problem-solving skills, concept formation, and cognitive flexibility, the specific cognitive skills assessed by each task have not been addressed, except by correlating error scores on one test with error scores on the other test. Bond and Buchtel (1984) argued that while the issue of identifying the cognitive skills assessed by neuropsychological measures is an important one, correlational studies which compare error scores on one test with those of another may be uninformative. In particular, they observed that correlations such as those which have been reported for the CT and WCST may reflect the unreliability of the tests rather than differences in abilities measured by these tests. This, however, is difficult to ascertain for several reasons. A considerable practice effect has been found to be characteristic of both measures, and those effects appear to vary from population to population (Matarazzo, Matarazzo, Wiens, Gallo, & Klonoff, 1976; Matarazzo, Wiens, Matarazzo, & Goldstein, 1974; Matarazzo, Matarazzo, Gallo, & Wiens, 1979), which renders test-retest reliability estimates essentially meaningless. Split-half reliability estimation for these measures is also not useful, because subsequent items are not equivalent in terms of difficulty or independent (Simmel & Counts, 1957), nor is the last half comparable to the first half. Since no alternate form exists for either measure and would be very difficult to construct, reliability estimates are not presently available.

Bond and Buchtel (1984) also described other problems attendant upon correlational analyses of error scores in describing common constructs measured by each of these tests. Among these are uncertainty concerning what processes given scores by specific individuals reflect. This is an especially important consideration when cognitive strategies may vary considerably. Bond and Buchtel recommended a process tracing approach, or a "think aloud" approach, as one means of minimizing these obstacles. This technique involves asking subjects to verbalize the rules or the reasoning underlying their responses.

Such an approach has the advantage of providing supplementary data that are useful to help clarify the hypothesis-testing process employed by subjects as they seek to identify relevant features of the stimuli comprising these tasks. Thereby, it is possible to differentiate types of errors subjects make in responding. It is not likely that subjects "tell all they know" in explaining the reasoning underlying their responses; rather verbal explanations are often incomplete and ambiguous. Yet, this technique permits more accurate description of cognitive processes characteristic of particular patients' problem-solving strategies. For example, some patients fail on these tasks due to an overly concrete perceptual focus, while others respond incorrectly due to failure to maintain cognitive set, while still others are deficient in idea fluency, or display perseverative responding. From the perspective of designing and implementing cognitive rehabilitation strategies, it is important to differentiate among these different types of errors, to the extent that these approaches to problem-solving tasks are characteristic of individual patients, or groups of patients.

This verbalization approach to assessment represents a modification of standard testing procedures, and is not without methodological problems of its own, but it has been used with some success in other contexts. Reitan has recommended this approach on an informal basis as a means of gathering additional information concerning patients' problem-solving strategies on the CT (Reitan & Wolfson, 1985), and it has been used also with concept formation tasks (Bourne, 1965).

However, this technique has not been previously reported in the literature. It is not known whether asking subjects to verbalize their reasoning may have changed the nature of cognitive processing involved in task performance. This question was addressed in the current study by comparing results of testing collected under standard administration conditions with test results of subjects who were asked to verbalize reasoning. Noncomparability of mean error scores would suggest that adding a verbalization component to these tasks has substantially modified the nature of the cognitive processes called upon in performing these problem-solving tasks. Moreover, if the strength of association between types of error scores between groups is significantly different statistically, it could concluded that the tasks presented under modified administration conditions (verbalization of reasoning) were not the same tasks presented under standard administration conditions (reasoning not verbalized). If asking patients to verbalize their reasoning on these tasks is found to substantially alter the nature of the task, obviously conclusions that were based on data collected under modified instructions could not be generalized to data collected under standard instructions.

Cognitive Process Variables

A primary concern of neuropsychological evaluation, particularly since the advent of more sophisticated neurodiagnostic procedures, is the prescription of treatment goals and the design of remediation strategies. This requires more specific definition of deficits than statements of "impaired problem solving skills," or "impaired cognitive flexibility," or other such general statements. However, evaluating performance of higher cognitive functions by counting errors on the CT and WCST fails to provide such specific information. What referral sources seek to know is, "What does this patient do that impairs performance on this problem-solving task and how can that behavior be changed?" Since the CT and WCST have been found to be among the most sensitive measures of brain damage, it is reasonable to believe that patients' performance characteristics contain the information necessary to respond to those questions. If that is so, patients' verbalization of the reasoning underlying their responses may well access that data. For that information to be useful, however, it must be translated into reliable scores, and those scores must reflect relatively consistent performance characteristics, independently of age and IQ covariates.

Asking patients to verbalize the reasoning underlying quantifiable responses introduces the problem of how one patient's qualitative responses may be compared to other patients' qualitative responses. Some anecdotal reports are available for the CT in which patients have been asked to verbalize their reasoning (Reitan & Wolfson, 1985), but these responses have been interpreted subjectively, which does not allow systematic comparisons across patients or groups of patients.

Two studies have more systematically addressed the issue of identifying item-by-item response determinants on the CT, however, and these may serve to provide some guidance in defining relevant response parameters. Simmel and Counts (1957) concluded that some form of counting was the most persistent response determinant for their subjects. They described counting responses based upon the number of stimuli or elements in the configuration, counting responses based upon perceptually similar stimuli, counting responses based upon perceptual dissimilarity, counting responses based upon spatial continuity of both similar and dissimilar stimuli, and counting responses based upon dominant perceptual features. Simmel and Counts also noted that subjects who are confused as to the correct principle may have perseverated to an obviously incorrect principle, or reverted to a preferred (incorrect) principle when the stimulus configuration changed or a dominant perceptual feature appeared, even though they may have been more recently reinforced for employing the correct principle.

Perrine (1984) referred to the concept formation literature (Bourne, Dominowski, & Loftus, 1979) for a cognitive process model which might help clarify the response tendencies described by Simmel and Counts. Following Bourne (Bourne et al., 1979), he differentiated between *attribute identification,* defined as characteristics of a stimulus configuration which convey information about membership or nonmembership in a class, and *rule-learning,* which describes the process whereby attributes are related logically. Perrine described the relationship between these components as hierarchical, having both concrete (attribute identification) and abstract (rule learning) stages. Concept formation tasks may involve only the concrete stage, or both the concrete and abstract stages. Using tasks devised to assess performance on these stages separately as independent variables, Perrine concluded that the WCST is primarily an attribute identification task (concrete) whereas the CT is more a rule learning task (both concrete and abstract). However, even though the relationship between these concept formation scores and CT subtest and WCST scores was significantly greater than the relationship between CT subtest and WCST scores from a statistical perspective, the improvement in predictive power was generally marginal from a clinical viewpoint. Further, even these findings may have been influenced by the greater similarity of the attribute identification tasks to the WCST than the CT, and vice versa for the rule learning tasks.

An assumption underlying such approaches is that the cognitive processes employed in problem-solving or concept formation tasks is "task-driven," wherein the strategies employed are dependent upon the nature of the task. However, as Bond and Buchtel (1984) have noted, individuals may select a common response by utilizing a variety of processes. Measuring only the end product of those processes may conceal as much as it illuminates.

What is needed is an approach which can measure the processes employed by subjects as they arrive at a solution to the task presented to them. Asking patients to verbalize their reasoning as they problem-solve may be the most direct means of accomplishing this goal. For the WCST those responses consist only of the perceptual attribute (utilizing Perrine's terminology) identified by the subject in determining his or her response. Thus, verbalized WCST responses serve to provide a less ambiguous means of scoring the responses than is true for conventional administration. However, for the CT, devising a coding system which permits comparison of verbalized responses across items and across subjects is a formidable task.

Perrine's Model of Concept Formation

Perrine's (1984) concept formation model for describing the process whereby patients select a response from the choices available to them on the problem-solving tasks represented by the CT provides a means for differentiating subjects' reasoning. In terms of this model, the attribute identification stage of response selection may be viewed as the perceptual component of that process, while the rule-learning stage appears more conceptual. Thus, when patients verbalized their reasoning on these tasks, attribute identification occurred when subjects discriminated features in the stimulus upon which their response was based, while the rule-learning component consisted of the significance subjects attributed to those features. For example, on Subtest III of the CT, many subjects chose "3" as the correct response because "three are not solid." In that instance, "openness" (nonsolidity) constituted the attribute identification portion of the response, while counting (elements that are not solid) comprised the rule-learning component.

However, the task of applying this model to subjects' verbal responses is somewhat equivalent to devising a valid and reliable coding system for the Rorschach Inkblot Technique. Different individuals describe the same reasoning differently, and on occasion similar language may describe very different reasoning processes. Similarly, one person may utilize different language to describe similar responses or may fail to use language which differentiates among different kinds of responses. For example, in Subtest IV, one person may respond, "Three because that part is missing," while another person, referring to the same perceptual features, may respond, "Three because that's where the line is." Even more problematic is when such language shifts occur within the same protocol. The question arises as to whether these differences in language relate to different cognitive processes. If so, coding subjects' reasoning on the CT may not be feasible since utilizing the language in subjects' descriptions of how stimuli relate to the number responses given has proven to be inherently unreliable, and therefore invalid.

The Attribute Identification Score

The alternative which was utilized in this study was the convention of identifying the perceptual attributes of stimuli to which the language used by subjects referred. For example, on Subtest III, subjects may have chosen "1" as the correct response because "One is different." In this instance, "different" referred to the solidity of one portion of the stimulus; consequently, the perceptual attribute would be "Solid" rather than "Different." Similarly, subjects may have chosen "2" as the correct response because "Two are alike" (referring to the fact that two stimuli have four sides). In that case, the perceptual attribute would be "Form." In this fashion it is possible to avoid the ambiguities inherent in attempting to code multiple ways of describing "difference" in a consistent fashion. In contrast, if the patient had stated in the first instance, "solidness," but the number response was "3," the perceptual attribute would be "Open," because the language of the verbal response clearly referred to the nonsolidity of three elements in the stimulus. Table 2 lists the Attribute Identification codes identified for the purposes of this study and brief definitions for each. These codes may be viewed as describing where the subject was looking in formulating the response, or what part of the stimulus was included in the response. This approach brings a greater degree of standardization to the task and permits increased reliability in coding responses.

An additional problem encountered is that due to the complexity of CT stimuli, a response may refer to two or more perceptual attributes, or a verbal response may be ambiguous as to which attribute takes precedence. Often it is

Table 2

Definitions of Attribute Identification (Concrete) Codes

1. Gestalt (G)	Those responses in which the perceptual focus is on the stimulus configuration as a whole a. does not differentiate between solid and dotted or inferred (missing) lines b. dissimilar elements are treated as equivalent.
2. Color (C)	Responses in which elements in the stimulus configuration are differentiated on the basis of color. a. Where achromatic colors are used to describe configural elements ("the long white one"), score for solidity not color. b. An exception to the above rule will be when responses are based upon the achromatic color of elements which are unlike with respect to solidity.
3. Form (F)	Responses based upon the form or shape of concrete nongeometric figures. If shape or form is used descriptively rather than to differentiate elements in the configuration ("3 small squares" to refer to dotted lines), form would not be scored.
4. Element (E)	Responses in which elements in the stimulus configuration which are not perceptually unique or distinct are arbitrarily differentiated from other perceptually similar elements.
5. Size (Z)	Responses in which elements in the stimulus configuration are differentiated on the basis of size.
6. Solidity (S)	Responses in which elements in the stimulus configuration are differentiated on the basis of solidity, or where the perceptual focus of attention is on elements outlined by a solid line.
7. Open (O)	Responses in which the openness of the stimulus or stimuli is identified as the perceptual focus of attention, in contrast to stimuli which are identified as "solid." Includes responses focusing attention on broken or dotted lines.
8. Missing (M)	Responses using inferred lines or components which are considered missing as the object of rule formulation. Includes dotted lines which are described as missing.
9. Back-ground(B)	Responses which include the background area beyond the borders of the stimulus.
10. Direction (D)	Responses which consist of elements representing a remote point or location pointed to or identified by another element in the stimulus.

not possible to differentiate which of these alternatives is true. Consequently, in the interest of increased reliability in coding responses, a decision tree approach was adopted for coding perceptual attributes. This approach establishes a hierarchy of coding rules, ordered in such a way that the most relevant or highest order perceptual attributes are listed first. (See Appendix E for the full model.) Use of this approach permitted achievement of a satisfactory level of test-retest reliability in recoding perceptual attributes. (See Procedures section for reliability coefficients.)

The Conceptual Abstraction Score

Perrine's model described a two-part process of concept formation in which Attribute Identification referred to the identification of relevant perceptual attributes of stimuli and Rule Learning referred to logical relationships between attributes, leading to hypothesis formation. With reference to the CT, Attribute Identification refers to identification of the relevant features of stimuli upon which the rule is based, whereas Rule Learning refers to the logical relationship between the attribute identified and the number response chosen. For example, on Subtest III, the Attribute Identification task is to discern that the stimulus which is different is the relevant dimension, while the Rule Learning task is to discover that the position of the different stimulus corresponds to the correct number.

However, during the course of this study, it was discovered that a three-stage coding system was required to reliably differentiate verbal responses:
(a) Concrete: subjects selected a perceptual feature which they believed relevant; (b) Abstract: the perceptual feature or features were organized conceptually to create a pattern; (c) Rule: this pattern was related to the number response given. For example, in Subtest IV, the quadrants are numbered in a clockwise direction, and the principle requires identifying the quadrant of the stimulus which is

missing or different. The Concrete (Attribute Identification) task is identification of the portion which is not there or which is different from the rest of the stimulus; the Abstract conceptual task is to perceive this attribute as one quadrant of the whole stimulus; the Rule Learning task is to select a number corresponding to the position of the missing or different quadrant when ordered in a clockwise direction.

Table 3 lists the Abstract scores identified for the purposes of this study and brief definitions for each. As was the case with the Concrete (Attribute Identification) score, reliability in coding was enhanced by utilizing a hierarchical decision-tree approach in coding the Abstract score. (See Appendix F for the full model.)

The Rule-Learning Score

Rule Learning is the third element in this model of concept formation. This is identical to Perrine's Rule-Learning model and refers to the logical relationship between the Abstract Pattern and the number response given for the item. For example, in the following response, "1 because there's 1 object" (referring to a solidly outlined square figure in Subtest V), the Attribute Identification score would be "S" because the stimulus included in the response is outlined in a solid line; the Abstract Pattern score would be "I" because the attribute identified occupies space corresponding to multiple quadrants of the stimulus and has been combined into a single percept; and the Rule Learning score would be "C" because the number response results from counting the Abstract Patterns identified. In contrast, if the response had been, "4 because there's 4 lines" (using the same stimulus), the Attribute Identification score would remain the same; the Abstract Pattern score would be "Q" because the attribute identified occupies

Table 3

Definitions of Abstract Pattern Codes

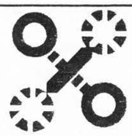
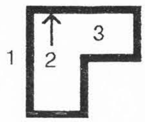
1.	Category Within (W)	Responses which combine elements with like attributes into a unitary set, then differentiate elements within the set.
2.	Category Between (K)	Responses which combine elements with like attributes into a unitary set, and apply the rule to the set without further subdivision.
3.	Location (L)	Responses based upon the position of concrete attributes, other than those responses which reflect the correct position of a specified quadrant.
4.	Unlike (U)	Responses which consist of one unique element which is described or treated as different than other elements in the stimulus array.
5.	Alike (A)	Responses which consist of elements described as "alike" or the "same" or differentiated from elements not included in the response on the basis of unique attributes.
6.	Number (N)	Responses which consist of attributes described as a shape representing a number value.
7.	Form (F)	Responses which consist of attributes described as a concrete, nongeometric shape other than a number.
8.	Missing (M)	Responses which consist of attributes which are physically present but are described as "missing."
9.	Quadrant (Q)	Responses which consist of attributes which occupy space corresponding to quadrants or fourths of the stimulus, or multiples thereof.
10.	Integrated (I)	Responses which consist of attributes occupying space corresponding to two or more quadrants which have been combined into a single percept.
11.	Division (D)	Responses which consist of attributes occupying space which does not correspond to quadrants of the stimulus, or multiples thereof.

space corresponding to quadrants of the stimulus, or multiples thereof, and the Rule Learning score would also be unchanged because the number response still results from counting the Abstract Patterns identified. (See Table 4 for additional scoring examples.)

In this fashion, it is possible to assess the hypothesis testing behavior of individual subjects, in response to confirmation or disconfirmation of hypotheses. Bourne (1965) reported that efficient hypothesis testing behavior was characterized by formulation of all-encompassing hypothesis initially, and

Table 4

Examples of Scoring Cognitive Process Variables

STIMULUS	NO.	VERBAL RESPONSE	C	A	R
	1	"One straight line"	F	D	C
	3	"There's the page, then this, and this"	B	D	C
	2	"There are 2 kinds"	G	K	C
	1	"The big box is pointing to No. 1"	D	Q	L
	4	"4th one is a different shape"	F	U	L

modifying elements of the hypothesis one at a time, until the correct principle is discovered. The coding system described here permits that type of analysis with respect to the CT. Table 5 lists the Rule Learning scores identified for the purposes of this study and brief definitions for each. As was the case with the Concrete and Abstract scores, reliability in coding was enhanced by utilizing a hierarchical decision-tree approach in coding the Rule-Learning score. (See Appendix G for the full model.)

Summary and Ratio Scores

The primary scores described above were transformed to a variety of ratio and summary scores which comprise the Cognitive Process variables. These scores reflect a smaller number of cognitive processes which are believed to be crucial for performance of problem-solving tasks. These include, but are not limited to,

Table 5

Definitions of Rule-Learning Codes

1.	Random (R)	Responses based on Guesses.
2.	Memory (M)	Responses based on a previous response without clarifying the reasoning.
3.	Counting (C)	Responses based upon counting Abstract Patterns identified.
4.	Location (L)	Responses corresponding to correct placement of a single quadrant or fourth of the stimulus.
5.	Location - Err (Le)	Responses corresponding to incorrect placement of a single quadrant or fourth of the stimulus.
6.	Proportion (P)	Responses based upon the proportion of the stimulus which was identified as the Abstract Pattern, when the number response given corresponds to the numerator of the fraction when the denominator = 4.
7.	Proportion - Err (Pe)	Responses based upon the proportion of the stimulus which was identified as the Abstract Pattern, when the number response given does not correspond to the numerator of the fraction or when the denominator $\neq 4$.
8.	Arithmetic (A)	Responses based upon a mathematical procedure other than simple counting of Abstract Patterns.
9.	Form (F)	Responses based upon the shape or form of the Abstract Pattern.
10.	Sequence (S)	Responses which are based upon a sequence other than described above.

the following: (a) ability to shift strategies when solutions are ineffective; (b) ability to maintain strategies when solutions are effective; (c) ability to generate alternative solutions; (d) ability to identify relevant perceptual attributes; (e) ability to apply abstract rules to variable perceptual stimuli. (See Appendix H for definitions of these variables.)

Hypotheses

If these or other similar cognitive processes predict performance on these tasks, the scores which have been derived from coding of patients' verbalized reasoning on the CT and WCST should form interpretable cognitive process factors, when entered into a factor analysis, which are at least moderately independent of age and IQ covariates. This may be stated in the form of the following research hypotheses:

1. Mean Total Error and Subtest Error scores on the CT and mean Error scores, Categories Completed, and Perseverative Responses on the WCST will not differ significantly (statistically) between a group of patients who verbalized the reasoning for their responses on the CT and WCST and a group of patients who did not verbalize their reasoning on the CT and WCST, and are comparable to the first group with respect to diagnosis, age and Full Scale WAIS-R IQ.

2. Correlations among CT Total Error scores, CT Subtest Error scores, WCST Error scores, WCST Categories Completed, WCST Perseverative Responses, WCST Nonperseverative Errors, and WCST Failure to Maintain Set will not differ significantly (statistically) between a group of patients who verbalized the reasoning for their responses on the CT and a group of patients who did not verbalize their reasoning on the CT, when both groups are comparable with respect to diagnosis, age and Full Scale WAIS-R IQ.

3. The relationships between coded variables derived from verbalized responses on the CT will produce three or more interpretable (cognitive process) factors, which are more independent of age, education, and IQ subtest and summary scores than CT error scores.

4. The relationships between coded variables derived from verbalized responses on the WCST (defined in Appendix I) will produce three or more interpretable (cognitive process) factors, which are more independent of age, education, and IQ subtest and summary scores than summary WCST scores (Categories Completed, Perseverative Responses, Nonperseverative Errors, Total Errors, Failure to Maintain Set).

5. The relationship between Total Error Scores and Subtest Error Scores on the CT and Perseverative Errors, Nonperseverative Errors, Categories Completed, Perseverative Responses, and Failure-to-Maintain Set scores on the WCST will not be statistically significant, when controlling for FSIQ, age, and education.

6. The relationship between interpretable factor scores from the WCST factor analyses (hypothesis 4) and each factor score derived from the CT factor analysis (hypothesis 3) will be statistically significant, when controlling for FSIQ, age, and education.

Purpose

In view of the need for more accurate description of the cognitive deficits that characterize the problem-solving behavior of patients with brain injuries or dysfunction, the purpose of this study was to devise a coding system to classify CT and WCST responses using a "think aloud" methodology, in order to achieve the following goals: (a) describe the cognitive strategies employed by individual patients in performing these tasks, and (b) differentiate the cognitive strategies assessed by the CT versus the WCST.

Objectives

The goals described above were addressed in terms of the following objectives:

1. To determine whether asking patients to verbalize their reasoning on WCST and CT responses substantially alters the nature of these tasks.

2. To devise a coding system derived from these verbal responses and determine if verbal responses obtained using this methodology can be reliably coded by trained examiners.

3. To determine the relationship between the cognitive processes reflected by scores on the WCST versus the cognitive processes reflected by scores on the CT.

Procedures

Subjects

Subjects receiving modified test instructions and those receiving nonmodified instructions (comparison group) were obtained from archival records in the Clinical Psychology Department at Iowa Methodist Medical Center (IMMC), a 710 bed privately-funded, nonprofit general medical center located in Des Moines, Iowa, affiliated with the only post-acute care rehabilitation center in central Iowa, Younkers Rehabilitation Center (YRC), which provides a full range of rehabilitation services for victims of stroke, traumatic injury, heart disease, neurological disorders, tumors, and a variety of other disabling conditions. In addition, IMMC offers inpatient and outpatient substance abuse treatment and inpatient psychiatric care. The Clinical Psychology Department has on staff two licensed neuropsychologists and receives a large number of referrals for neuropsychological evaluation from neurologists, physiatrists, psychiatrists, and other medical specialists associated with these treatment programs and others. Cognitive rehabilitation is also provided as an adjunct service by the Clinical Psychology Department, and in that connection since July, 1990, technicians providing neuropsychological testing have routinely asked patients to verbalize their reasoning on the Wisconsin Card Sorting Test and the Category Test, when these measures were administered and time permitted. These responses were recorded verbatim to supplement the clinical data available to the neuropsychologist to provide baseline data for treatment planning purposes.

Records of neuropsychological testing since July, 1990 were reviewed in order to identify patients whose charts contained verbalized records of both complete WCST and Category Test administration, and completed WAIS-R protocols. Charts of 132 such patients were identified. Additionally, patients were identified whose charts contained completed CT and WCST protocols, administered according to standard instructions (nonverbalized), and who also completed the WAIS-R. Charts of 88 patients who met these criteria were identified. However, patients in this comparison group were tested over a much wider time period for several reasons. First, most patients referred since July, 1990 received verbalized test administration on the CT and WCST for clinical reasons; second, staff changes occurred just prior to that time, and previous staff members apparently administered both of these test instruments in their entirety much less frequently than was true after July, 1990. As a result, testing records were searched to mid-1986 in order to identify a sufficient number of patients for the comparison group.

Confidentiality of these patients was protected by duplicating only those portions of the protocols which contained no personally identifying information, with the exception of date of testing, date of birth, and age. Duplicated protocols were coded with a sequential identifying number. At the same time, the Report of Neuropsychological Evaluation was consulted to obtain other data relevant to assessing comparability of the two groups, which was entered on a coding sheet identified by the assigned identification number. These data included sex, handedness, diagnosis, and medical information relevant to neuropsychological functioning (history of hospitalization for head injury, substance abuse history, treatment for neurological or psychiatric disorder, and learning disabilities). After these data were entered onto the coding sheet and protocols were photoduplicated as described above, original protocols were returned to the files. Coded protocols were then divided into the comparison group (nonmodified instructions) and experimental group (modified instructions).

Although patients whose protocols were selected for the purposes of this study did not sign informed consent forms, the designated custodian of the records, the Director of the Psychology Department at IMMC, on the recommendation of a five-member research committee appointed by the president of IMMC, and chaired by the research director, approved the use of these records for the purposes described, with the limitation that patient confidentiality would be protected, according to the procedures described above.

No attempt was made to ensure that patients whose protocols were selected for inclusion in this study were representative with respect to demographic characteristics of the general population or the population of patients referred to the Psychology Department at IMMC for neuropsychological evaluation. In fact, it is unlikely that these groups of patients are comparable with the general population in terms of these characteristics, since young males are at increased risk of head injuries, and closed head injury is the primary diagnostic classification for both groups of patients.

These groups of patients are also unlikely to be representative of all patients referred for neuropsychological assessment since the selection criteria (completion of the WAIS-R, CT, and WCST) excluded most severely impaired patients, patients with very poor frustration tolerance who could not persist in completing these difficult tasks, and aphasic patients who were unable to adequately verbalize responses on the Verbal Scale of the WAIS-R or verbalize their reasoning on the CT and WCST. As a result, it is likely that patients selected for inclusion in the study included a disproportionate number of patients with right cerebral hemisphere damage, since patients with left hemisphere damage are much more likely to be aphasic.

Further analysis of the influence of these variables on outcome was not attempted. Table 6 lists primary, secondary, and tertiary diagnoses (when available) for patients in each group.

It will be noted that a history of head injury was present in 73% of all patients, at least by patient self-report, although the reason for referral may have been another diagnosis. As a result, reliably differentiating among this group of patients on the basis of diagnosis is not practical. Consequently, diagnosis is not included as a variable or covariate in this study.

Since previous researchers have identified age and IQ, and to a lesser degree education, as covariates of CT and WCST performance, mean values of these demographic variables were computed for each group, and comparability of

Table 6

Diagnostic Classification Within Each Patient Group

DIAGNOSIS	EXPERIMENTAL GRP			COMPARISON GRP		
	1st	2nd	3rd	1st	2nd	3rd
Traumatic Brain Injury	88	6	3	59	4	
CVA	7	1		2		1
Toxic Exposure	3	1		3		
Brain Tumour	3			1		
Dementia					1	
Psychiatric	8	6	1	3	4	
Chemical Dependency	5	23	5	4	5	3
Seizure Disorder		5	2	1	4	1
Other Neurological	6	1		10		
Unspecified	7			4	3	

group means were assessed using *t* tests. Table 7 lists mean age, education, and IQ scores of patients by sex and handedness for each patient group and for both groups together. It will be noted that though the groups were comparable with respect to IQ and education, the age comparison revealed that the experimental group was significantly older ($p < .03$) than the comparison group.

Inspection of decade age ranges in the two groups revealed that the proportions of subjects in both extremes of the continuum were quite dissimilar for the two groups. (See Table 8.) Consequently, covarying age was necessary on all group comparisons.

Test Administration

For both of these groups of patients, testing was administered in a standard manner by qualified technicians under the appropriate supervision of licensed psychologists who are credentialed as clinical neuropsychologists. Each examiner has completed at least an M.A. or M.S. degree in psychology, but each group was tested by different examiners, with the exception of the small number of patients in the Comparison Group who were tested subsequent to July, 1990 and were not asked to verbalize reasoning on the CT and WCST. Among the

Table 7

Demographic Characteristics of Patient Groups

GROUP	N	AGE		ED		PIQ		VIQ		FSIQ	
					Mean (S.D.)						
Exp. Grp Total	132	*37.1	(14.3)	13.2	(2.7)	95.9	(14.1)	97.3	(13.6)	96.2	(13.5)
Comp. Grp Tot.	88	*33.1	(12.4)	12.8	(2.5)	97.6	(13.5)	98.4	(14.0)	98.1	(13.7)
Combined Tot.	220	35.5	(13.7)	13.1	(3.0)	96.5	(13.9)	97.7	(13.7)	96.9	(13.6)

*$p < .05$

reasons for employing standard administration for these subjects were time limitations, impaired expressive language skills, or assignment to an examiner who was reluctant to employ modified test instructions.

The Booklet Category Test (BCT) was substituted for the Halstead Category Test for both groups, even for those tested prior to 1990. The BCT (DeFilippis & McCampbell, 1979) consists of either one or two large loose-leaf binders containing reproductions of the CT stimuli on a black background to correspond to the smoked screen on which the CT slides are projected. Patients are presented a strip of paper on which numbers one through four have been printed, and rather than responding by pulling a lever on the projection apparatus, they point to the number they have chosen as their response. Feedback consists of the examiner's verbal response, "correct" or "incorrect" rather than the bell or buzzer which provides feedback on the CT apparatus. The BCT has the advantage that it is much less cumbersome than the CT and may be readily transported to patients' rooms and other testing sites. It is also less intimidating for many patients and facilitates examiner interaction to help maintain patient collaboration.

Validation of the BCT has been hampered by practice effects when both the Halstead CT and the BCT are administered within a short time period. To some extent, however, this has been compensated for by counter-balancing

Table 8

Distribution of Decade Age Ranges Within Patient Groups

GROUP	<20	20-29	30-39	40-49	50-59	60-69	>69
Experimental	7.58%	24.24%	29.55%	21.21%	9.85%	3.03%	4.55%
Comparison	15.91%	30.68%	22.73%	19.32%	10.23%	2.27%	0%

administration order in a variety of populations. These studies have consistently demonstrated a high degree of consistency between scores on the two forms (r=.913 and r=.893, for example) and equivalent patterns of correlation and group discrimination (DeFilippis & McCampbell, 1979). Although validation research continues, the BCT is generally viewed as an equivalent form of the CT.

Coding of Responses

Verbalized CT responses for the experimental group were coded by the experimenter, employing the decision tree procedures in Appendices E through G. These procedures were developed, also by the experimenter, using patient protocols which were not included in this study for the purpose of statistical analysis. These protocols were among those which were discarded due to missing data, generally because the WCST was not available or was not administered according to modified instructions.

Criteria employed in the development of the coding system were based upon the theoretical approach described by Perrine (1984), as outlined in the description of Coding Process Variables in the section above. Definitions and coding procedures were constructed in such a way that discrimination among verbal responses which varied in terms of Concrete Attribute Identification, Abstract Pattern Identification, or Rule Learning was enhanced. Such discrimination was measured on the basis of test-retest (recoding) reliability for recoding by the examiner of four protocols which had been initially coded, also by the examiner, four months previously[5]. Since verbal responses of subjects who made few errors on the CT provided little variety of scoring determinants or

[5] Although interscorer reliability would have provided a more accurate measure of reliability in coding responses, the complexity of the coding system precluded use of trained volunteers for the purpose of this study.

coding challenges, protocols which contained fewer than 50 errors were eliminated from the selection pool, and protocols used for assessing reliability were randomly selected from among those which remained. Although it is not typically used in this manner, an adaptation of the point-by-point agreement ratio was computed, based upon a total sample of 2,160 scores (Kazdin, 1982). Reliability for recoding responses utilizing this procedure was .943. Consequently, it was concluded that use of coding rules provided adequate reliability.

Data Entry

After all CT protocols were coded following the procedures described above, the experimenter entered all data into appropriate cells on a spreadsheet software program (Excel 4.0) running on an Apple Macintosh IIci computer. This spreadsheet contained formulas for computation of frequency and ratio scores for all variables defined in Appendix I. WCST responses, both those scored conventionally and those based upon patients' verbal responses, were entered into a second spreadsheet program, which contained formulas for computation of WCST variables listed in Appendix H. After all data were entered and all summary and ratio scores for both the CT and WCST were computed, these data were combined with WAIS-R IQ scores and demographic variables, with each subject identified by sequentially assigned ID numbers.

Data for the Comparison Group were summarized and collated in the same manner, except that CT cognitive process variables and WCST verbalized scores were not available for this group. Consequently, only CT error scores and conventionally scored WCST variables were combined. At this point, the combined scores for each group were exported to files for statistical analysis using SPSS for the Macintosh, version 4.0 (SPSS, 1990).

CHAPTER IV
RESULTS

Data analysis for this study proceeded in three distinct phases, reflecting the separate purposes of the three sets of hypotheses described in the previous chapter. First, error scores on the CT and conventional scores on the WCST, and relationships among those scores, for the Experimental Group were compared to the same scores for the Comparison Group in order to determine if there were differential performance characteristics between these groups, which might have suggested that asking patients to verbalize their reasoning altered the nature of the task. Since it had been established previously that these groups were not comparable with respect to some variables which have been found to influence performance on these tasks (age and some subtests on the Performance IQ scale), it was necessary to statistically control for the influence of these variables in performing these comparisons. This was accomplished by use of analysis of covariance (ANCOVA), in which scores reflecting performance comprised the dependent variables with group membership comprising the independent variable, while covarying age and WAIS-R performance. Since IQ scores were age-corrected, and age was already entered as a covariate, nonage-corrected IQ scores (summed scale scores) were utilized to avoid confounding correction terms. Comparing correlations between pairs of scores was accomplished in the same manner, partialing out the influence of age and IQ variables, then covarying one score on another score, with group membership as the independent variable.

The second set of analyses related to hypotheses three and four, which concerned the identification of separate cognitive process factors comprised of subsets of cognitive process variables from the CT and WCST. It was hypothesized that at least three interpretable factors for each test would emerge from factor analyses of the ratio and summary variables defined in Appendices H and I. It was further hypothesized that these factors would be at least moderately independent of age and IQ variables. Therefore, this set of analyses was accomplished using principal components factor analysis with oblique rotation since it was not expected that these factors would be independent of each other. The relationship of these factors to age and IQ variables was assessed through the use of multiple regression techniques, with age and WAIS-R subtest scores regressed on each interpretable factor which emerged from factor analysis of variables from each test.

The third set of analyses was intended to test hypotheses 5 and 6. It was hypothesized that when age and IQ covariates were partialed out, the correlation between CT error scores and WCST conventional scores would not be statistically significant, but that CT factor scores would be significantly correlated with WCST factor scores, when age and IQ covariates are partialed. This set of analyses was performed using multiple regression techniques, regressing the combined sets of covariates and independent variables onto each dependent variable, forcing covariates into the equation first, in order to partial out their influence.

Results of Group Comparisons

Hypothesis 1

It was anticipated that mean CT error scores and Categories Completed, Perseverative Errors, and Total Errors on the WCST for the Experimental Group would not be significantly different statistically from the same scores for the Comparison Group, when controlling for variance due to age, education, and

FSIQ (nonage corrected) covariates. This expectation was tested using ANCOVA, with summary scores on both the CT and WCST serving as dependent variables, and group membership as the independent variable, covarying blocked age, education, and summed WAIS-R scale scores (FSSUM). The results of these comparisons are listed in Table 9. Of the 11 comparisons, only mean scores for CT Subtest IV Errors (C4Err) were significantly different ($p = .017$).

The relevance of this single significant comparison for conclusions concerning the comparability of the scores for the two groups was explored further, covarying all possible combinations of age, education, and FSSUM variables. The two groups were found to be comparable only when age by itself was covaried ($p = .075$), but treating age and group membership both as independent variables, without covariates, produced the least differentiation between groups with respect to C4Err. Adding education and FSSUM as covariates decreased comparability of the groups.

All other comparisons listed in Table 9 were also recomputed using this procedure, and group differences were found only for CT Subtest VI (C6Err) and WCST Total Errors (WTE). Covarying education and FSSUM was found to

Table 9

Comparison of Mean CT and WCST Summary Scores

GROUP	CT						WCST				
	III	IV	V	VI	VII	TOT	CAT	PSV	NPS	TE	FTM
Experimental	16.0	*13.4	13.7	10.7	4.8	58.6	4.7	19.7	14.1	31.1	.98
Comparison	16.4	*10.7	11.9	8.3	4.3	51.9	4.6	20.3	19.2	36.9	.86

*$p < .02$

produce comparability between groups on C6Err (p = .16) and covarying education by itself produced comparability between groups on WTE (p = .24). No interaction terms were found for any of these analyses; consequently, although groups were found to be not comparable with respect to age and covariates for individual variables was sometimes dissimilar, as might be expected, no evidence was found for group differences as a function of administration condition.

Hypothesis 2

Not only were CT and WCST mean scores expected to be comparable between the experimental group and the comparison group, when controlling for age, education, and IQ covariates, but the relationships among those scores were also expected to be comparable. This hypothesis was tested utilizing analysis of covariance to evaluate the covariance between pairs of variables, after partialing out age, education, and FSSUM, with group membership comprising the independent variable. The main effect in each comparison for group membership is listed in Table 10.

Only 3 of 55 comparisons revealed statistically significant differences in covariance between variables as a function of group membership. Two of those involved WCST Total Errors, while the third difference was the correlation between CT Subtest III Errors (C4Err) and CT Total Errors (CTE).

Multiple regression analysis in each instance revealed that patterns of covariation for age, education, and IQ scores for these variables were dissimilar between groups. Consequently, it would not be expected that relationships among these scores would be similar across groups. As a result of these findings,

Table 10

Probabilities of Group Differences Between Variables

VARIABLES	III	IV	V	VI	VII	TOT	CAT	PSV	NPV	FTM	TE
	CT						WCST				
Subtest III	N/A	.14	.25	.28	.22	*.03	.32	.35	.31	.32	.58
Subtest IV		N/A	.18	.16	.06	.15	.09	.08	.11	.11	*.03
Subtest V			N/A	.31	.09	.18	.11	.11	.08	.12	.07
Subtest VI				N/A	.18	.35	.20	.21	.13	.24	.11
Subtest VII					N/A	.38	.83	.77	.91	.86	.45
Total Errors						N/A	.35	.33	.34	.41	.12
Cat. Completed							N/A	.80	.74	.65	**.003
PSV Responses								N/A	.06	.51	.11
NonPSV Resp.									N/A	.13	.40
Fail to Maintain										N/A	.28
Total Errors											N/A

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

covariate relationships for all variables were computed for both groups. These results are shown in Table 11.

It can be seen that age, education, and IQ covariates for all scores are quite dissimilar between the groups. Consequently, although most correlations between pairs of CT and WCST variables were comparable between groups, when age, education, and Full Scale IQ were partialed out, the three correlations which were not comparable were sufficient to raise questions concerning the comparability of groups. However, given the group differences which were previously identified, it was unclear whether dissimilar relationships among variables across groups reflected differences due to administration procedures or due to other variables.

This question was explored further by analyzing relationships among other sets of scores on which administration procedures did not vary between groups. It was reasoned that if groups were comparable with respect to diagnosis, severity of impairment, and other relevant variables, relationships among WAIS-R subtest and summary scores would be similar across groups, since these

Table 11

R^2 for Age, Education, and IQ Regressed on CT/WCST Variables

Covar.	Age	Ed	I	DS	A	V	C	S	PC	PA	BD	OA	DSy	VS	PS	FS	Tot.
EXPERIMENTAL GROUP																	
C3Err	.06								.11				.03				**.18**
C4Err		.03					.04			.17						.05	**.27**
C5Err											.04				.21		**.24**
C6Err																.18	**.18**
CTE	.02					.08										.27	**.37**
WCAT	.07	.03														.20	**.28**
WPSV								.02					.05		.26		**.31**
WNPV	.03										.06					.10	**.16**
WFTM																	**0**
WTE	.07												.02			.32	**.40**
COMPARISON GROUP																	
Covar.	Age	Ed	I	DS	A	V	C	S	PC	PA	BD	OA	DSy	VS	PS	FS	Tot.
C3Err	.05														.40		**.44**
C4Err	.05							.08			.32						**.41**
C5Err											.23						**.22**
C6Err											.30						**.29**
CTE	.08										.50					.03	**.59**
WCAT									.17								**.15**
WPSV	.07								.09								**.13**
WNPV																	**0**
WFTM											.07						**.05**
WTE									.14								**.13**

tasks were administered in the same way for all subjects. The results of these multiple regression analyses are shown in Table 12.

These analyses revealed that relationships among WAIS-R subtest and summary scores were quite dissimilar, with the exception of Wechsler's (1944) "Hold" scores on the verbal scale, Information, Vocabulary, and Comprehension subtests, plus Similarities. However, the so-called Hold scores from the performance scale, Picture Completion and Object Assembly, were among those which were most dissimilar across groups.

Table 12

R^2 Change for Age, Education, and IQ Interrelationships

Covar.	Age	Ed	I	DS	A	V	C	S	PC	PA	BD	OA	DSy	VS	PS	FS	Tot.
EXPERIMENTAL GROUP																	
Info.		.03				.52								.02	.02	.02	**.59**
DS					.25	.04					.03	.02	.08	.08	.02		**.50**
Arith.				.02		.36					.09			.08	.03	.02	**.58**
Voc	.02		.52	.04	.01		.12	.02						.04	.02	.05	**.82**
Comp.	.02					.42			.05	.03				.03			**.53**
Sim.	.09					.39	.04						.02	.10			**.62**
PC							.09				.24			.07	.02	.09	**.50**
PA	.04						.02	.07				.28			.13		**.53**
BD				.07					.02			.48			.01		**.57**
OA				.02						.06	.48		.02	.12	.06	.02	**.77**
DSY				.12				.02				.23			.05		**.40**
COMPARISON GROUP																	
Covar.	Age	Ed	I	DS	A	V	C	S	PC	PA	BD	OA	DSy	VS	PS	FS	Tot.
Info.						.55		.04						.06			**.63**
DS					.25												**.24**
Arith.				.11				.44						.13			**.66**
Voc	.03		.55				.03	.08						.08			**.76**
Comp.						.43			.13				.04	.08			**.67**
Sim.					.09	.50								.18			**.75**
PC							.33			.09					.12		**.52**
PA									.06		.39				.25		**.68**
BD				.04						.39			.08		.22		**.71**
OA										.20			.05		.22		**.45**
DSY		.22									.09				.18		**.46**

These findings strongly suggested that the comparison group was not comparable to the experimental group with respect to variables influencing CT and WCST performance. The pattern of dissimilarity among WAIS-R scores raised the possibility that groups differed with respect to severity, but sufficient data to evaluate this possibility were not available. More importantly, these differences precluded determination of the effect of administration procedure modifications on CT and WCST performance.

Factor Analyses of Cognitive Process Variables

Hypothesis 3

A primary goal of this study was to determine if scores derived from a theoretically based approach to concept formation could differentiate among the problem-solving approaches displayed by individual patients on the CT. It was expected that accomplishment of this goal would be reflected in the emergence of at least three distinct and interpretable factors in factor analysis of the cognitive process variables defined in Appendix H. It was further anticipated that these factors would be relatively independent of age, education, and IQ variables, unlike summary error scores.

Mean cognitive process scores were computed for each subtest individually, as well as across all subtests used in the analysis. However, Leonberger, Nicks, Larrabee, and Goldfader (1992) noted that multicolinearity results when summary measures are entered in the same analysis with subtests used to compute them; consequently, mean total variable scores were factor analyzed separately from subtest variable scores. An additional problem cited by Leonberger et al. is the use of multiple variables assessed by the same or similar methods, which tends to inflate correlations among them. An attempt was made to minimize this problem by converting variable scores to proportions, to reduce the dependence of variables within subtests upon error scores on that subtest.

Since 19 process variables were computed for each subtest and factor analysis of all variables simultaneously would require considerably more subjects than were presently available for this study (Gorsuch, 1974), each subtest was analyzed separately, and separate analysis was conducted for mean scores across all subtests. Additionally, since normative data were not available for these scores, and cannot be derived from the results of this neuropsychologically impaired group (see mean scores in Table 9), in order to facilitate the interpretability of factor loadings on each factor analysis, the group was divided into impaired and unimpaired subgroups, using Reitan's criteria for impaired performance on CT total errors (>50), and the factor analysis for each subtest was replicated for each subgroup. Table 13 lists the mean scores and standard deviations of each subgroup on each CT subtest and for total errors.

It should be noted that these secondary analyses violated the recommended minimum ratio for subjects to variables for factor analysis (5:1). Consequently, these results were utilized only for the purpose of clarifying how factor loadings differed for impaired versus unimpaired subjects, thereby clarifying relationships among variables which comprised the factors.

Each retained factor from the primary individual analyses was transformed to factor scores for each subject, and factor scores were then subjected to second order factor analysis in order to assess the relationships among cognitive processes across subtests. Subsequently, the relationship of each primary and higher order factor to age, education, and IQ covariates was assessed utilizing multiple regression analysis.

Table 13

Mean CT Scores for Unimpaired Versus Impaired Subgroups

GROUP	N	III		IV		V		VI		TOTAL	
		Mean (S.D.)									
Unimpaired	58	7.6	(8.5)	4.9	(5.3)	10.0	(4.9)	6.8	(5.0)	31.5	(12.7)
Impaired	74	22.6	(11.1)	20.1	(10.5)	16.6	(6.4)	13.7	(7.0)	79.9	(21.0)

Subtest III

Factor pattern. The factor pattern for the results of factor analysis of Subtest III variables is shown in Table 14 . Six factors with eigenvalues above 1.0 were extracted, accounting for 79.9% of the variance, but the scree test and inspection of the factor loadings indicated that only the first four factors, accounting for 60.2% of the variance, should be retained.

Factor 1 in this analysis relates to maintaining set when the previous response was correct. Four of the seven variables which loaded highly on this factor were based on maintaining the determinants in the previous response, although,

Table 14

Factor Pattern Matrix for Full Group on Subtest III

Variable Number:	1	2	3	4	5	6
Percent of Variance:	28.1%	15.4%	9.8%	6.9%	6.2%	5.7%
6. RKA	.78					
4. SLR	.74					
8. RPCK	.72			-.39		
16. NC3	-.69				.35	
3. LR	-.65		.40			
7. RKR	.53					.34
12. WS1		-.88				
10. WSA		.83				
14. WS3			-.79			
15. WPCS		.53	-.71			
13. WS2		.57	.66			
11. WSR	.42		-.64			
9. WSC			-.62			
5. RKC				-.95		
1. MSIS				-.93		
19. PSV5-6					.71	
18. PRAC4					.69	
2. MSBS						.73
17. NC4						.70

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

interestingly, maintaining the concrete determinant (C) from the previous response loaded on a separate factor and did not load on Factor 1. Not surprisingly, Learning Rate (represented by the item number of the third consecutive conceptually correct response) also loaded on this factor, as did Near Correct, which is comprised of responses in which the correct principle is applied incorrectly. Thus, this factor appeared to represent efficient learning to the correct principle and maintaining that principle on subsequent items.

Factor 2, accounting for 15.4% of the variance, reflected an inefficient approach to problem-solving characterized by shifting more than one determinant simultaneously when the previous response was incorrect, which usually included the abstract pattern (A) as one of the determinants shifted, with no clear pattern emerging for selecting the other determinant to be shifted. This factor appeared to represent variance attributed to subjects who repeatedly tried different combinations of determinants without any systematic plan or incorporation of feedback.

Factor 3, which accounted for 9.8% of the variance, was also represented by variables which reflected somewhat inefficient and unsystematic learning. However, this factor was differentiated from the previous factor by a moderate positive loading for Learning Rate, indicating that the correct principle was learned but following numerous learning trials.

Factor 4, which accounted for 6.9% of the variance, reflected variance attributable to subjects who perseverated on concrete features of the stimuli, but who failed to apprehend that identifying the pattern which was most different determined which concrete feature was relevant. Thus, RKC loaded highly on this factor (in a negative direction), as did, by definition, MSIS.

Factor 5 consisted primarily of two variables derived from other subtests, PRAC4 and PSV5-6, which both related to some degree of cognitive rigidity or shifting set from one subtest to another, although scores on these variables imply learning of the Subtest IV principle. This variable accounted for 6.2% of the variance.

Factor 6 was ambiguous, but accounted for 5.7% of the variance. MSBS (maintaining set between sets when the stimuli change within the subtest) loaded highly on this factor, but it showed no relationship with MSIS (maintaining set within sets). NC4, which related to a rigid adherence to an almost correct principle that resulted in many errors, also loaded highly on this factor. Since variance attributable to adaptive set maintenance behavior was accounted for by Factor 1, it appeared that this factor represented an automatic form of repetition which is nonadaptive.

Comparison with unimpaired subgroup. Table 15 shows the factor pattern for the unimpaired subgroup. Results for these groups were quite similar, but also differ in significant ways. Most importantly, whereas for the full group, responses involving the Abstract Pattern were highly salient and loaded highly on the first two factors, responses involving Concrete Attributes were much less discriminating.

In contrast, within the group of subjects whose performance on the CT was unimpaired, the variable which loaded first, and thus accounted for the greatest proportion of variance, was RKC, and, as previously, loaded with MSIS. MSBS also became a much more discriminative variable than was previously true. The only other notable change in the factor patterns between these groups was the dropping out of RKA and RKR as discriminative variables. Thus, the primary difference between the full group and the unimpaired subgroup was the

Table 15

Factor Pattern Matrix for Unimpaired Subgroup on Subtest III

Variable Number:	1	2	3	4	5	6
Percent of Variance:	32.8%	14.8%	10.3%	7.3%	6.2%	5.5%
5 RKC	.99					
1 MSIS	.86					
4 SLR	.81					
2 MSBS	.74					
8 RPCK	.60			.45		
3 LR	-.59		.38			
16. NC3	-.38					
10. WSA		.94				
12. WS1		-.87				
15. WPCS		.70	-.51			
13. WS2		.32	.85			
14. WS3			-.74			
9 WSC			-.56		-.54	
11. WSR	.36	.39	-.52			
17. NC4				.86		
7 RKR				.53		-.37
6 RKA	.41			.48		-.33
18. PRAC4					.91	
19. PSV5-6						-.90

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

change of position of RKC and MSIS with RKA and RKR, with increased discriminative power for MSBS.

Comparison with impaired subgroup. The factor pattern for the impaired subgroup is shown in Table 16. This pattern was dissimilar from the other groups in many ways. Whereas in the other groups Maintaining Set accounted for the largest proportion of the variance, in this subgroup it accounted for only 68% of the variance. Interestingly, NC3 did not load on this factor, as for the other groups, but PSV5-6 replaced it, although it loaded in the opposite direction. Although the correlation between these variables was null (-.01), probably

Table 16

Factor Pattern Matrix for Impaired Subgroup on Subtest III

Variable Number:	1	2	3	4	5	6	7
Percent of Variance:	23.5%	19.4%	11.6%	7.3%	6.8%	5.9%	5.5%
13. WS2	-.84	-.39					
12. WS1	.82						
10. WSA	-.78						
14. WS3		.94					
11. WSR		.88					
15. WPCS	-.45	.77					
9. WSC		.49		-.47			.42
3. LR		-.48			-.42		
5. RKC			.94				
1. MSIS			.90				
17. NC4				.91			
6. RKA					.80		
8. RPCK			.40		.77		
19. PSV5-6					.59	.40	
7. RKR		-.37			.59		
4. SLR		.35		.36	.49		
16. NC3						.83	
2. MSBS					.42	-.51	.44
18. PRAC4							.78

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

because they were each unique to unrelated subtests, both appeared to relate to persisting with unsuccessful hypotheses. Thus, for the other groups, NC3 loaded negatively on a factor associated with maintaining set when the previous response was correct, but for this impaired subgroup, PSV5-6 loaded positively on that same factor. This loading may well reflect the reason this factor accounted for little variance within this subgroup. Whereas for the other groups maintaining set appeared to reflect a purposeful and adaptive problem-solving strategy, for this subgroup it may represent cognitive rigidity and failure to adopt a purposeful strategy.

Factor 1, which accounted for 23.5% of the variance, was a mirror image of Factor 2 in the full group, with the addition of WS2 loading negatively with WSA. Thus, not surprisingly, for this impaired subgroup, the largest proportion of the variance was accounted for by systematic problem-solving strategies when the previous response was incorrect.

Factor 2 for the impaired subgroup, accounting for 19.4% of the variance, was similar to factor composition of Factor 3 in the other groups which related to less efficient problem-solving strategies when the previous response was incorrect, but for this subgroup many of the variables, including LR, loaded in the opposite direction. Furthermore, for this subgroup SLR loaded somewhat moderately on this factor. Thus, this factor represented usually inefficient problem-solving strategies which in this instance were associated with some degree of success. Therefore, the variance accounted for by this factor appeared to be due to flashes of insight in which the correct rule was suddenly apprehended.

Factor 3 for this subgroup, accounting for 11.6% of the variance, was identical in variable composition to Factor 4 for the full group, consisting primarily of RKC and MSIS, but variables loaded in the opposite direction for this subgroup. These are the variables which were highly discriminating in the unimpaired subgroup, but for this subgroup, as for the full group, accounted for a much smaller proportion of the variance. For the full group, this factor represented subjects who perseverated on identifying a common physical feature on succeeding items rather than identifying the element within each item which was most different. For this subgroup, however, this factor represented subjects who were able to successfully identify the relevant physical feature, but apparently this was not translated to the correct rule.

Factor 4 for this subgroup, accounting for 7.3% of the variance, has no correlate in the other groups. It is comprised predominantly of NC4, WSC loaded moderately in a negative direction, and it was somewhat predictive of maintaining set when the principle was learned. Thus, it appeared that this factor represented variance attributed to subjects whose concrete responses were often relevant, but the rule was probably based on counting. Similarly, these subjects achieved only incomplete rule learning in Subtest IV.

Factor 5 was similar to Factor 1 in the other groups, and represented variance attributed to maintaining set. However, for this subgroup, this factor showed a reduced loading for LR and SLR, reflecting the increased learning trials necessary to establish set. As a result, it accounted for only 6.8% of the variance, and was associated with increased perseveration on Subtests V and VI, suggesting that these impaired subjects established set more slowly and were less flexible in shifting to alternative principles when the previously learned principle became ineffective.

Factor 6 reflected partial learning in Subtest III, probably due to perseveration of the counting principle from Subtest II, and was also associated with perseveration of the Subtest IV principle on Subtests V and VI. This factor also had no correlate in the other groups.

The final factor for this subgroup was comprised predominantly of PRAC4 variance, which was attributed to slow learning of the demonstration items on Subtest IV, but with success achieved within six trials. Consistent with that pattern, this factor was associated with some degree of perseveration on concrete physical features of stimuli in Subtest III.

Subtest IV

Factor pattern. The factor pattern for Subtest IV, shown in Table 17, was somewhat different from that of Subtest III. Factors were much more dichotomized between those comprised of variables related to maintaining set and those related to shifting determinants. Factor 1 again was related to maintaining set, although for this subtest, all set maintenance variables, including RKC, MSIS, and MSBS, as well as PRAC4, loaded on this factor. In that sense, this factor represented a more pure measure of set maintenance.

Table 17

Factor Pattern Matrix for Full Group on Subtest IV

Variable Number:	1	2	3	4	5
Percent of Variance:	38.1%	19.0%	8.5%	6.4%	5.4%
8. RPCK	.93				
6. RKA	.90				
5. RKC	.89				
1. MSIS	.89				
3. LR	-.78				
4. SLR	.75				
2. MSBS	.71				
7. RKR	.68		-.58		
18. PRAC4	.63				
15. WPCS		.95			
10. WSA		.88			
9. WSC		.82			
12. WS1		-.75		-.34	
14. WS3		.73			-.31
17. NC4			.88		
11. WSR	.41	.36	.42	-.39	
13. WS2				.80	
19. PSV5-6				-.54	
16. NC3					.87

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

Factor 2 was comprised of variables related to ineffective problem-solving when the previous response was incorrect. Shifting A and C determinants (WSA, WSC) were prominent on this factor, as was the total percentage of determinants shifted (WPCS). Not surprisingly, shifting all three determinants simultaneously (WS3) also loaded highly. In contrast, shifting just one determinant following an incorrect response (WS1) loaded negatively on this factor, reinforcing the interpretation of this variable from Subtest III. Rule shifting showed low loading on this factor, as was expected on this subtest, where the primary task was to discover the clockwise rotational pattern which comprises the rule. Interestingly, variance on this score (WSR) was distributed to an approximately equal degree across the first four factors, presumably reflecting the variety of hypotheses utilized by subgroups of patients in seeking to discover the correct rule.

Factor 3 was comprised almost exclusively of variance associated with incomplete rule learning. NC4, which reflected utilization of a left/right location scheme for this subtest, or an alternative nonclockwise rotation, loaded highly on this factor. Subjects with high NC4 scores usually were expected to respond correctly, at least initially when the missing quadrant was located in the first two positions (the first line in the stimulus moving from left to right), but incorrectly when the missing quadrant was located on the second line (where clockwise rotation reverses the expected order of quadrants three and four). Consequently, rule learning necessarily involves shifting following correct feedback to quadrants one and two, if a left/right location scheme was utilized. Consistent with this expectation, RKR (repeating the previously used rule when the previous response was correct) loaded to a moderate level in a negative direction on this factor. Consistent with that finding, WSR also loaded moderately.

Factor 4 is predominantly comprised of WS2 (shifting two determinants simultaneously following an incorrect response), with a low negative loading for WS1 and WSR, and a more moderate negative loading for PSV5-6. This pattern suggested that subjects who simultaneously shifted two determinants were somewhat unlikely to utilize a more systematic approach of shifting only one determinant (although this negative correlation is not as pronounced as one might expect) and were also somewhat unlikely to shift the rule utilized. Thus, the variance associated with this factor appeared to be generated by subjects who failed to learn the location principle for this subtest and probably continued to utilize the counting principle articulated by Simmel and Counts (1957): "If in doubt, count." Consistent with this hypothesis, PSV5-6 loaded negatively on this factor at a moderate level, since these subjects were unlikely to have learned the location principle to perseverate to on subsequent subtests.

Factor 5 consisted predominantly of NC3 with a low negative loading for WS3. Thus, most of the variance associated with this factor arose from the previous subtest.

Comparison with impaired subgroup. The factor pattern for the impaired subgroup is shown in Table 18. Results for this subgroup were almost identical to that for the full group. The only notable changes were in the direction predicted in the analysis above.

For this subgroup, RKR loaded to a slightly less degree on Factor 1, Maintaining Set, and to a somewhat greater degree on Factor 3, Nonclockwise Rotation. The relationships among variables for Factor 4 were also slightly stronger, with the addition of low negative loading for MSBS, LR, and SLR, as might be expected given the learning failure associated with this factor. Factor 5

Table 18

Factor Pattern Matrix for Impaired Subgroup on Subtest IV

Variable Number:	1	2	3	4	5
Percent of Variance:	33.7%	20.9%	10.4%	5.8%	5.6%
RPCK	.88				
MSIS	.86		-.31		
RKA	.86				
RKC	.85				
PRAC4	.69				
LR	-.57			.34	
SLR	.56			-.31	
MSBS	.46			-.36	-.32
WPCS		.93			
WSC		.83			
WS3		.77			.36
WSA		.77			
WS1		-.61		-.35	
WSR		.55	.48	-.51	
NC4			.90		
RKR	.50		-.73		
PSV5-6				-.82	
WS2				.65	
NC3					-.89

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

was also virtually identical, although SLR now loaded to a slightly greater degree.

Comparison with unimpaired subgroup. When the factor pattern for these groups was compared with that for the unimpaired subgroup (see Table 19), somewhat more differences were apparent but the changes were relatively minor, and in the anticipated direction.

PRAC4 for this subgroup (learning the principle on the first six practice items) was not associated with maintaining set, probably because of reduced variability on this variable among these subjects. Instead PRAC4 now comprised the fifth

Table 19

Factor Pattern Matrix for Unimpaired Subgroup on Subtest IV

Variable Number:	1	2	3	4	5
Percent of Variance:	38.1%	19.0%	8.5%	6.4%	5.4%
RKR	.98				
RPCK	.94				
MSBS	.92				
LR	-.90				
MSIS	.86				
SLR	.85				
NC4	-.83				
RKC	.82				
RKA	.74				
WPCS		.94		.34	
WSA		.92			
WS1		-.84			
WSC		.84			
WS3		.69			
WS2			.81		
PSV5-6			-.62		.50
NC3			.53		
WSR				.91	
PRAC4					.85

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

factor, on which the only other variable loading above .3 was PSV5.6. Thus, among subjects who performed relatively well on the CT overall, but showed inefficient learning of the Subtest IV principle on the demonstration items, inefficient learning of the principle on subsequent subtests was also likely.

Additionally, Factor 3 for this subgroup was similar to Factor 4 for the other groups, with the addition of NC3 and a weaker negative association with adaptive problem-solving strategies (WSR and WS1). For this subgroup, unlike the other groups, WSR was not associated with nonclockwise rotation but showed a weak relationship with nonadaptive problem-solving (WPCS).

Subtest V

Factor pattern. The factor pattern for Subtest V is shown in Table 20. Again, this pattern was similar in many respects to that of Subtests III and IV, but also uniquely reflected the task differences presented by this subtest. Factor 1 was again comprised of set maintenance scores, but RKR (maintaining rule) loaded only weakly on this factor, probably reflecting the reduced variability of this score among subjects who consistently maintained set on other scores when the previous response was correct. That hypothesis was supported by the weak loading for WSR, reflecting the low probability of rule shifting when the previous response was incorrect, among subjects who consistently maintained set.

Factor 2 was comprised predominantly of ineffective problem-solving strategies, with WS3 and WPCS loading highly, and WSR loading more moderately, supporting the interpretation above for this variable. WS1 loaded negatively on this factor, as expected, with low negative loading also for RKR, probably reflecting the randomness of problem-solving strategies suggested by high scores on this factor.

Factor 3 represented more systematic problem solving when the previous response was incorrect, with WSA loading highly, and a low negative loading for RKA and WSR. This factor appeared to reflect variance attributable to subjects who correctly discerned that the task on this subtest was accurate pattern identification, but failed to perceive the relationship between the number choices presented them and subdivision of the stimulus. The adaptive role of this factor was further supported by the moderate loading for RKR and the high negative loading for PSV5-6, reflecting the low probability of rule perseveration from the previous subtest of subjects who effectively problem solved.

Table 20

Factor Pattern Matrix for Full Group on Subtest V

Variable Number: / Percent of Variance:	1 36.2%	2 16.4%	3 11.3%	4 6.8%	5 5.8%	6 5.4%
1. MSIS	.97					
8. RPCK	.96					
5. RKC	.95					
6. RKA	.91		-.31			
3. LR	-.90					
4. SLR	.89					
2. MSBS	.71					
14. WS3		.96		.35		
15. WPCS		.90		-.30		
11. WSR	.33	.68	-.42			
12. WS1		-.63		.41		
10. WSA		.32	.83			
19. PSV5-6			-.75			.31
7. RKR	.38	-.39	.58			
13. WS2				-.98		
9. WSC		.31		-.47	.30	-.32
16. NC3					.77	
17. NC4					.68	.31
18. PRAC4						.84

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

Factor 4 was difficult to interpret. High scores on this factor were associated with a very low probability of shifting two scores simultaneously when the previous response was incorrect. High scorers on this factor were also somewhat unlikely to shift C, but did often shift three scores simultaneously. These subjects tended to have a moderately low proportion of total scores shifted or may have shifted only one score. This pattern appeared to be characteristic of patients who were not responding randomly, but were not systematic or well organized in seeking to discover the correct principle. This factor might be described as "confused problem-solving."

Factor 5 was predominantly comprised of NC3 and NC4, with a low loading by WSC. Thus, high scores on this factor appeared to relate to incomplete learning, which may result from rigid thinking, since NC3 and NC4 describe the correct principle used incorrectly. Subjects with high scores on NC3 had difficulty giving up the counting principle from practice Subtest II and thus counted the number of elements located in specific positions, and subjects with high scores on NC4 were usually unable to give up a conventional left/right orientation in positioning elements in the stimulus, even when the six practice items demonstrated an alternate placement. Not surprisingly, this factor had a low loading on WSC, which probably represented some perseveration from the missing element in Subtest IV to the dotted lines in Subtest V. The failure to PSV5-6 to load on this factor was probably due to poor learning of the Subtest IV principle, which was a prerequisite for perseveration to this subtest.

Factor 6 also appeared related to rigid thinking, but whereas Factor 5 was associated with failure to learn, this factor was associated with slow learning. PRAC4 (slow learning of the correct principle on the practice items) loaded highly on this factor, but NC4 showed a weak loading, as does PSV5-6. For each of those variables, the factor loading reflected a relatively high number of trials required to learn the principle, but the principle was learned in each case, unlike the variance associated with Factor 5. The weak negative loading of WSC, which was almost a reverse, mirror image of the loading on Factor 5, supported that interpretation.

Comparison with impaired subgroup. The factor pattern for the impaired subgroup is shown in Table 21. These results replicated those for the full group with respect to the first two factors, except that the low loading for WSA on Factor 2 in the full group was replaced by low loadings for WSC and PSV5-6 in

Table 21

Factor Pattern Matrix for Impaired Subgroup on Subtest V

Variable Number:	1	2	3	4	5
Percent of Variance:	32.8%	20.1%	10.8%	6.7%	6.4%
MSIS	.95				
RPCK	.94				
RKC	.92				
SLR	.87				
LR	-.81				
RKA	.79				
MSBS	.71				
WPCS		.96			
WS3		.94			
WSR		.87			
WS1		-.74			
WSA			.80		
NC3			-.64		
WSC		.41	-.47	.38	-.36
WS2				.75	
NC4				.70	.37
PRAC4					.66
PSV5-6		.36	-.30		.61
RKR		-.50			-.52

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

this subgroup. Additionally, the proportion of the variance accounted for by this factor was slightly greater. These changes all reflected the somewhat greater likelihood of ineffective problem-solving in this subgroup.

The remaining three factors for this subgroup were somewhat different from the pattern in the full group, but still recognizable. Factor 3 was comprised predominantly of high scores on WSA and a moderate likelihood of a low score on WSC, with a weak negative loading for PSV5-6. Thus, this appeared to represent variance attributed to subjects who recognized the correct concrete attribute but had difficulty learning the quadrant principle for the abstract pattern. The negative loadings on NC3 and PSV5-6 suggested that cognitive rigidity did not characterize the performance of these subjects, however. This factor was somewhat similar to Factor 3 for the full group, but the negative correlation between WSA and PSV5-6 was much weaker for this subgroup and the moderate loading for RKR was much weaker, and moderate negative loadings for WSC and NC3 were present. Thus, this factor appeared to represent less efficient problem solving for this subgroup than was true for the full group, but rigid thinking was negatively associated with this factor for this subgroup.

Factor 4 for this subgroup was very different from that of the full group. Whereas in the latter case this factor characterized subjects with intentional but nonsystematic and confused problem-solving strategies who were unlikely to shift two scores simultaneously when the previous response was incorrect, for this subgroup, high scorers on this factor were likely to shift two scores simultaneously. NC4 also loaded highly on this factor, which may characterize subjects who became frustrated by this point and were responding randomly.

Factor 5 for the full group had no correlate in the impaired subgroup, but Factor 5 in this subgroup corresponded to Factor 6 in the full group, with the addition of a moderate loading in the negative direction for RKR. Thus, in this subgroup, this factor continued to suggest slow learning, but with less success than was true for the full group.

Comparison with unimpaired subgroup. The factor pattern for the unimpaired subgroup (see Table 22) was similar to that for the other groups, but composition of factors was slightly different. Factor 1 replicated the first factor in the other groups in most respects, but with the addition of low to moderate negative loadings for PSV5-6 and WSC, reflecting the decreased incidence of rigid problem solving in this subgroup. Additionally, WSR was not associated

Table 22

Factor Pattern Matrix for Unimpaired Subgroup on Subtest V

Variable Number:	1	2	3	4	5
Percent of Variance:	39.8%	13.9%	10.8%	7.0%	5.7%
RKC	.98				
MSIS	.96				
RPCK	.95				
RKA	.92				
LR	-.87				
SLR	.86				
MSBS	.58				
WS2		.91		-.48	
WS1		-.86			
WPCS		.83		.40	
WSC	-.34	.61			
WSR			.79	.35	
PSV5-6	-.42		.74		
RKR	.54		-.64		
WSA			-.50		.36
NC4		.48	-.39		
WS3				.88	
PRAC4					.74
NC3					.64

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

with this factor, probably also as a function of decreased perseveration of the rule from Subtest IV for this subgroup.

Factor 2 continued to reflect ineffective problem solving, but it contributed a slightly smaller proportion of the variance. Whereas shifting three determinants simultaneously (WS3) loaded highly for the other groups on this factor, for this subgroup shifting two determinants simultaneously (WS2) loaded highly, and there was no association for WS3, probably due to limited variability on this score. Both WSA and WSC showed a stronger association with this factor for this subgroup, but WSR, which loaded moderately for the other groups, showed no association, and WS1 was more negatively associated. Thus, for the unimpaired subgroup, ineffective problem-solving, though still accounting for substantial variance, was less maladaptive than for the other group.

For this subgroup, Factor 3 appeared to be almost a mirror image of this factor in the other groups, particularly the full group. Whereas previously this factor represented effective problem-solving when the previous response was incorrect, for this subgroup it reflected very ineffective problem-solving. Variable loadings for all groups were very similar, but the direction of the loadings for this subgroup was opposite. Consistent with that change, the moderate loading for RKR was replaced in this subgroup with a low negative loading for NC4, reflecting rigid thinking.

Factor 4 for this subgroup was also in the opposite direction than for the full group, but like the impaired subgroup, this factor appeared to represent random or nonsystematic responding, with a high loading for WS3.

Factor 5 for this subgroup had no precise correlate in the other groups in terms of variable loadings, but like the other groups this appeared to represent slow learning due to rigid thinking, with high loadings for PRAC4 and NC3, and a low loading for WSA.

Subtest VI

Factor pattern. The factor pattern for Subtest VI, shown in Table 23, was very similar to that for Subtest V, as might be expected for the only two subtests sharing the same principle. The only notable difference for Factor 1, reflecting set maintenance, was that WSR showed a weak association on Subtest V but not Subtest VI, and RKR showed a somewhat higher loading on Subtest VI. Presumably, that difference was related to the rule shift between Subtests IV and V which was not present between Subtests V and VI.

Table 23

Factor Pattern Matrix Subtest VI

Variable Number:	1	2	3	4	5
Percent of Variance:	36.2%	18.0%	10.3%	6.9%	5.8%
4. SLR	.97				
8. RPCK	.97				
5. RKC	.95				
6. RKA	.93				
1. MSIS	.92				
2. MSBS	.91				
3. LR	-.86				
15. WPCS		.95			
12. WS1		-.87			
14. WS3		.63		-.46	
9. WSC		.62			
10. WSA		.60			
11. WSR		.57	-.49	-.32	
19. PSV5-6			-.85		
7. RKR	.53		.66		
18. PRAC4			-.54	.49	
13. WS2				.87	
16. NC3					.80
17. NC4					.64

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

Factor 2 also replicated Factor 2 from Subtest V, but represented even more pronounced ineffective problem solving, with WSC, WSA, and WSR all loading to an approximately equal degree, suggesting a random approach to discovering the correct principle. Unlike Subtest V, RKR did not load on this factor in a negative direction, probably due to reduced variability on this score.

The variable loading pattern on Factor 3 was also very similar to the variable loading pattern on the third factor in Subtest V, with the exception that in the latter case, WSA was the predominant feature associated with adaptive problem-solving when the previous response was incorrect, but on this subtest WSA did not load on this factor. Rather, maintaining the previous rule is the predominant feature here, and was associated with negative loadings on PSV5-6 and PRAC4. Thus, adaptive problem-solving associated with these scores apparently acted to restrict variability on abstract pattern identification after learning the principle in Subtest V and maintaining set with respect to formulating the rule based on the quadrant principle was the differentiating variable.

Factor 4 reflected ineffective problem-solving, as in Subtest V, although the loading pattern was somewhat different. For this subtest, this factor was associated with shifting two determinants simultaneously, although WSR loaded weakly in the negative direction indicating that this shifting was not entirely random. PRAC4 loaded moderately on this factor, reflecting slow learning or a moderate degree of cognitive rigidity.

Factor 5 replicated Factor 5 in the previous subtest and was represented almost entirely by high loadings on NC3 and NC4. Thus, this factor reflected cognitive inflexibility associated with tasks on previous subtests but was not associated with any variables on this subtest to any substantial degree.

Comparison with impaired subgroup. The factor pattern for the impaired subgroup is shown in Table 24. Factor 1 essentially replicated factor composition of Factor 1 for the full group, with the addition of weak loadings for shifting just one determinant at a time (WS1) but not two (WS2). Thus, for this impaired subgroup, maintaining set was associated with utilization of a systematic approach, as might be anticipated.

Factor 2 also was very similar to Factor 2 in the full group, reflecting nonsystematic responses when the previous response was incorrect. Factor composition was almost identical in the two groups, but small differences emerged with respect to factor loadings. WSR loaded to a substantially greater

Table 24

Factor Pattern Matrix for Impaired Subgroup on Subtest VI

Variable Number:	1	2	3	4	5
Percent of Variance:	37.0%	17.1%	10.9%	6.9%	6.3%
SLR	.97				
RKA	.95		-.31		
RPCK	.95				
MSIS	.90				
RKC	.90				
MSBS	.89				
LR	-.79				
WPCS		.98			
WSR		.85	-.35		
WS3		.68			
WS1	.32	-.66		-.39	
PSV5-6			-.86		
RKR	.44		.69		
WSA		.53	.55		-.37
WS2	-.33			.76	
PRAC4			-.42	.57	
NC4				.43	
WSC		.41			.85
NC3		.34			-.53

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

degree in the impaired subgroup than was true for the full group, and WSC loaded to a lesser degree. Further, NC3, reflecting a rigid approach to task performance in Subtest III, loaded weakly on this factor for the impaired subgroup. However, these differences between groups with respect to this factor, which reflected random and nonsystematic problem solving, were generally minor.

Factor 3 for this subgroup is also very similar to the third factor for the full group, except that WSA now loaded to a moderate degree on this factor, as it did for Factor 3 in Subtest V. Thus, for this subgroup, when the previous response was incorrect, this factor reflected subjects who tended to appropriately shift the abstract pattern, rather than the rule.

Factor 4 in this subgroup was somewhat different than Factor 4 in the full group, although still reflecting cognitive rigidity with moderate loadings on PRAC4 and NC4 and a moderate loading in the negative direction on WS1.

Factor 5 was very different than the fifth factor in the full group, which reflected only variance attributable to cognitive rigidity on previous subtests. For this subgroup, this factor loaded principally on WSC, with negative loading on NC3 and WSA. Despite these differences, this pattern still reflected rigid thinking in attempting to shift the concrete features rather than the abstract pattern.

Comparison with unimpaired subgroup. The factor pattern for the unimpaired subgroup is shown in Table 25. This pattern was almost identical to the other groups with respect to Factors 1 and 2, reflecting, respectively, maintaining set and responding nonsystematically. Factor 3 for this subgroup was very similar to Factor 4 for the full group, reflecting a nonsystematic approach when the previous response was wrong, and was also associated with slow or somewhat rigid thinking.

Factor 4 loaded on most of the same variables as Factor 3 in the full group, but loadings were in the opposite direction. Consequently, whereas for the full group this factor reflected a more systematic approach when the previous response was incorrect, and was negatively associated with rigid thinking, for this subgroup the opposite was true. This factor related to a rigid, perseverative approach.

Factor 5 for this subgroup was also similar to Factor 5 for the full group with respect to factor composition, but whereas for the full group a rigid approach on

Subtests III and IV varied together, for this subgroup those variables (NC3 and NC4) were negatively related. Consequently, this factor appeared to represent subjects who identified the location principle in Subtest III but experienced difficulty giving up the counting rule from Subtest II, so used the location and counting principles together, failing to benefit from feedback in shifting to a more effective approach. This confused thinking, however, did not extend to Subtest IV for this group.

Table 25

Factor Pattern Matrix for Unimpaired Subgroup on Subtest VI

Variable Number:	1	2	3	4	5
Percent of Variance:	37.0%	19.9%	9.0%	6.6%	5.4%
RKA	.98				
RPCK	.95				
SLR	.95				
RKC	.94				
MSBS	.90				
MSIS	.89				
LR	-.86				
WS1		-.86			
WPCS		.85		.33	
WSC		.81			
WSA		.67		-.35	
WS3		.54	-.44	.35	
WS2			.81		
PRAC4			.72		.31
WSR	.31			.82	
PSV5-6	-.40			.68	
RKR	.56			-.57	
NC3					.78
NC4					-.46

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

Total Score

Factor analysis of cognitive process variables from each CT subtest individually resulted in factor solutions of five or six factors in each instance which were quite similar in many respects, although rather discrepant results were obtained for Subtest III. Many of the similarities within subtests remained even when separate factor analyses were conducted for impaired versus unimpaired subjects, again except for Subtest III, where several unique factors relating to ineffective problem-solving emerged for the impaired subgroup. However, this was not surprising since this was the most difficult subtest (see Table 9), and it might be anticipated that subjects would encounter difficulty on this task for a variety of reasons.

What was more surprising was that the remaining subtests showed as much similarity as was evident, both across subtests and across levels of impairment. Consequently, mean scores across subtests were also factor analyzed to determine if these patterns would emerge with respect to total scores for the CT as a whole.

The factor pattern for mean scores across all subtests is shown in Table 26. Factor 1 for this group was comprised of substantial loadings on all maintaining set variables, with a more moderate loading for RKR. PRAC4 also showed a weak loading on this factor, perhaps since a high score on that variable, though suggesting slow learning, was also associated with successful learning within six trials. Further, WSR loaded moderately on this factor, perhaps reflecting the rule shifting across subtests which was related to effective performance in terms of total errors.

Factor 2 represented nonsystematic shifting of determinants when the previous response was incorrect, as was true of the second factor in all other subtests as well (except Subtest III). Factor 3 appeared to represent a mixture of responses types, with WS2 loading moderately, reflecting a nonsystematic approach, but also PSV5-6 loaded prominently in a negative direction, suggesting that this factor did not reflect random responding. All of the rigid thinking variables loaded to a substantial degree on Factor 4; thus, this appeared to reflect some perseverative responding.

Since factors which emerged from mean cognitive process scores were not appropriate to enter in subsequent factor analyses with subtest specific factors

Table 26

Factor Pattern Matrix for Full Group on Total Score

Variable Number: Percent of Variance:	1 38.0%	2 17.6%	3 11.0%	4 6.6%
RPCK	.97			
RKC	.96			
MSIS	.95			
SLR	.93			
MSBS	.90			
RKA	.87			
LR	-.87			
PRAC4	.35			.32
WPCS		.92		
WSA		.82	.34	
WSC		.73		
WS1		-.68		
WS3		.66	-.41	
PSV5-6			-.83	
RKR	.58		.70	
WSR	.50	.36	-.65	
NC4				.73
NC3				.52
WS2			.47	.50

Note. Variable numbers and labels in left hand column refer to variable definitions listed in Appendix H.

(since they share common variance), this group was not subdivided into high and low scorers for further analysis.

<u>Second Order Factor Analysis Across Subtests</u>

Although intuitively many similarities were evident among factors which emerged in the analyses, it did not follow that factors composed of similar variables across subtests were necessarily related. To assess the relationship among cognitive process factors from one subtest to another, second order factor analyses of subtest factors was performed. However, it was necessary to recompute factor scores, deleting the marker variables which were entered in each analysis described above in order to prevent the repetitive scores on these variables form masking the relationships among the remaining variables. Such recomputation did not substantially alter factor composition or loadings, except where those variables which were deleted comprised the principal determinant of the factor. Thus, Factor 5 was eliminated from the Subtest III analysis, Factor 5 was eliminated from Subtest IV, Factors 5 and 6 were eliminated from Subtest V, Factor 5 was eliminated from Subtest VI, Factor 3 was merged with Factor 2, and Factor 4 became the third factor. Additionally, the direction of factor loadings changed in two instances: Factor 4 in Subtest III, which previously reflected variance attributed to subjects who perseverated on the concrete features of the stimuli, changed to reflect variance attributed to subjects who did not perseverate on concrete features, and a similar change occurred on Factor 4 in Subtest V, which now became characterized by those who shifted two scores simultaneously, often tending to shift C. None of these changes represents substantial interpretive significance.

The results of second order factor analysis of the remaining factors are listed in Table 27. Variable numbers comprising the subtest specific factors (from

Table 27

Second Order Factor Analysis of All Subtest Factors

Var. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Grd. F1	**-.9**	**-.8**	**-.4**	**.4**	**.3**														
(V: F3)						-.3	.7			.7	-.6								-.9
(VI: F3)							.5				-.7		.8	-.6	-.4				
(IV: F4)												-.4	.8					.4	
IV: F1	.9	.8	-.9	.8	.9	.9	.8	1.0			.5							.5	
III: F1	.3		-.7	.8		.9	.6	.8	-.3		.5		-.3			-.5			
Grd. F2	**.9**	**.9**	**.5**	**-.3**															
V: F1	1.0	.7	-.9	.9	1.0	.9	.3	1.0	-.4		.4								
VI: F1	.9	.9	-.9	1.0	1.0	.9	.6	1.0											
IV: F1	.9	.8	-.9	.8	.9	.9	.8	1.0			.5							.5	
(III: F3)			.3				.3		-.7	-.3	-.6		.6	-.8	-.7				
Grd. F3	**-.8**	**-.3**	**-.3**	**-.3**	**.5**	**.5**	**.3**												
(V: F2)									.3	.4	.6	-.6		1.0	.9				
(IV: F2)									.8	.9	.4	-.8		.7	1.0				
(V: F 3)						-.3	.7			.7	-.6								-.9
(IV: F3)							-.5				.6							.8	
V: F4									.4			-.4	1.0	-.4					
IV: F4												-.4	.8					.4	
VI: F3							.5				-.7		.8	-.6	-.4				
Grd. F4	**.7**	**.6**	**.6**	**.3**															
III: F3			.3				.3		-.7	-.3	-.6		.6	-.8	-.7				
III: F2										.8	.6	-.9			.5				
III: F1	.3		-.7	.8		.9	.6	.8	-.3		.5		-.3			-.5			
V: F4									.4			-.4	1.0	-.4					
Grd. F5	**-.6**	**-.5**	**-.4**	**.6**															
(IV: F2)									.8	.9	.4	-.8		.7	1.0				
(IV: F4)												-.4	.8					.4	
(IV: F1)	.9	.8	-.9	.8	.9	.9	.8	1.0			.5							.5	
VI: F2									.6	.7	.4	-.9	.4	.5	.9				
Grd. F6	**-.7**	**-.7**																	
(III: F5)		.9					.3									-.3			
(IV: F3)							-.5				.6							.8	
Grd. F7	**.9**	**.4**	**.3**																
III: F4	.9				.9			.4											
III: F3			.3				.3		-.7	-.3	-.6		.6	-.8	-.7				
IV: F3							-.5				.6							.8	

Note. All subtest factor loadings above .3 are listed. Negative loadings are in parentheses. 2nd order factor loadings are in bold print in the order listed.

Appendix H) entered in the analysis are listed across the top row of the table. Similar variables in Appendix H are numbered consecutively. Thus, if similar variables loaded on common factors across subtests, several subtests should be represented, with factor loadings forming a cluster, under each second order factor.

It can be seen that some clustering was evident but the dispersion of loadings was more prominent than the clustering. Further, many of the clustering patterns which were evident tended to be specific to sets of similar subtests, rather than common to all subtests.

Thus, this analysis suggested that many of the cognitive process variables were task specific rather than generalizable across subtests. For example, on second order Factor 1, Variable 11 (WSR) loaded positively on Factor 3 for Subtests V and VI but negatively on Factor 1 for Subtests III and IV. Statistical analysis of the correlates of these second order factors, therefore, appeared appropriate.

Relationship of Cognitive Process Factors to Error Scores

The relationship between second order factor scores and performance on each CT subtest and for CT Total Errors was assessed using stepwise multiple regression analyses. Those results revealed that second order factors significantly predicted error scores in each case ($p < .00009$), but the proportion of the variance which was accounted for varied substantially, although Factor 2, consisting predominantly of Maintaining Set scores, accounted for most of the variance across all scores except for Subtest III. (See Table 28.)

Table 28

Adjusted R^2 for 2nd Order Factors on CT Error Scores

Predictors	III		IV		V		VI		TOTAL	
Factor 1	(1)	.17	(1)	.26					(2)	.25
Factor 2	(2)	.12	(2)	.30	(1)	.47	(1)	.64	(1)	.48
Factor 3										
Factor 4	(3)	.09							(3)	.02
Factor 5			(3)	.04						
Factor 6							(2)	.01		
Factor 7										
Total Var.		.36		.58		.46		.65		.75

Note. Step at which predictor entered equation is given in parentheses.

These results represented some improvement in prediction of subtest error scores from error scores on corresponding subtests, but the improvement was not dramatic, as shown by Table 29.

Second order factor scores did show a very significant relationship with CT performance, but, as Table 28 demonstrates, this relationship was comprised almost entirely of variance due to Maintaining Set scores. The previously documented relationship of age, education, and IQ variables with CT performance might have accounted for these findings if Maintaining Set factor scores were influenced to a substantially greater degree by covariates, as is likely. Consequently, the foregoing analyses were recomputed, with age, education, WAIS-R scale scores, and Full Scale sum, Verbal Scale sum, and Performance Scale sum entered into the equation first, to account for variance from these sources on CT error scores. Results are listed in Table 30.

Although partialing variance due to age, education, and IQ scores resulted in marginally improved predictability of error scores, the proportion of variance accounted for by each factor remained essentially unchanged. Thus, Maintaining

Table 29

Adjusted R^2 for Subtest Error Scores with Covariates Partialed Out

Predictors		III		IV		V		VI
Total Covar.		.18		.27		.24		.18
III			(1)	.32				
IV	(1)	.24						
V							(1)	.52
VI			(2)	.37	(1)	.56		
Total Var.		.24		.37		.56		.52

Set factor scores continued to account for nearly all of the variance in error scores.

This pattern was also evident among cognitive process factors specific to individual subtests, as shown by the results of multiple regression analysis listed in Table 31. In each case, age, education, and IQ covariates, plus Maintaining Set scores, accounted for almost all of the predicted variance in error scores.

It is not surprising that discovering the correct principle for each subtest and consistently applying that principle across sets of dissimilar stimuli accounted for most of the variance in CT performance (although to a lesser degree in

Table 30

R^2 Change for CT Error Scores with Covariates Partialed Out

Predictors		III		IV		V		VI		TOTAL
Total Covar.		.18		.27		.24		.18		.36
Factor 1	(1)	.10	(1)	.15					(2)	.16
Factor 2	(3)	.07	(2)	.19	(1)	.28	(1)	.48	(1)	.23
Factor 3									(4)	.01
Factor 4	(2)	.07	(4)	.01			(2)	.01	(3)	.01
Factor 5			(3)	.02						
Factor 6										
Factor 7										
Total Adj. R^2		.41		.63		.52		.66		.77

Note. Number in parentheses reflects step at which factor entered equation.

Table 31

R^2 Change for Subtest Error Scores with Covariates Partialed Out

Predictors		III		IV		V		VI		TOTAL
Total Covar.		.18		.27		.24		.18		.35
Factor 1	(1)	.48	(1)	.47	(1)	.33	(1)	.63	(1)	.33
Factor 2	(4)	.01			(3)	.04			(4)	.01
Factor 3	(2)	.10			(2)	.03	(2)	.01	(2)	.07
Factor 4	(3)	.03	(2)	.05	(4)	.03			(3)	.04
Factor 5										
Total Adj. R^2		.81		.80		.66		.82		.82

Note. Number in parentheses reflects step at which factor entered equation.

Subtest V). However, a much more relevant issue from a treatment perspective is how patients fail to discover the correct principle and maintain set.

Although the analyses described above revealed some patterns characteristic of such failures, it was unclear how those patterns related to CT performance. For that reason, variables related only to failures were factor analyzed separately from Maintaining Set variables. Those results were very similar in factor composition and factor loadings to the full model described above, but the relationship of these factors and CT performance were rather different, as shown by the results of multiple regression analyses listed in Table 32.

Since previous analyses had shown that Maintaining Set scores were strongly related to the dependent variable, in order to assess the relationship between error scores and factors related to incorrect responding independently of Maintaining Set factor scores, the former factors were entered into the analyses first, following covariates, with Maintaining Set factors entered last.

This procedure revealed a substantially greater relationship between error scores and factors related to incorrect responding than was true when the same procedure was applied to the full model. In the latter instance, for each subtest

Table 32

R^2 Change for Subtest Error Scores with Covariates Partialed Out

Predictors		III		IV		V		VI		TOTAL
Total Covar.		.18		.27		.24		.18		.36
Factor 1 Wrong	(1)	.09	(3)	.02	(2)	.04			(3)	.03
Factor 2 Wrong			(1)	.28	(1)	.08			(1)	.12
Factor 3 Wrong	(2)	.11	(2)	.06			(2)	.05	(2)	.04
Factor 4 Wrong	(3)	.12	(4)	.01	(3)	.02				
Factor 5 Wrong							(1)	.11		
Factor 1 Right	(4)	.26	(5)	.17	(4)	.27	(3)	.47	(4)	.25
Factor 2 Right	(5)	.03								
Total Adj. R^2		.79		.82		.65		.81		.81

and for Total Errors, even when Maintaining Set scores were forced to enter the equation last, following the stepwise block for factors related to incorrect responding, all of the variance in error score prediction was accounted for by Maintaining Set factors and none of the factors related to incorrect responding entered the equation, with probability to enter set at .05. Nonetheless, these factors were virtually equivalent, as shown by the correlations in Table 33 between factors in the full model (across the top row) and corresponding factors from the separate models.

This discrepancy appeared to be the result of differential relationships between these sets of factors and Maintaining Set factor scores due to the effects of oblique axes rotation in the full model. Even though in oblique rotation (oblimin) factors are not constrained to independence, the effect of maximizing factor loadings served to minimize correlations between these factors. Thus, due to the strong relationship between the dependent variable (error scores) and Maintaining Set factor scores, the correlation between error scores and Incorrect Responding factor scores was also minimized.

Table 33

Correlation of Full Model With Separate Right/Wrong Factors

Subtest	III					IV				V				VI			TOTAL			
Factors	1	2	3	4	5	1	2	3	4	1	2	3	4	1	2	3	1	2	3	4
F1 Wrong			.89				.98				.94				.98			.97		
F2 Wrong		.97											.87			.95			.86	
F3 Wrong								.88				.93								.97
F4 Wrong									.96											
F1 Right	.91					.97				.95				1.0			.96			
F2 Right				.94																
F3 Right					.91															

In fact, when factor scores for the full model were recomputed without rotation, a much stronger pattern of relationships between error scores and Incorrect Responding factor scores emerged (except for Subtest VI), as shown in Table 34.

Thus, the failure of Incorrect Responding factor scores to significantly predict error scores appeared to be artifactual rather than reflecting the absence of relationships. Nonetheless, the cognitive processes reflected by these factor scores, and to an even greater extent the second order factor scores, remained ambiguous. In part, this was because specific variable scores did not necessarily reflect a uniform approach within subtests, even for an individual subject. Further, task demands varied from subtest to subtest; consequently, variable scores did not reflect similar processes across subtests. Despite this ambiguity, however, factor scores may be more accurate predictors of CT performance than were the error scores themselves since error due to "incidentally correct" responses was partialed out. Further, error scores were substantially influenced by age, education, and IQ covariates, while the previous analyses have shown that partialing covariates did not substantially moderate the predictive power of

Table 34

R^2 Change for Unrotated Factor Scores with Covariates Partialed Out

Predictors	III	IV	V	VI	TOTAL
Total Covar.	.18	.27	.24	.18	.36
Factor 1 Wrong	(3) .08	(4) .01	(1) .10		
Factor 2 Wrong	(1) .20	(1) .24			(1) .19
Factor 3 Wrong	(2) .08	(3) .03	(2) .04		
Factor 4 Wrong		(2) .09			
Factor 1 Right	(4) .24	(5) .17	(3) .28	(1) .63	(2) .25
Factor 2 Right	(5) .02				
Total Adj. R^2	.81	.82	.65	.81	.81

factor scores. Thus, further investigation of covariate relationships appeared warranted.

Covariates of Cognitive Process Factors

All covariates in the previous analyses (age, education, WAIS-R scale scores, Verbal Scale Sum, Performance Scale Sum, and Full Scale Sum) were regressed on each first and second order cognitive process factor. These results are listed in Table 35. Covariates accounted for a slightly greater proportion of the variance in error scores for each subtest than was true for the cognitive process factors. This discrepancy was not substantially greater for the Maintaining Set factor in each subtest, but was greater for all other process variables by a significant margin, except in the case of Factor 3 in Subtest VI. This pattern suggested the possibility that the relationship of covariates with cognitive processes related to maintaining set may be principally responsible for the relationship of covariates with error scores, since Maintaining Set factors (Factor 1) in each subtest accounted for the largest single portion of the variance in subtest error scores in the multiple regression analyses listed in Table 35.

Table 35

R^2 for Age, Education, and IQ Regressed on Cognitive Factors

Covar.	Age	Ed	I	DS	A	V	C	S	PC	PA	BD	OA	DSy	VS	PS	FS	Tot.
III	.17								.11				.20				**.20**
Fact. 1	.12							.07									**.12**
Fact. 2																	**0**
Fact. 3		.10						.07				.04					**.10**
Fact. 4												.07					**.07**
Fact. 5										.04		.11					**.11**
IV		.20					.24			.17						.29	**.29**
Fact. 1	.16	.21			.24		.26								.13		**.26**
Fact. 2										.04				.11			**.11**
Fact. 3																	**0**
Fact. 4										.04							**.04**
V											.25				.21		**.25**
Fact. 1					.18						.15						**.18**
Fact. 2																	**0**
Fact. 3					.06												**.06**
Fact. 4																	**0**
VI																.18	**.18**
Fact. 1	.17								.13							.10	**.17**
Fact. 2																	**0**
Fact. 3						.16							.12			.09	**.16**
Total	.37					.35										.27	**.37**
Fact. 1	.28								.30							.22	**.30**
Fact. 2	.04																**.04**
Fact. 3																	**0**
Fact. 4																	**0**
2/O F1									.08								**.08**
2/O F2	.17								.20							.14	**.20**
2/O F3																	**0**
2/O F4								.05				.10					**.10**
2/O F5																	**0**
2/O F6																	**0**
2/O F7							.03										**.03**

Note. All subtest factor loadings above .3 are listed. Negative loadings are in parentheses. 2nd order factor loadings are in bold print in the order listed.

The other principal finding to emerge with respect to these analyses related to the dispersion of correlates across subtests and across factors within subtests. Given the weak relationship of many of these variables, considerable inconsistency should be expected. However, even when the relationship was stronger, considerable variability was evident.

The influence of age was a notable exception to that general rule, since age significantly predicted Subtest III errors and Factor I scores (Maintaining Set), as well as Maintaining Set scores on Subtests IV and VI, Maintaining Set Scores Total, and second order Factor 2 scores (Maintaining Set on Subtests IV, V, and VI). However, all other covariates were quite inconsistent across factors and across subtests. This finding appeared to reflect the ambiguity inherent in interpreting the cognitive process factors.

Hypothesis 4

It was also anticipated that factor analysis of verbalized responses on the WCST would produce three or more interpretable cognitive process factors which would differentiate among problem-solving approaches employed on this task as well. Coding verbalized responses on the WCST did not present the challenges which were present for the CT. Conventional scoring for the WCST is based upon a process analysis of subjects' problem-solving approaches on this task. The only ambiguities present on this task relate to uncertainty which determinant is instrumental in determining card placement when cards match on two or three determinants. Asking patients to verbalize their reasoning was an attempt to better clarify process analysis by reducing the proportion of ambiguous responses without simplifying the task. However, in practice it was found that many subjects verbalized two or more determinants for their responses; consequently, such verbalized responses conveyed no more information than did conventional responses.

Comparison of Verbalized versus

Nonverbalized Scoring

If verbalized administration was successful in clarifying reasoning on this task, verbal process scores should correlate only moderately or below with conventionally scored responses but summary scores (Categories Completed, Total Errors) would show very high correlation coefficients. Consequently, these sets of scores were correlated. Results are listed in Table 36. Due to space limitations, scores are designated by the order in which they were listed in Appendix I. These correlations were all reasonably high, with the notable exception of Right-Shift responses (#19), and to a lesser degree Missed Learning Opportunities (#5) and Percent Nonperseverative Errors (#6).

Consequently these scores were examined to determine the source of these discrepancies. (See Table 37.) Inspection of the data revealed that the low correlation between verbalized and conventional Right-Shift scores resulted from restricted range on this score under both scoring conditions. Consequently, this score was dropped from further analysis. Reduced correlations for the other two variables, however, appeared to reflect legitimate differences in the values of these scores between scoring conditions. Little difference in mean scores for Nonperseverative Error Percent was evident, but considerably greater variability was present under verbalized scoring conditions. Thus, it appeared that for some subjects verbalizing their reasoning substantially reduced ambiguity in

Table 36

Correlations Between Verbalized and Conventional WCST Scores

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
.92	.95	.82	.90	.73	.79	.84	.87	.92	.86	.93	.97	.86	.94	.91	.92	.69	.90	.26	.89

Table 37

Descriptive Data for Right-Shift, Missed Lrng Oppor., and NPSV%

Variables	N	Mean	S.D.	Min.	Max.
Right-Shift Conventional	132	.64	.76	0	3
Right-Shift Verbalized	132	.10	.30	0	1
Missed Lrng Oppor., Conven.	132	7.69	8.08	1	33
Missed Lrng Oppor., Verbal.	132	13.58	11.11	2	42
NonPSV Error %, Conven.	132	12	8	1	41
NonPSV Error %, Verbal.	132	13	12	0	60

scoring perseverative versus nonperseverative responses, but this was not true for all subjects. Conventional scoring resulted in fewer missed learning opportunities apparently because of a higher proportion of ambiguous responses, which reduced scoring opportunities. Therefore, these latter two scores were retained for further analysis.

Factor Analysis of WCST Scores

Verbalized administration. The factor pattern for factor analysis of all retained scores from verbalized WCST administration is described in Table 38. Only three factors emerged in this analysis, accounting for 82.5% of the variance.

Factor 1, which accounted for the largest proportion of the variance by far was comprised of substantial loadings on all major variables, and appeared to reflect ineffective, perseverative responding. Summary scores reflecting nonperseverative responding (Conceptual Level Responses, Categories Completed, and Learning to Learn) loaded negatively on this factor. Interestingly, Unambiguous Form responses and Unambiguous S.D. both loaded very highly on this factor, and Unambiguous Color responses loaded more moderately, but Unambiguous Number showed only a weak loading, below .30. This pattern was also evident with respect to maximum classification (Maximum Color, Form, Number), but at a lower level. In contrast, Factor 2 was comprised

Table 38

Factor Pattern Matrix for WCST Verbalized Administration

Factor Number: Percent of Variance:	Factor 1 57.4%	Factor 2 16.9%	Factor 3 8.2%
Total Errors %	.99		
Unambiguous Form	.96		
Unambiguous S.D.	.93		
Concept. Level Resp.	-.92		
Perseverative Error %	.91		
Perseverative Resp. %	.91		
Wrong-Stay	.90	.41	
Max. Determ. S.D.	.86	-.29	
NonPSV Error %	.79		
Categories Completed	-.77		.32
Unambiguous Color	.70	-.56	
Maximum Form	.69		-.26
Learning to Learn	-.67	.37	
Unambiguous No.		.80	
Maximum Number		.78	-.42
Missed Lrng. Oppor.	.31	.56	-.41
Fail to Maintain Set			-.88
Total Correct %		.30	-.87
Maximum Color	.52		-.61

predominantly of Unambiguous Number and Maximum Number, but with a negative loading for Unambiguous Color (at a moderate level), and almost no relationship with Unambiguous Form.

This pattern strongly suggested that when the proportions of determinant responses were unequal, whether due to ambiguous or unambiguous responses, color and form responses were associated with perseveration, while unequal number responses tended not to be associated with clear performance patterns, other than a mild inclination to repeat unsuccessful responses, in a nonperseverative manner.

Factor 3 accounted for a much smaller proportion of the variance and related very mildly to effective nonrigid problem-solving, with high negative loadings on variables relating to failure to maintain set.

Conventional administration. The factor pattern for conventional WCST scoring is shown in Table 39. Three factors also emerged in this analysis, which were almost identical to the verbal scoring factors, and accounted for similar proportions of the variance.

Consistent with the correlation patterns described in Table 36, differences between Factor 1 for the two sets of scores (r = .85) emerged only with respect to Missed Learning Opportunities and Nonperseverative Error Percent. In the former case, the loading for Factor 1 increased substantially, from .31 to .71, and in the latter case the loading decreased from .79 to .59. Both of these changes were in the predicted direction, consistent with greater accuracy in scoring responses with verbalized administration, although hardly sufficient to justify

Table 39

Factor Pattern Matrix for WCST Conventional Scores

Variables	Factor 1 66.2%	Factor 2 11.1%	Factor 3 9.2%
Perseverative Error %	.99		
Perseverative Resp. %	.99		
Unambiguous Form	.95		
Unambiguous S.D.	.94		
Total Errors %	.90		
Wrong-Stay	.88	.31	
Concept. Level Resp.	-.84		
Max. Determ. S.D.	.82		
Unambiguous Color	.72	-.36	
Missed Lrng. Oppor.	.71	.34	
Maximum Form	.68		-.36
Learning to Learn	-.68		.35
Categories Completed	-.66		.36
NonPSV Error %	.59	.34	
Unambiguous No.		.93	
Maximum Number		.90	
Fail to Maintain Set			-.98
Total Correct %		.32	-.76
Maximum Color	.51		-.55

the change in administration procedure. Loading and factor composition for Factors 2 and 3 for the two administration conditions were also very similar (r = .77; r = .77), and in the predicted direction. In neither case did loading changes significantly modify interpretive conclusions.

Covariates for WCST Cognitive Process Factors

Despite the similarities between factors which emerged from the two sets of scores, the relationships with age, education, and IQ covariates were considerably different for Factor 1 across scoring conditions, although the remaining factors in each condition did not show any substantial relationship with covariates. Results of multiple regression analysis for covariates on each factor are listed in Table 40.

The proportion of the variance accounted for was somewhat greater for verbal scores, but more striking was the very different pattern of covariation between corresponding pairs of scores, suggesting that these factors may not be equivalent measures. In fact, analysis of covariance, partialing out all age, education, and IQ covariates, did not substantially change these correlation

Table 40

R^2 for Age, Education, and IQ Regressed on WCST Factors

Covar.	Age	Ed	I	DS	A	V	C	S	PC	PA	BD	OA	DSy	VS	PS	FS	Tot.
Verbal																	
Fact. 1	.24				.30						.27			.21		.14	.30
Fact. 2																	0
Fact. 3		.05															.05
Conv.																	
Fact. 1									.19				.15				.19
Fact. 2									.04								.04
Fact. 3																	0

coefficients between similar factors from different scoring conditions (Factor 1, r = .84; Factor 2, r = .76; Factor 3, r = .80). The relative equivalence of these coefficients suggested that similar factor scores did not share common covariance.

In an attempt to discern the source of differences in covariation between the two sets of scores, multiple regression analyses were performed. Age, education, and IQ covariates were regressed onto each of the original WCST variables which loaded on these factors. These results are listed in Table 41. Somewhat different patterns of covariation emerged from these analyses as well. Only 7 of 20 comparisons showed very similar patterns of covariation, and an additional 3 comparisons were somewhat similar. The remainder were quite dissimilar, although some similarities were evident.

The proportion of dissimilar findings is far beyond chance and strongly suggested that these sets of scores may reflect different cognitive processes. This interpretation was strengthened by the fact that Verbal Scale Sum (VS) did not significantly predict variance on any variable from conventional scoring but did in four instances for verbalized scores (Categories Completed, Conceptual Level Response, Perseverative Responses, and Total Errors). It should be noted that these latter variables were all among the major contributors to Factor 1 under both sets of scores. The opposite pattern was obtained with respect to covariation with Performance Scale Sum, which covaried with three conventional scores, but not verbalized scores.

Additionally, age and Digit Symbol Scale Score (DSySS) covaried to a substantially greater degree with conventional scores than verbal scores. These differences all led to the conclusion that asking patients to verbalize their reasoning on the WCST reflected very different skills than asking them to

Table 41

R^2 for Age, Education, and IQ Regressed on WCST Factors

Covar.	Age	Ed	I	DS	A	V	C	S	PC	PA	BD	OA	DSy	VS	PS	FS	Tot.
Conv.																	
Cat.															.12		
CLR	.28												.56			.24	
FTMS																	
LTL															.09		
MLOP	.23															.20	
NPSVE	.25															.22	
PSVR													.20		.24		
TE	.29												.31			.24	
TC																.12	
Max C																.20	
Max F	.23												.26			.17	
Max N																.11	
Max SD	.16															.13	
Sin. C																.13	
Sin. F								.17					.14				
Sin. N																.10	
Sin. SD	.20															.17	
Wrg-St									.26		.31	.28	.19				
Verbal																	
Cat.														.19		.15	
CLR											.57			.30		.24	
FTMS		.06															
LTL															.06		
MLOP												.10					
NPSVE																.16	
PSVR	.31										.33			.28		.17	
TE														.27		.21	
TC																.09	
Max C																.18	
Max F									.14		.25	.16	.22				
Max N												.04					
Max SD	.19															.15	
Sin. C																.13	
Sin. F													.10				
Sin. N																.07	
Sin. SD																.15	
Wrg-St								.19				.14					

perform the same task without verbalization. It was not clear, however, whether performance varied as a function of verbalizing reasoning, or whether only the verbal scores varied. Consequently, it is not known at this point which procedure may be best suited to measuring the cognitive processes involved in attribute identification and cognitive flexibility. Some light may be shed concerning those relationships by investigating the relationships between the WCST and CT factors, while controlling for covariates.

Comparisons Between the Category Test and the WCST

Previous studies have found that the CT and WCST share approximately 30% common variance, using conventional scoring. Other studies have found less significant relationships when partialing out age, education, and IQ covariates. In this study, the relationship between performance on these individual tasks was assessed using multiple regression analysis, entering covariates first to partial out their influence on the relationship between CT and WCST scores. Those results are shown in Table 42.

Although each dependent variable (with only two exceptions) on both the WCST and CT were significantly predicted statistically by at least one variable from the corresponding test, the increase in predictive power in each case was

Table 42

Prediction of CT/WCST Variables, Controlling for Covariates

Dependent Var.	Covar. Adj. R^2	Independent Var.	R^2 Change	Sig F Chg
CT Total Errors	.31	WCST Categories Completed	.03	.0139
CT Subtest III Errors	.13	None Entered		
CT Subtest IV Errors	.22	WCST Max. Class SD	.05	.0040
CT Subtest V Errors	.25	WCST Total Correct	.04	.0416
CT Subtest VI Errors	.20	WCST Categories Completed	.09	.0002
WCST Categories Comp.	.16	CT Subtest VI Errors	.09	.0003
WCST Total Errors %	.29	None Entered		
WCST Non PSV Errors %	.24	CT Subtest VI Errors	.03	.0440
WCST PSV Responses %	.23	CT Subtest IV Errors	.02	.0484
WCST PSV Errors %	.24	CT Subtest IV Errors	.02	.0485
WCST FTM Set	.03	CT Subtest VI Errors	.04	.0176
WCST Learn to Learn	.12	CT Subtest VI Errors	.06	.0027

nonsubstantial. These findings are very consistent with other reports in the literature.

It was anticipated that factor scores derived from cognitive process variables on each of these tasks would predict a significantly greater proportion of the variance than using conventional scores alone on the other task. This prediction

Table 43

Prediction of CT/WCST Factor Scores, Controlling for Covariates

Dependent Var.	Covar. R^2	Independent Var.	R^2 Change	Sig F Chg
CT Subtest III Factor 1	.08	WCST Maximum Color	.03	.0394
CT Subtest III Factor 2		None Entered		
CT Subtest III Factor 3	.05	None Entered		
CT Subtest III Factor 4	.04	None Entered		
CT Subtest III Factor 5	.14	None Entered		
CT Subtest IV Factor 1	.19	WCST Verbal Factor 1	.03	.0453
		WCST Verbal Missed Lrng Oppor.	.03	.0255
CT Subtest IV Factor 2	.08	WCST Conventional Factor 3	.03	.0464
CT Subtest IV Factor 3		None Entered		
CT Subtest IV Factor 4	.12	None Entered		
CT Subtest V Factor 1	.19	WCST Verbal Factor 1	.03	.0238
		WCST Verbal Unambiguous Form	.04	.0101
CT Subtest V Factor 2		None Entered		
CT Subtest V Factor 3	.04	None Entered		
CT Subtest V Factor 4		None Entered		
CT Subtest VI Factor 1	.14	WCST Verbal Learning to Learn	.06	.0025
CT Subtest VI Factor 2		None Entered		
CT Subtest VI Factor 3	.14	None Entered		
CT Total Factor 1	.20	WCST Verbal Factor 1	.05	.0052
		WCST Verbal Factor 3	.04	.0122
		WCST Verbal Total Errors Percent	.03	.0178
CT Total Factor 2		None Entered		
CT Total Factor 3		None Entered		
CT Total Factor 4		None Entered		
CT 2nd Order Factor 1	.06	None Entered		
CT 2nd Order Factor 2	.13	WCST Verbal Factor 1	.04	.0141
		WCST Verbal Factor 3	.03	.0389
CT 2nd Order Factor 3		None Entered		
CT 2nd Order Factor 4		None Entered		
CT 2nd Order Factor 5		None Entered		
CT 2nd Order Factor 6	.11	None Entered		
CT 2nd Order Factor 7	.03	None Entered		
WCST Factor 1	.26	CT 2nd Order Factor 2	.03	.0354
WCST Factor 2		None Entered		
WCST Factor 3	.04	None Entered		

was assessed utilizing multiple regression procedures, shown in Table 43, controlling for the influence of covariates, as described above. For CT factors as dependent variables, covariates were regressed first, followed by WCST factor scores from verbal administration, then WCST factor scores from conventional administration, followed by all individual WCST variables.

Results of these analyses showed even less relationship between factor scores from one task and factor scores from the other task. Moreover, using variables as predictors for factor scores failed to improve prediction. These analyses provided no evidence that the tasks assessed by these tests are comparable in substantive ways.

CHAPTER V
DISCUSSION

The objectives of this study were threefold: (a) determine whether asking patients to verbalize their reasoning on WCST and CT responses substantially alters the nature of these tasks; (b) devise a coding system derived from these verbal responses and determine if verbal responses obtained using this methodology can be reliably coded by trained examiners; and (c) determine the relationship between the cognitive processes reflected by scores on the WCST versus the cognitive processes reflected by scores on the CT. The results described in the previous chapter will be described in terms of the research hypotheses which address each of these objectives.

Comparison of Verbalized versus Conventional Administration

Prior validation studies have shown that the CT and WCST are among the most sensitive indicators of brain damage (Kløve, 1974). Departures from standard administration procedures for these tasks, therefore, must be undertaken cautiously to avoid modifying the nature of the skills being assessed. For that reason, it was necessary to establish that the results obtained using standard administration procedures were comparable to results obtained when patients were asked to verbalize their reasoning on these tasks, when both groups were comparable in other relevant respects. The specific hypotheses tested were (a) Mean Total Error and subtest error scores on the CT and mean Categories Completed, Total Errors, Nonperseverative Errors, Perseverative Errors, Perseverative Responses, and Failure to Maintain Set scores on the WCST will not differ significantly (statistically) between a group of patients who verbalized the reasoning for their responses and a group of patients who did not verbalize their reasoning, and are comparable to the first group with respect to diagnosis, age, education, and Full Scale WAIS-R IQ; and (b) Correlations between Mean Total Error and subtest error scores on the CT and mean Categories Completed, Total Errors, Nonperseverative Errors, Perseverative Errors, Perseverative Responses, and Failure to Maintain Set scores on the WCST will not differ significantly (statistically) between a group of patients who verbalized the reasoning for their responses and a group of patients who did not verbalize their reasoning, and are comparable to the first group with respect to diagnosis, age, education, and Full Scale WAIS-R IQ.

Hypothesis 1 was tested using ANCOVA to compare scores between groups while controlling for age, education, and IQ covariates. Results revealed that only CT Subtest IV Errors (C4Err) significantly differed ($p < .02$) between the groups. Further analysis of covariates for C4Err revealed that covarying only age increased comparability on C4Err across groups. Treating both age and group membership as independent variables further increased comparability between groups, so that main effects were found only for age, as expected. The effects of group membership were nonsignificant ($p = .16$), and no interaction effects for age and group membership were apparent. Consequently, comparison of mean scores across groups did not reveal group differences, when controlling for age, education, and IQ covariates.

Rather different results were found for Hypothesis 2, however. ANCOVA was also utilized for testing differences in correlation coefficients between all possible pairs of CT and WCST variables, while controlling for age, education, and IQ covariates. Significant differences were found for correlations between three pairs of variables. Fifty-five comparisons were computed; consequently, finding significant differences for three of these may have been due to chance.

However, further analysis of these differences revealed different patterns of age, education, and IQ covariation between pairs of identical scores across groups. These results suggested that groups may not be comparable, despite finding no significant differences between mean scores and between most correlations for sets of scores across groups. Comparing relationships among WAIS-R subtest and summary scores also revealed substantial differences between groups with respect to relationships among scores.

Since administration conditions for the WAIS-R were identical for both groups, in view of differences among relationships for these scores between groups, it cannot be concluded that differences in relationships for CT and WCST variables is the result of administration conditions. Rather, the groups may not be comparable with respect to diagnosis, severity, or other relevant variables.

These results are not surprising, since data were obtained from archival records, and subjects were not randomly assigned to administration conditions. Systematic differences between groups may have occurred because different examiners tested subjects in each group, and decisions as to which test instruments to administer were made by individual examiners based on clinical judgment. It is likely that these judgments were not based on consistent criteria. Therefore, generalization of the results of this study to subjects who were administered these test instruments under conventional administration conditions should be considered tentative.

Reliability of Coding Verbalized Responses

Asking patients to verbalize the reasoning underlying quantifiable responses introduced the problem of how one patient's qualitative responses may be compared to other patients' qualitative responses. Some anecdotal reports are available for the CT in which patients have been asked to verbalize their reasoning (Reitan & Wolfson, 1985), but these responses have been interpreted subjectively, which does not allow systematic comparisons across patients or across responses.

Perrine (1984) referred to the concept formation literature (Bourne et al., 1979) for a cognitive process model which might help clarify the cognitive processes on the CT and WCST. For the purposes of this study, Perrine's model was utilized for the development of a system of coding verbal responses on the CT. This model described a two-part process of concept formation in which Attribute Identification referred to identification of the relevant features of stimuli upon which the rule is based, whereas Rule Learning refers to the logical relationship between the attribute identified and the number response chosen. For example, on Subtest III, the Attribute Identification task is to discern that the stimulus which is different is the relevant dimension, while the Rule Learning task is to discover that the position of the different stimulus corresponds to the correct number.

However, during the course of this study, it was discovered that a three-stage coding system was required to reliably differentiate verbal responses:
(a) Concrete: Subjects selected a perceptual feature which they believed relevant;
(b) Abstract: The perceptual feature or features were organized conceptually to create a pattern; and (c) Rule: This pattern was related to the number response given. For example, in Subtest IV, the quadrants are numbered in a clockwise direction, and the principle requires identifying the quadrant of the stimulus which is missing or different. The Concrete (Attribute Identification) task is identification of the portion which is not there or which is different from the rest of the stimulus; the Abstract conceptual task is to perceive this attribute as one quadrant of the whole stimulus; the Rule Learning task is to select a number

corresponding to the position of the missing or different quadrant when ordered in a clockwise direction.

Often, however, responses referred to two or more perceptual attributes, or verbal responses were ambiguous as to which attribute took precedence. Consequently, in the interest of increased reliability in coding responses, a decision tree approach was adopted for coding verbal responses. This approach established a hierarchy of coding rules, ordered in such a way that the most relevant or highest order perceptual attributes and rules were listed first. (See Appendices E through G for the full model.) Use of this approach permitted achievement of a satisfactory level of test-retest (recoding) reliability (.94) for recoding verbal responses after an interval of four months (for one examiner). Achievement of this level of reliability permitted quantitative comparison of verbalized responses across subjects and across responses.

Factor Analysis of Cognitive Process Variables

Category Test

The CT has been described as a test of complex, problem-solving skills (Reitan & Wolfson, 1985) on the basis of previous factor analytic studies in which CT error scores loaded with other variables known to predict abstract reasoning, cognitive flexibility, and problem-solving skills (Barnes & Lucas, 1974; Corrigan & Hinkeldey, 1988; Goldstein & Shelly, 1972; Halstead, 1947a; Haynes & Sells, 1963; Royce et al., 1976; Russell, 1982). However, in only one of these studies (Royce et al., 1976) have CT subtest scores been entered into the analysis individually, and in none of these instances has it been possible to evaluate the nature of the errors committed by subjects in solving the task presented to them by the CT. The methodology developed for the purpose of the present study has afforded that opportunity by devising a standardized coding system based upon concept formation literature (Bourne et al., 1979) for describing how subjects relate the perceptual stimuli presented to them to the numbers which constitute their responses.

Since the focus of interest on this task is not how quickly the correct principle on each subtest was discovered, but rather on how effectively the feedback provided by the examiner (or the apparatus) was utilized, concept formation scores were transformed to change scores, or process scores, to identify how feedback was incorporated into subsequent responses. Thus, cognitive process variables were obtained which reflected the degree to which response determinants were repeated when correct feedback was received and response determinants were modified when incorrect feedback was received. These scores (see Appendix H) were tabulated for each CT subtest and for CT Total Errors.

It was hypothesized that factor analysis of these process scores would produce three or more interpretable cognitive process factors which were relatively independent of age, education, and IQ subtest and summary scores. Since preliminary analysis revealed that, except for Subtests V and VI which share a common principle, individual subtest error scores were relatively independent of error scores on other subtests (although some relationships were statistically significant), and that the small N precluded analyzing all data simultaneously, process scores for each subtest were analyzed separately. In order to enhance interpretability of results, the data analysis for the full group was replicated for impaired and unimpaired subgroups, using Reitan's definition for impairment on the CT (>50 errors).

Maintaining Set

These analyses revealed fairly consistent results across all subtests and for Total Errors. The factor in all subtests accounting for the largest proportion of the variance (28.1% for Subtest III to 38.1% for Subtest IV) was a Maintaining Set factor on which most variables that loaded highly related to learning the correct subtest principle and maintaining that principle on subsequent items. This was true also on all subgroup analyses, with the exception of the impaired subgroup on Subtest III. This factor accounted for the greatest portion of the variability in subtest error scores as well, ranging from 33% of the variance in error scores for Total Errors and Subtest V to 63% of the variance on Subtest VI (see Table 31 for details).

This factor also showed the greatest relationship with age, education, and IQ covariates across all subtests and for Total Errors. In most instances this relationship was nearly as great as the relationship between error scores and covariates (see Table 35). It may therefore be postulated that this factor accounts for most of the discriminability of the CT in differentiating brain-damaged from nonbrain-damaged subjects. In fact, this may well be a more reliable indicator of effective problem-solving skills than error scores since this factor score is comprised almost exclusively of variance due to repeating determinants which are correct and is not inflated by "incidentally correct" responses (Simmel & Counts, 1957).

Shifting Determinants

However, from a clinical perspective, despite the importance of assessing effective problem-solving skills, treatment must proceed from evaluation of how patients fail problem-solving tasks. The remaining factors which emerged from factor analysis of process variables across all subtests and for Total Errors predominantly related to how response determinants were changed when the previous response was incorrect. These factors were often ambiguous but were of three general types: (a) random or irrelevant shifting characterized by nonsystematic shifting of multiple determinants simultaneously; (b) systematic shifting of determinants in order to identify the relevant dimensions; and (c) cognitively rigid shifting characterized by persisting with a previous principle which was no longer effective or by difficulty abandoning assumptions.

Factor 2 for all analyses was a nonsystematic factor characterized by simultaneously shifting multiple determinants, often all three, although the degree to which each determinant loaded on this factor varied somewhat from subtest to subtest. Shifting just one determinant consistently loaded negatively on this factor, reflecting the absence of a systematic approach.

Factor 3 showed somewhat more variability across subtests. For Subtest III, this factor represented somewhat more effective and systematic problem-solving than the previous factor, with a moderate positive loading for Learning Rate, relating to slow but successful learning. This factor was not represented in Subtest IV, but Subtests V and VI were consistent with Subtest III, although the determinants shifted varied somewhat.

The remaining factors were less consistent across subtests, but related to some degree with cognitive rigidity and difficulty abandoning assumptions.

The relationship of these factors to CT performance was unclear, however, since almost all of the variance in subtest error scores was predicted by Maintaining Set factors. (See Table 31.) It was hypothesized that the restricted relationship of Determinant Shifting factors with subtest error scores may have been due to artifacts associated with oblique rotation, which apparently served to minimize the relationship between factors. Given the strong relationship between Maintaining Set factors and subtest error scores, this apparently served also to minimize the relationship between Determinant Shifting factors and subtest error scores. Consequently, Maintaining Set variables were factor analyzed separately from Determinant Shifting variables.

This strategy produced almost identical factors in each case (see Table 33), but multiple regression analysis revealed that the relationship of the Determinant Shifting factors to error scores was rather different. For the full model, even when Maintaining Set scores were forced to enter the equation last, following the stepwise block for factors related to incorrect responding, all of the variance in error score prediction was accounted for by Maintaining Set factors and none of the factors related to Determinant Shifting entered the equation, with probability to enter set at .05. However, when each set of scores was factor analyzed separately, the relationship of Determinant Shifting factors to subtest error scores was much more substantial.

Thus, the failure of Determinant Shifting factor scores to significantly predict error scores appeared to be artifactual rather than reflecting the absence of relationships. Nonetheless, the cognitive processes reflected by these factor scores remained ambiguous.

Second Order Factor Analysis

In view of the consistency of factor loadings across subtests and Total Errors, it appeared that these primary factors may be descriptive of a general approach to problem-solving tasks. Consequently, second order factor analysis was performed, entering all factors across subtests, to determine if primary factors of each type related to the same cognitive process. Results are shown in Table 27, with variables identified across the top of the table by number (from the definitions in Appendix H) and primary factors loading on each second order factor listed in the left hand column. It was expected that if primary factors related to cognitive processes which were consistently applied across problem-solving tasks, variable loadings on primary factors would cluster within each second order factor.

Seven second order factors emerged in this analysis, but although some clustering of variables is evident, only the first two second order factors have primary loadings from all subtests and variable clustering is not consistent. Second Order Factor 2 shows the greatest degree of variable clustering, with all maintaining set variables and primary factors from Subtests IV, V, and VI loading substantially. Subtest III showed a very different pattern, however. Although all subtests also loaded on Second Order Factor 1, the variable clustering pattern on this factor was very discrepant, with maintaining set variables from Subtests III and IV loading consistently (although comprising a relatively small proportion of the variance) but with unsystematic shifting variables from Subtests V and VI loading negatively. Thus, low scores on Unsystematic Shifting variables on Subtests IV, V, and VI appeared to predict high Maintaining Set variable scores on Subtests III and IV. The remaining second order factors are specific to primary factors from subsets of two subtests.

This analysis strongly suggested that despite the consistency with which similar factors emerged across subtests in the initial factor analysis of process variables, second order factor analysis indicated that these factors were primarily descriptive of variance unique to individual subtests or sets of similar subtests. Factor patterns were reminiscent of the relationships among subtests described in Table 29 where multiple regression analysis revealed that error scores on Subtest III were significantly predicted only by error scores on Subtest IV, and error scores on Subtest V were significantly predicted only by error scores on Subtest

VI, and vice versa. Only Subtest IV showed a significant relationship with more than one other subtest. As expected, error scores on Subtest IV were significantly predicted by Subtest III, but also by error scores on Subtest VI. However, even though these relationships were statistically significant ($r < .05$), they accounted for only a minor portion of the variance, except in the case of Subtests V and VI, which are based upon the same principle. It was expected that second order factor analysis of process variables would reveal stronger relationships among primary process factors across subtests, but this did not prove to be the case.

The failure to find second order factors which generalized across subtests may reflect characteristics of the cognitive process variables which were devised to measure hypothesized cognitive processes. Each of these scores was based upon determinant shifting following feedback from the examiner. However, perceptual differences between subtests may modify the significance of given determinant shifts. For example, in Subtest III, the most salient determinant may be R, while in Subtest IV the determinant which is most salient may be A. Consequently, the relevance of R shifts is likely to be very different across subtests. For that reason, intersubtest comparisons with respect to determinant shifting are not likely to be fruitful.

In contrast, the failure to find second order factors for cognitive processes which generalized across subtests may also reflect the uniqueness of task demands presented by each subtest (again except for Subtests V and VI). In fact, this explanation appears likely in view of the failure of Maintaining Set factors to show stronger relationships across subtests. These factors show less dependence upon perceptual dissimilarities between subtests in that they reflect a "win-stay" strategy, without regard to perceptual shifts.

Only two previous studies have addressed correlates of differential performance on individual CT subtests. Royce et al. (1976) also found that Subtests III and IV comprised a factor separate from Subtests V and VI. They referred to these as Halstead Abstraction I and Halstead Abstraction II, and found somewhat different patterns of correlations with sites of cerebral damage for each of these factors. Holland and Wadsworth (1976) also found evidence for a different pattern of correlates for individual CT subtests in a study which compared the performance of schizophrenic subjects with brain-damaged subjects on a battery of neuropsychological tests. They found that only CT Subtest IV minus Subtest V significantly differentiated the groups. However, the significance of these findings for cognitive correlates of CT subtest performance is not known.

Covariates of Cognitive Process Factors

One means of assessing correlates of subtest-specific cognitive process factors is to investigate the relationships of predicted age, education, and IQ covariates. These results are shown in Table 35. As might be expected in view of the restricted relationships among subtest errors scores and among cognitive process factors across subtests, relationships of age, education, and IQ covariates with both error scores and cognitive process factors were quite variable. These results suggested that scores across subtests are quite clearly measures of different cognitive processes.

More importantly, however, these findings suggested that most of the relationship of CT error scores with age, education, and IQ covariates was reflected in Maintaining Set factors, but that most Determinant Shifting factors were relatively independent of covariates. Previous research has found the CT to be highly sensitive to brain dysfunction, but specificity has been less reliable in discriminating nonbrain-damaged subjects due to the influence of covariates on CT performance. If relationships of covariates with Determinant Shifting factor scores are confirmed in cross-validation research , these scores may be more useful than error scores in discriminating nonbrain-damaged subjects.

Summary of CT Factor Analyses

It was anticipated that factor analysis of CT cognitive process variables would produce three or more cognitive process factors which were relatively independent of age, education, and IQ covariates. This expectation was fulfilled for all CT subtests in most respects. The requisite number of factors did emerge in all analyses and most were interpretable, although often ambiguous. However, the first factor to emerge for all subtests related to Maintaining Set variables, which predicted most of the variance in subtest error scores and showed almost as much covariation with age, education, and IQ variables as did the error scores themselves. Nonetheless, these scores may be more accurate reflectors of CT subtest performance than error scores because they are not influenced by "incidentally correct responses."

The remaining factors which emerged reflected Determinant Shifting variables and were quite consistent across subtests, showing much less relationship with age, education, and IQ covariates, although subtest specific factors did not predict similar factors on other subtests. Similarly, Maintaining Set factors from one subtest showed only minimal relationship with Maintaining Set factors from other subtests. These relationships were reminiscent of relationships between subtest error scores. It therefore appeared that CT subtest scores, whether error scores, Maintaining Set factor scores, or Determinant Shifting scores, were not substantially related, suggesting that CT subtests are measures of different cognitive processes. Previous research offers little guidance as to what the correlates of individual subtests may be, and the factors which emerged in this study do not shed much light on this question.

Factors which emerged in each subtest analysis were comprised of similar variables, but contrary to expectations the relationship between similar factors across subtests was minimal. Thus, second order factor analysis of primary cognitive process factors from each subtest did not reveal consistent relationships across subtests, and the pattern of covariation with age, education, and IQ variables across subtest factor scores was also quite inconsistent. It was, therefore, apparent that the cognitive process variables which were devised for this study did not reflect the cognitive skills required to perform each subtest task. For the cognitive shifting variables, this might be due to perceptual differences between subtest stimuli which served to modify the significance of given determinant shifts from subtest to subtest. That is, if a given determinant was of unequal salience across subtests, shifting that determinant in response to feedback would convey different interpretations from subtest to subtest.

The Maintaining Set factors are comprised of two components. First, scores on this factor reflect a "win-stay" strategy, in which determinant scores that were successful are repeated. Second, use of this strategy presupposes that the correct determinants have been surmised. The latter component is reflected in Set Learning Rate (SLR) scores, which are measures of how consistently determinant scores were repeated after the correct principle was learned. It might be expected that this strategy would generalize across problem-solving tasks, including different CT subtests. Discovery of the correct principle within each subtest is reflected in Learning Rate (LR) scores. If individual subtests reflect dissimilar cognitive skills, it would not be expected that LR scores would show substantial relationships across subtests.

However, multiple regression analysis revealed that both subtest specific SLR and LR scores did not generalize across subtests, except for Subtests V and VI (which are based upon the same principle), contrary to expectations. Inspection of the data suggested that the failure of SLR scores to generalize may have been due to learning failures which were not expected to generalize across subtests. Consequently, the data were recomputed after eliminating these scores. However, this adjustment did not substantially affect the generalizability of this score.

These results are very puzzling. Task differences across subtests are not relevant for this score, when learning failures are eliminated; consequently, SLR should reflect only failures of working memory, or attentional lapses, and this would not be expected to vary substantially as a result of task differences. Unreliability of scores might account for these findings, but in that case Subtests V and VI would not be expected to show discrepant patterns of relationship.

However, this is an impaired subgroup of predominantly head-injured subjects whose performance on learning tasks would be characterized by frequent proactive intrusions from previously learned material (Butters, 1992). The probability of such intrusions would vary with respect to strength of prior learning and task novelty. Thus, previous material which was overlearned, counting stimuli for instance, may intrude more frequently on tasks which are novel, such as Subtests III or IV, than on tasks which are more familiar, as in Subtests V and VI. This principle would predict that mean SLR for Subtests III and IV would be much smaller than mean SLR for Subtests V and VI. In fact, this pattern was partially confirmed. Mean SLR for Subtest III was .18 (S.D. = .04), while for Subtests V and VI mean SLR was .64 (S.D. = .22) and .64 (S.D. = .24), respectively. However, mean SLR for Subtest IV was greatest at .82 (S.D. = .31).

Task novelty for the latter subtest, however, may have been influenced by the six practice items on this subtest.

Further research concerning these patterns would be necessary to confirm this explanation, but certainly these findings point to the need for more systematic task analysis to define variables which reflect discriminable cognitive processes utilized by subgroups of subjects in concept formation tasks, particularly CT subtests. Fortunately, the determinant coding system devised for the purpose of this study makes such analyses possible.

Such an approach must be based on the premise that because the CT is a problem-solving test in which the central task is to surmise the correct principle for each subtest through use of feedback from the examiner (or the apparatus), not all responses are of equal salience. For example, incidentally correct responses are of particular value in helping to identify the correct principle. These are responses which are correct by chance rather than by design. These offer subjects the opportunity to contrast the perceptual attributes of the item which was incidentally correct with the perceptual attributes of previous items which were incorrect to identify differences, which must logically be descriptive of the correct principle. Consequently, determinant shifts following incidentally correct responses are particularly relevant in identifying cognitive approaches to this task.

Similarly, responses following an incorrect response within a series of "almost correct" responses are particularly salient in helping subjects discover the correct principle. These are responses in which the current operating principle is very similar to the correct principle, but it is ineffective in some instances. For example, in Subtests V and VI, many subjects count quadrants which are outlined rather than basing their responses on the proportion of the stimulus which is outlined. This results in correct responses for most but not all items. Thus, determinant changes following these disconfirming responses are especially relevant in identifying problem-solving characteristics of subjects.

Careful item analysis of determinant changes employing such a process approach to develop characteristic patterns of performance for various subgroups of subjects will be of much value in identifying cognitive processes descriptive of impaired performance on the CT.

Wisconsin Card Sorting Test

The WCST did not present the coding challenges which were present for the CT, since this test has been traditionally scored utilizing a process approach to analyze subjects' performance. The ambiguities for WCST conventional scores stem from ambiguous responses in which cards match on more than one perceptual attribute. In such cases, it is not possible to identify which feature determined card placement. As a result of such ambiguity, Osman (1992) noted that perseverative responses reflect a variety of processes, including (a) intrusions of a prior mental set; (b) continuation of on-going behavior; (c) loss of mental set; and (d) loss of conceptual power. Asking subjects to verbalize their reasoning was an attempt to better clarify process analysis by reducing the proportion of ambiguous responses without task simplification. It was expected that factor analysis of verbalized scores would produce three or more cognitive process factors which were relatively independent of age, education, and IQ covariates.

Comparison of Verbalized and
Nonverbalized Scores

Comparison of results of different scoring conditions revealed that when subjects verbalized their responses they often responded by utilizing multiple determinants. In such cases, verbalized responses conveyed no more information than did conventional scoring. As a result, the correlations between verbalized scores and conventional scores were generally quite high. Exceptions were for Right-Shift responses, Missed Learning Opportunities, and Percent Nonperseverative Errors. Inspection of the data revealed that the low correlation for Right-Shift responses resulted from restricted range of scores. Consequently, this score was dropped from further analysis. Further analysis revealed that for the remaining two variables, lower correlations between scoring conditions resulted from reduced ambiguity on verbalized scores. Consequently, these variables were retained for further analysis.

Separate factor analyses of both sets of retained scores revealed that the factor pattern for both sets of data was very similar. In both cases, Factor 1 reflected ineffective, perseverative responding and summary scores related to nonperseverative responding (Conceptual Level Responses, Categories Completed, and Learning to Learn) loaded negatively on this factor. Primary differences between the scoring conditions related to the variables which showed the lowest correlation between scoring conditions, Missed Learning Opportunities and Percent Nonperseverative Errors. Both changes were in the predicted direction, consistent with greater accuracy in scoring responses utilizing verbalized reasoning.

Factor 2 for both scoring conditions related principally to Unambiguous Number responses and Maximum Number responses, with lesser loadings for

Wrong-Stay, Missed Learning Opportunities, and Total Correct Percent, with a negative loading for Unambiguous Color. This pattern suggested that when the proportions of attribute responses (Color, Form, Number) were unequal, whether due to ambiguous or unambiguous responses, color and form responses were associated with perseveration, while unequal number responses were associated with nonperseverative responses. Thus, identifying the number attribute appeared to represent a different cognitive process than color or form attributes.

Factor 3 for both scoring conditions contributed the smallest proportion of the variance, and ironically was the only factor which related principally to more effective problem-solving strategies, with high negative loadings for variables related to maintaining set.

This pattern was very different from the factor pattern which emerged for the CT. In the latter case, effective problem-solving (Maintaining Set factor scores) accounted for the largest proportion of the variance, with less relationship for ineffective strategies. In contrast, on the WCST, ineffective strategies accounted for a much larger proportion of the variance, with less relationship for effective strategies. This distinction may be a function of the complexity of the tasks.

<u>Covariates for WCST Cognitive Process Factors</u>

Despite the similarity between factors for the two scoring conditions, relationships with age, education, and IQ covariates were quite different for Factor 1, although the remaining factors showed little relationship with covariates under both conditions. Consequently, the relationship of the original WCST variables with covariates was analyzed further, regressing covariates onto each variable for each scoring condition. Only 7 of 20 comparisons showed very similar patterns of covariation, and 3 other variables were somewhat similar. This proportion of dissimilar findings across scoring conditions is well beyond chance levels, and suggested that each set of scores may be measuring different cognitive processes. This view is strengthened by the differential relationship of Verbal IQ subtest scores versus Performance IQ subtest scores across scoring conditions. This pattern suggested that asking subjects to verbalize reasoning may have produced scores which reflected verbal skills rather than problem-solving skills. However, it is not known whether performance varied as a function of verbalizing reasoning, or whether only verbal scores varied in comparison with conventional scores.

Comparisons Between the CT and WCST

It was anticipated that factor scores derived from cognitive process variables on each of these tasks would predict a significantly greater proportion of the variance on the other task than using conventional scores alone. This prediction was tested utilizing multiple regression analysis, regressing all factor scores from one test onto each factor score from the other test, and partialing out variance due to age, education, and IQ covariates. However, the results of these analyses showed even less relationship between tasks than utilizing conventional scores. Consequently, no evidence was found that the tasks assessed by these tests are comparable.

It should be noted, however, that these analyses were dependent upon the success achieved in fractionating the cognitive processes on each task reflected by conventional scores. As noted previously, this assumption is probably not warranted in view of the ambiguity of factor and variable scores on CT Determinant Shifting. Osman (1992) has suggested that this may well also be true for WCST variable scores. Consequently, the relationship between cognitive processes reflected by CT and WCST performance remains an open question.

Conclusions

Assessment of concept formation and problem-solving skills is an essential part of any comprehensive neuropsychological evaluation since these high-level skills are among the functions most likely to be impaired following brain damage. The Halstead Category Test, or the equivalent form, the Booklet Category Test, and the Wisconsin Card Sorting Test have been found to be among the most sensitive instruments for identifying patients with impaired problem-solving abilities. However, for development of treatment plans for remediating these deficits, it is not sufficient to know that patients are impaired with respect to these skills; what is more important is to specify how patients are impaired. Heretofore, such information could be obtained only inferentially or anecdotally, since both CT and WCST performance have been conventionally assessed only by summary error scores, which evaluate only the end product of problem-solving. For the purposes of this study, a modified administration technique was devised which relied on patient verbalization of reasoning as they performed these tasks, which provided a means for observing how patients erred on these problem-solving tasks.

WCST Results

In the case of the WCST, although most such verbalized scores correlated highly with conventional scores, this technique permitted less ambiguous quantification of performance. However, patterns of age, education, and IQ covariates for each scoring condition were found to be very different, raising questions concerning what such verbalized scores measured. In particular, verbalized scores correlated more highly with WAIS-R verbal scale subtest scores, while conventional scores correlated more highly with WAIS-R performance scale subtest scores. These relationships may suggest that verbalized scores reflected ability to accurately describe behavior rather than reflecting dimensions of problem-solving abilities. However, since each set of scores was based upon the same objective performance (scored once conventionally and scored again based on patients' verbal descriptions of the reason for their responses), the present data are insufficient for determining how verbalization of reasoning may have affected performance.

Data were obtained from a comparison group to address this question, and patterns of covariation for primary scores on the WCST and CT were quite different between the comparison group and the experimental group, but the two groups were significantly different with respect to age ($p < .03$), and patterns of covariation were different between groups among WAIS-R subtest scores, for which no differences in administration conditions were present. Consequently, data from the comparison group were not useful for comparing performance characteristics of verbalized responses versus nonverbalized responses. Further research based upon a prospective design in which subjects are randomly assigned to administration conditions, and matched with respect to age, education, and IQ is needed to address this question.

Factor analysis of WCST scores for each scoring condition resulted in almost identical three-factor solutions in each case, differentiated primarily on the basis of the two scores with the lowest correlation between scoring conditions loading somewhat differently. This was in the predicted direction and appeared due to reduced ambiguity for these scores, Missed Learning Opportunities and Percent Nonperseverative Errors, in the verbalized condition. Factors which emerged for both scoring conditions were interpretable and related to (a) ineffective, perseverative responding; (b) nonperseverative number errors; and (c) Maintaining Set.

CT Results

Verbalized responses on the CT were much more difficult to quantify than was true for the WCST. Perrine (1984) suggested that the WCST is primarily a task of attribute identification, whereas the CT involves both attribute identification and rule learning. Perhaps for that reason, the CT represents a more complex task and reliable coding of verbal responses is therefore more multidimensional. Perrine's model of concept formation, derived from Bourne (1965), was utilized to describe a three-part hierarchy of response determinants, consisting of (a) concrete perceptual attributes; (b) cognitive organization of perceptual attributes into abstract patterns; and (c) relating abstract patterns to the corresponding number responses. Decision trees were devised to prescribe a set of rules for coding each score. Utilization of this approach yielded adequate test-retest reliability for recoding responses. Such an approach permitted quantification of the cognitive processes utilized by individual subjects as they performed this task.

For the purposes of quantifying how patients incorporated feedback from the examiner as to whether each response was correct or incorrect, concept formation scores were transformed to change scores, so that for each CT subtest patients' performances could be quantified with respect to the proportion of determinants which were repeated following "correct" feedback, proportion of each determinant which was changed following "incorrect" feedback, proportion of responses in which only one determinant was changed simultaneously versus those in which multiple determinants were changed simultaneously, as well as other more specific summary scores. Sets of variables for each subtest were then factor analyzed, with second order factor analysis of all factors from each subtest in order to determine if common cognitive process scores on each subtest described cognitive process scores on other subtests.

Results of these analyses revealed similar factor solutions for each subtest, but despite the apparent similarity of these factors across subtests, subtest specific factors were not predictive of similar factor scores on other subtests, except for Subtests V and VI, which are based upon the same principle. Similarly, second order factor analysis did not result in general factors across subtests; rather, the factors which emerged tended to be specific to, at most, sets of similar subtests.

For each subtest, Maintaining Set factors predicted most of the variance in subtest error scores, and these factor scores were influenced to approximately an equivalent degree by age, education, and IQ covariates as were error scores. Determinant Shifting factor scores were predictive of error scores to a much lesser degree than Maintaining Set factor scores. In fact, it was necessary to enter Determinant Shifting variables separately in order to discover any relationship with error scores at all. Thus, these change scores appeared to be independent of Maintaining Set scores, and also showed much more independence from age, education, and IQ covariates.

These results appear to suggest that fractionating error scores on the CT by utilizing a concept formation approach to cognitive processing is a promising approach for describing subsets of cognitive failures in a group of brain-injured patients. However, the cognitive process scores developed for the present study achieved only limited success in accomplishing this goal because the change scores utilized were too much dependent upon specific task characteristics and too multiply-determined. Further research utilizing careful task analysis for specific categories of responses may be more successful. For example, responses following "incidentally correct" responses (responses correct by chance) are especially salient in discerning how patients utilize feedback in modifying behavior. Responses which receive feedback of "incorrect" within a series of "almost conceptually correct" responses will be of value in assessing idea fluency and systematic organization of multiple data in problem-solving. These and other concept formation scores may perhaps be best derived through item analysis of CT responses, utilizing the concept formation approach devised for the purpose of this study.

Comparison of WCST and CT Factor Scores

Factor scores from the CT were expected to predict a greater proportion of the variance in WCST factor scores than the proportion of the variance in WCST summary scores predicted by CT error scores. However, this hypothesis was not supported. In fact, the relationship between CT and WCST factor scores was slightly lower than the relationship between CT error scores and WCST summary scores. However, this may have been due to the fact that CT factor scores tended to be task specific rather than generalizable across subtests. Since these scores were not predictive of factor scores across the much more similar CT subtests, it was not surprising that they were also not predictive of WCST factor scores.

Limitations of the Study and Suggestions for Further Research

While the results of this study were promising with respect to the development of a reliable coding system for comparing verbalized responses on the CT and WCST across subjects and across responses, and in providing greater clarification of the processes utilized by brain-injured patients in solving the problems represented by these tasks, this study must be considered exploratory. As with any study based upon archival data where subjects may not be randomly assigned to groups, these results may be confounded by subject selection procedures. Perhaps more importantly, recording of verbal responses without prior knowledge of scoring procedures may have resulted in ambiguous responses which would not be coded similarly by other scorers. Thus, although adequate test-retest (for recoding) reliability was achieved, interscorer reliability is unknown. Further, in view of the questionable comparibility of groups, it could not be established in this study to what degree verbalization of reasoning may have affected performance on these tasks.

Utilization of process scores of unknown reliability or validity also raises many questions concerning the significance of these results. Much further investigation of the correlates of these scores is needed. In particular, normative research is necessary to establish the relevance of these scores for various cognitive and demographic correlates, and item analyses of results will be needed to derive an appropriate means of measuring cognitive process variables.

Further, Osman (1992) suggested that WCST perseverative responses also reflect a variety of processes, including (a) intrusions of a prior mental set; (b) continuation of on-going behavior; (c) loss of mental set; and 4) loss of conceptual power. Consequently, he has also utilized a verbalized administration procedure for the WCST from which he has developed three additional scores to fractionate the cognitive processes reflected by perseverative responses on this task. Further research comparing the relationship between CT cognitive process scores derived from task analysis of specific categories of CT response items and fractionated WCST responses is necessary to address this question.

Such research has the potential to clarify the relationship between brain injury and executive and problem-solving functions, which is a prerequisite of effective remediation of impaired higher level skills in brain-injured patients.

REFERENCES

Aftanas, M. S., & Royce, J. R. (1969). A factor analysis of brain damage tests administered to normal subjects with factor score comparisons across ages. Multivariate Behavioral Research, 4, 459-479.

Allen, R. M. (1948). A note on the use of the Bellevue-Wechsler Scale Mental Deterioration Index with brain injured patients. Journal of Clinical Psychology, 4, 88-89.

Anderson, S. W., Damasio, H., Hones, R. D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. Journal of Clinical and Experimental Neuropsychology.

Barnes, G. W., & Lucas, G. J. (1974). Cerebral dysfunction vs. psychogenesis in Halstead-Reitan tests. Journal of Nervous and Mental Disease, 158, 50-60.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. Journal of General Psychology, 39, 15-22.

Bertram, K. W. (1984). Clarification of the role and configuration of learning as they are manifest in performance at Halstead's Category Test. (Doctoral dissertation, Michigan State University). Dissertation Abstracts International, 45, 1276B-1277B.

Blake, R. R., & McCarty, B. S. (1948). A comparative evaluation of the Bellevue-Wechsler Mental Deterioration Index distributions of Allen's brain injured patients and of normal subjects. Journal of Clinical Psychology, 4, 415-418.

Boll, T. J. (1981). The Halstead-Reitan Neuropsychological Battery. In S. B. Filskov & T. J. Boll (Eds.), Handbook of clinical neuropsychology (pp. 577-607). New York: John Wiley & Sons.

Bond, J. A., & Buchtel, H. A. (1984). Comparison of the Wisconsin Card Sorting Test and the Halstead Category Test. Journal of Clinical Psychology, 40, 1251-1254.

Bornstein, R. A. (1986). Contribution of various neuropsychological measures to detection of frontal lobe impairment. International Journal of Clinical Neuropsychology, VIII, 18-22.

Bourne, L. E., Jr. (1965). Hypotheses and hypothesis shifts in classification learning. Journal of General Psychology, 72, 251-261.

Bourne, L. E., Jr. (1974). An inference model for conceptual rule learning. In R. L. Solso (Ed.), Theories in cognitive psychology: The Loyola Symposium (pp. 231-256). Potomac, MD: Lawrence Erlbaum.

Bourne, L. E., Jr., Dominowski, R. L., & Loftus, E. F. (1979). Cognitive processes. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Boyle, G. J. (1988). What does the neuropsychological Category Test measure? Archives of Clinical Neuropsychology, 3, 69-76.

Brandon, A. D., & Chavez, E. L. (1985). Order and delay effects on neuropsychological test presentation: The Halstead Category and Wisconsin Card Sorting Tests. International Journal of Clinical Neuropsychology, VII, 152-153.

Brown, A. (1974). Examination of nypothesis-sampling theory. Psychological Bulletin, 11, 773-790.

Buss, A. H. (1953). Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. Journal of Experimental Psychology, 45, 75-81.

Buss, A. H. (1956). Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. Journal of Experimental Psychology, 52, 162-166.

Butters, N. (1992, November). Memory disorders associated with anmesia and dementia. Paper presented at the annual meeting of the National Academy of Neuropsychology.

Chapman, L. F., & Wolff, H. G. (1959). The cerebral hemispheres and the highest integrative functions of man. Archives of Neurology, 1, 357-424.

Chelune, G. J. (1983). Effects of partialing out p[ostmorbid WAIS scores in a heterogenous sample: Comment on Golden et al. Journal of Consulting and Clinical Psychology, 51, 932-933.

Cohen, J. (1957). The factorial structure of the WAIS between early adulthood and old age. Journal of Consulting Psychology, 21, 283-290.

Corrigan, J. D., Agresti, A. A., & Hinkeldey, N. S. (1987). Psychometric characteristics of the Category Test: Replication and extension. Journal of Clinical Psychology, 43, 368-376.

Corrigan, J. D., & Hinkeldey, N. S. (1988). Patterns of performance within the Halstead-Reitan Neuropsychological Test Battery. International Journal of Clinical Neuropsychology, X, 25-34.

Cullum, C. M., Steinman, D. R., & Bigler, E. D. (1984). Relationship between fluid and crystallized cognitive functions using Category Test and WAIS scores. International Journal of Clinical Neuropsychology, VI, 172-174.

DeFilippis, N. A., & McCampbell, E. (1979). Manual for the Booklet Category Test: Research and clinical form. Odessa, FL: Psychological Assessment Resources.

DeFilippis, N. A., McCampbell, E., & Rogers, P. (1979). Development of a booklet form of the Category Test: Normative and validity data. Neuropsychology, 11, 339-342.

De Renzi, E., Fagliori, P., Savoiardo, M., & Vignolo, L. A. (1966). The influence of aphasia and the hemispheric side of the cerebral lesion on abstract thinking. Cortex, 2, 399-420.

Diller, L. (1987). Neuropsychological rehabilitation. In M. Meier, A. Benton, & L. Diller (Eds.), Neuropsychological rehabilitation. New York: Guilford Press.

Doehring, D. G., & Reitan, R. M. (1961). Behavioral consequences of brain damage associated with homonymous visual field defects. Journal of Comparative and Physiological Psychology, 54, 489-492.

Doehring, D. G., & Reitan, R. M. (1962). Concept attainment of human adults with lateralized cerebral lesions. Perceptual and Motor Skills, 14, 27-33.

Doehring, D. G., Reitan, R. M., & Kløve, H. (1961). Changes in patterns of intelligence test performance associated with homonymous visual field defects. Journal of Nervous and Mental Disease, 132, 227-233.

Dominowski, R. L. (1973). Requiring hypotheses and the identification of unidimensional, conjunctive, and disjunctive concepts. Journal of Experimental Psychology, 100, 387-394.

Dominowski, R. L. (1974). How do people discover concepts? In R. L. Solso (Ed.), Theories in cognitive psychology: The Loyola Symposium (pp. 257-287). Potomac, MD: Lawrence Erlbaum.

Drewe, E. A. (1974). The effect of type and area of brain lesion on Wisconsin Card Sorting Test performance. Cortex, 10, 159-170.

Ernst, J. (1987). Neuropsychological problem-solving skills in the elderly. Psychology and Aging, 2, 363-365.

Fey, E. T. (1951). The performance of young schizophrenics and young normals on the Wisconsin Card Sorting Test. Journal of Consulting Psychology, 15, 311-319.

Filskov, S. D., Grimm, B. H., & Lewis, J. A. (1981). Brain-behavior relationships. In S. B. Filskov & T. J. Boll (Eds.), Handbook of clinical neuropsychology (pp. 39-73). New York: John Wiley & Sons.

Finlayson, M. A. J., Johnson, K. A., & Reitan, R. M. (1977). Relationship of level of education to neuropsychological measures in brain-damaged and nonbrain-damaged adults. Journal of Consulting and Clinical Psychology, 45, 538-542.

Fitzhugh, K. B., Fitzhugh, L. C., & Reitan, R. M. (1962). Wechsler-Bellevue comparisons in groups with "chronic" and "current" lateralized and diffuse brain lesions. Journal of Consulting Psychology, 26, 306-310.

Fitzhugh, K. B., Fitzhugh, L. C., & Reitan, R. M. (1964). Influence of age upon measures of problem solving and experiential background in subjects with longstanding cerebral dysfunction. Journal of Gerontology, 19, 132-134.

Fowler, P. C., Zillmer, E., & Newman, A. (1987). Abstract. A factor model of a modified Halstead neuropsychological test battery and its relationship to general level of cognitive functioning and neuropsychiatric status. Journal of Clinical and Experimental Neuropsychology, 9, 38.

Goldstein, F. C., & Levin, H. S. (1987). Disorders of reasoning and problem-solving ability. In M. Meier, A. Benton, & L. Diller (Eds.), Neuropsychological rehabilitation (pp. 327-354). New York: Guilford Press.

Goldstein, G. (1987). Neuropsychological assessment for rehabilitation: Fixed batteries, automated systems, and nonpsychometric methods. In M. Meier, A. Benton, & L. Diller (Eds.), Neuropsychological rehabilitation (pp. 18-40). New York: Guildford Press.

Goldstein, G., & Shelly, C. H. (1972). Statistical and normative studies of the halstead Neuropsychological Test Battery relevant to a neuropsychiatric hospital setting. Perceptual and Motor Skills, 34, 603-620.

Goldstein, G., & Shelly, C. H. (1973). Univeriate vs. multivariate analysis in neuropsychological test assessment of laterialized brain damage. Cortex, 9, 204-216.

Goldstein, G., & Shelly, C. H. (1984). Discriminative validity of various intelligence and neuropsychological tests. Journal of Consulting and Clinical Psychology, 52, 383-389.

Goldstein, K. (1940). Human nature. Cambridge, MA: Harvard University Press.

Goldstein, K., & Scheerer, M. (1941). Abstract and concrete behavior: An experimental study with special tests. Psychology Monographs, 53, (2, Whole No. 239).

Goldstein, S. G. (1969). Strategies of research in human neuropsychology. Journal of Learning Disabilities, 2, 403-406.

Gormezano, I., & Grant, D. A. (1958). Progressive ambiguity in the attainment of concepts on the Wisconsin Card sorting Test. Journal of Experimental Psychology, 55, 621-627.

Gorsuch, R. L. (1974). Factor analysis. Philadelphia: Saunders.

Grant, D. A. (1951). Perceptual versus analytic responses to the number concept of the Weigl-type card sorting test. Journal of Experimental Psychology, 41, 23-29.

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. Journal of Experimental Psychology, 38, 404-411.

Grant, D. A., & Cost, J. R. (1954). Continuities and discontinuities in conceptual behavior in a card sorting problem. Journal of General Psychology, 50, 237-244.

Grant, D. A., & Curran, J. F. (1952). Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. Journal of Experimental Psychology, 43, 408-413.

Grant, D. A., Jones, O. R., & Tallantis, B. (1949). The relative difficulty of the number, form, and color concepts of a Weigl-type problem. Journal of Experimental Psychology, 39, 552-557.

Guilmette, T. J., Faust, D., Hart, K., & Arkes, H. R. (1990). A national survey of psychologists who offer neuropsychological services. Archives of Clinical Neuropsychology, 5, 373-392.

Gutman, B. (1950). The application of the Wechsler-Bellevue scale in the diagnosis of organic brain disorders. Journal of Clinical Psychology, 6, 195-198.

Halstead, W. C. (1939). Behavioral effects of lesions of the frontal lobe in man. Archives of Neurology and Psychiatry, 42, 780-783.

Halstead, W. C. (1940). Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and nonequivalent stimuli. American Journal of Psychiatry, 96, 1263-1294.

Halstead, W. C. (1945). A power factor (P) in general intelligence: The effect of brain injuries. Journal of Psychology, 20, 57-64.

Halstead, W. C. (1947a). Brain and intelligence: A quantitative study of the frontal lobes. Chicago: Chicago University Press.

Halstead, W. C. (1947b). Specialization of behavioral functions and the frontal lobes. Journal of Nervous and Mental Disease, 27-59.

Halstead, W. C., & Settlage, P. H. (1943). Grouping behavior of normal persons and of persons with lesions of the brain. Archives of Neurology and Psychiatry, 49, 489-506.

Hanfmann, E., & Kasanin, J. (1936). A method for the study of concept formation. Journal of Psychology, 3, 521-540.

Haynes, J. R., & Sells, S. B. (1963). Assessment of organic brain damage by psychological tests. Psychological Bulletin, 60, 316-325.

Heaton, R. K. (1981). Wisconsin Card Sorting Test manual. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Grant, I., & Matthews, C. G. (1986). Differences in neuropsychological tests performance associated with age, education, and sex. In I. Grant & K. M. Adams (Eds.), Neuropsychological assessment of neuropsychiatric disorders (pp. 100-120). New York: Oxford University Press.

Heidbreder, E. (1948). The attainment of concepts: VI. Exploratory experiments on conceptualization at perceptual levels. Journal of Psychology, 26, 193-216.

Heidbreder, E., & Overstreet, P. (1948). The attainment of concepts: V. Critical features and contexts. Journal of Psychology, 26, 45-69.

Heilbrun, A. B. (1962). Issues in the assessment of organic brain damage. Psychological Reports, 10, 511-515.

Heinrichs, R. W. (1990). Variables associated with Wisconsin Card Sorting Test performance in neuropsychiatric patients referred for assessment. Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 3, 107-112.

Hevern, V. W. (1980). Recent validity studies of the Halstead-Reitan approach to clinical neuropsychological assessment: A critical review. Clinical Neuropsychology, 2, 49-61.

Hewson, L. R. (1949). The Wechsler Belevue Scale and the Substitution Test as aids in neuro-psychiatric diagnosis. Journal of Nervous and Mental Disease, 109, 158-193, 246-266.

Holland, T. R., & Wadsworth, H. M. (1976). Assessment of conceptual deficits in brain-damaged and schizophrenic patients. Perceptual and Motor Skills, 43, 951-957.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. Journal of Educational Psychology, 57, 253-270.

Kazdin, A. E. (1982). Single-case research designs. New York: Oxford University Press.

Kendler, H. H., Glucksberg, S., & Keston, R. (1961). Perception and mediation in concept learning. Journal of Experimental Psychology, 61, 186-191.

King, M. C., & Snow, W. G. (1981). Problem-solving task performance in brain-damaged subjects. Journal of Clinical Psychology, 37, 400-404.

Kløve, H. (1959a). Relationship of differential electroencephalographic patterns to distribution of Wechsler-Bellevue scores. Neurology, 9, 871-876.

Kløve, H. (1959b, May). The relationship of sensory suppression to distribution of Wechsler-Bellevue scores. Paper presented at Midwest Psychological Association Meeting.

Kløve, H. (1974). Validation studies in adult clinical neuropsychology: Current status and applications. In R. M. Reitan & L. A. Davison (Eds.), Clinical neuropsychology: Current status and applications. Washington, DC: V. H. Winston and Sons.

Kløve, H., & Reitan, R. M. (1958). Effect of dysphasia and spatial distortion on Wechsler-Bellevue results. Archives of Neurology and Psychiatry, 80, 708-713.

Lansdell, H., & Donnelly, E. F. (1977). Factor analysis of the Wechsler Adult Intelligence Scale subtests and the Halstead-Reitan Category and Tapping Tests. Journal of Consulting and Clinical Psychology, 45, 412-416.

Leckliter, I. N., Matarazzo, J. D., & Silverstein, A. B. (1986). A literature review of factor analystic studies of the WAIS-R. Journal of Clinical Psychology, 42, 332-342.

Leonberger, F. T., Nicks, S. D., Larrabee, G. J., & Goldfader, P. (1992). Factor structure of the Wechsler Memory Scale - Revised within a comprehensive neuropsychological battery. Neuropsychology, 6, 239-249.

Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 71, 331-338.

Lezak, M. D. (1983). Neuropsychological assessment (2nd ed.). New York: Oxford University Press.

Lin, Y. G., & Rennick, P. M. (1974). Correlations between performance on the Category Test and the Wechsler Adult Intelligence Scale; in an epileptic sample. Journal of Clinical Psychology, 30, 62-65.

Logue, P. E., & Allen, K. (1971). WAIS-predicted Category Test scores with the Halstead Neuropsychological Battery. Perceptual and Motor Skills, 33, 1095-1096.

Loranger, A. W., & Misiak, H. (1960). The performance of aged females on five nonlanguage tests of intellectual functions. Journal of Clinical Psychology, 16, 189-191.

Mack, J. L., & Carlson, N. J. (1978). Conceptual deficits and aging: The Category Test. Perceptual and Motor Skills, 46, 123-128.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of adult intelligence (5th ed.). New York: Oxford University Press.

Matarazzo, J. D., Matarazzo, R. G., Wiens, A. N., Gallo, A. E., Jr., & 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, J. D., Wiens, A. N., Matarazzo, R. G., & Goldstein, S. G. (1974). Psychometric and clinical test-retest reliability of the Halstead Impairment Index in a sample of healthy, young, normal men. Journal of Nervous and Mental Disease, 158, 37-49.

Matarazzo, R. G., Matarazzo, J. D., Gallo, A. E., Jr., & Wiens, A. N. (1979). IQ and neuropsychological changes following carotid endarterectomy. Journal of Clinical Neuropsychology, 1, 97-116.

Matthews, C. G., & Booker, H. F. (1972). Pneumoencephalographic measurements and neuropsychological test performance in human adults. Cortex, 8, 69-92.

Matthews, C. G., & Reitan, R. M. (1964). Correlations of Wechsler-Bellevue rank orders of subtest means in lateralized and nonlateralized brain-damaged groups. Perceptual and Motor Skills, 19, 391-399.

Matthews, C. G., Shaw, D. J., & Kløve, H. (1966). Psychological test performance in neurologic and "pseudoneurologic" subjects. Cortex, 2, 244-253.

Matthews, L. J., & Patton, J. H. (1975). Failure to shift following disconfirmation in concept identification. Journal of Experimental Psychology: Human Learning and Memory, 104, 91-94.

McFie, J., & Piercy, M. F. (1952a). Relations of laterality of lesions to performance on Weigl's sorting test. Journal of Mental Science, 98, 299-305.

McFie, J., & Piercy, M. F. (1952b). Intellectual impairment with localized cerebral lesions. Brain, 75, 292-311.

McKenna, P., & Warrington, E. K. (1986). The analytical approach to neuropsychological assessment. In I. Grant & K. M. Adams (Eds.), Neuropsychological assessment of neuropsychiatric disorders, (pp. 31-47). New York: Oxford University Press.

Meyerink, L. H. (1982). Intellectual functioning: The nature and pattern of change with aging. (Doctoral dissertation, University of Arizona). Dissertation Abstracts International, 43, 855B.

Milberg, W. P., Hebren, N., & Kaplan, E. (1986). The Boston Process Approach to neuropsychological assessment. In I. Grant & K. M. Adams (Eds.), Neuropsychological assessment of neuropsychiatric disorders (pp. 65-99). New York: Oxford University Press.

Milner, B. (1963). Effects of different brain lesions on card sorting. Archives of Neurology, 9, 90-100.

Milner, B. (1971). Interhemispheric differences in the localization of psychological process in man. British Medical Bulletin, 27, 272-277.

Mishkin, M. (1957). Effects of small frontal lesions on delayed alternation in monkeys. Journal of Neurophysiology, 20, 615-622.

Misiak, H., & Loranger, A. W. (1961). Cerebral dysfunction and intellectual impairment in old age. Science, 134, 1518-1519.

Moses, J. A., Jr. (1985). The relative contributions of Halstead-Reitan Neuropsychological Battery and WAIS subtest variables to cognitive performance level. International Journal of Clinical Neuropsychology, VII, 117-122.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. Cortex, 12, 313-324.

Osman, D. C. (1992, November). Modified Wisconsin Card Sorting Test with scores to fractionate executive processes. A paper presented to the annual meeting of the National Academy of Neuropsychology.

Parsons, O. A., Jones, B., & Vega, A. (1971). Halstead's Category Test and lateralized brain damage. Perceptual and Motor Skills, 33, 1245-1246.

Parsons, O. A., & Prigatano, G. P. (1978). Methodological considerations in clinical neuropsychological research. Journal of Consulting and Clinical Psychology, 46, 608-619.

Pendleton, M. G., & Heaton, R. K. (1982). A comparison of the Wisconsin Card Sorting Test and the Category Test. Journal of Clinical Psychology, 38, 392-396.

Perrine, K. R. (1984). Concept formation in the Halstead Category Test and Wisconsin Card Sorting Test. (Doctoral dissertation, University of Kentucky, 1983). Dissertation Abstracts International, 44, 2904B.

Peterson, M. J., & Colavita, F. B. (1964). Strategies, type of solution and stage of learning. Journal of Experimental Psychology, 68, 578-587.

Pishkin, V., & Williams, W. V. (1977). Cognitive rigidity in information processing of chronic undifferentiated schizophrenics. Journal of Clinical Psychology, 33, 625-630.

Pribram, K. H., Mishkin, M., Rosvold, H. E., & Kaplan, S. J. (1952). Effects on delayed-response performance of lesions of dorsolateral and ventromedial frontal cortex of baboons. Journal of Comparative Physiology and Psychology, 45, 465-575.

Prigatano, G. P., & Parsons, O. A. (1976). Relationship of age and education to halstead Test Performance in different patient populations. Journal of Consulting and Clinical Psychology, 44, 527-533.

Rattan, G. (1986). The role of response time in a complex learning task of the Halstead-Reitan Neuropsychological Test Battery. (Doctoral dissertation, Ball State University, 1985). Dissertation Abstracts International, 47, 2143B.

Reed, H. B. C., & Reitan, R. M. (1963a). A comparison of the effects of the normal aging process with the effects of organic brain-damage on adaptive abilities. Journal of Gerontology, 18, 177-179.

Reed, H. B. C., & Reitan, R. M. (1963b). Changes in psychological test performance associated with the normal aging process. Journal of Gerontology, 18, 271-274.

Reitan, R. M. (1955a). Investigation of the validity of Halstead's measures of biological intelligence. AMA Archives of Neurology and Psychiatry, 73, 28-35.

Reitan, R. M. (1955b). The distribution according to age of a psychologic measure dependent upon organic brain functions. Journal of Gerontology, 10, 338-340.

Reitan, R. M. (1955c). Certain differential effects of left and right cerebral lesions in human adults. Journal of Comparative and Physiological Psychology, 48 474-477.

Reitan, R. M. (1956). Investigation of relationships between "psychometric" and "biological" intelligence. Journal of Nervous and Mental Disease, 123, 536-541.

Reitan, R. M. (1958). Qualitative versus quantitative mental changes following brain damage. Journal of Psychology, 46, 339-346.

Reitan, R. M. (1959a). Impairment of abstraction ability in brain damage: Quantitative versus qualitative changes. Journal of Psychology, 48, 97-102.

Reitan, R. M. (1959b). The comparative effects of brain damage on the Halstead Impairment Index and the Wechsler-Bellevue Scale. Journal of Clinical Psychology, 15, 281-285.

Reitan, R. M. (1960). The significance of dysphasia for intelligence and adaptive abilities. Journal of Psychology, 50, 355-376.

Reitan, R. M. (1966a). Problems and prospects in studying the psychological correlates of brain lesions. Cortex, 2, 127-154.

Reitan, R. M. (1966b). A research program on the psychological effects of brain lesions in human beings. In N. R. Ellis (Ed.), International review of research in mental retardation (Vol. 1, pp. 53-218). New York: Academic Press.

Reitan, R. M. (1967). Psychological assessment of deficits associated with brain lesions in subjects with normal and subnormal intelligence. In J. L. Khanna (Ed.), Brain damage and mental retardation: A psychological evaluation (pp. 44-87). Springfield, IL: Charles C. Thomas.

Reitan, R. M. (1979). Manual of administration of neuropsychological test batteries for children and adults. Tucson, AZ: Neuropsychology Laboratory.

Reitan, R. M. (1986). Theoretical and methodological bases of the Halstead-Reitan Neuropsychological Test Battery. In I. Grant & K. M. Adams (Eds.), Neuropsychological assessment of neuropsychiatric disorders (pp. 3-30). New York: Oxford University Press.

Reitan, R. M., & Shipley, R. E. (1963). The relationship of serum cholesterol changes on psychological abilities. Journal of Gerontology, 18, 350-356.

Reitan, R. M., & Wolfson, D. (1985). The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation. Tucson, AZ: Neuropsychology Press.

Restle, F. (1962). The selection of strategies in cue learning. Psychological Review, 69, 329-343.

Reynell, W. R. (1944). A psychometric method of determining intellectual loss following head injury. Journal of Mental Science, 90, 710-719.

Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. Journal of Consulting and Clinical Psychology, 48, 605-614.

Rogers, L. S. (1950a). A comparative evaluation of the Wechsler-Bellevue Mental Deterioration Index for various adult groups. Journal of Clinical Psychology, 6, 199-202.

Rogers, L. S. (1950b). A note on Allen's Index of Deterioration. Journal of Clinical Psychology, 6, 203.

Rothke, S. (1986). The role of set shifting cues on the Wisconsin Card Sorting Test and Halstead Category Test. International Journal of Clinical and Neuropsychology, VIII, 11-14.

Royce, J. R., Yeudall, L. T., & Bock, C. (1976). Factor analytic studies of human brain damage: I. First and second-order factors and their brain correlates. Multivariate Behavioral Research, 11, 318-418.

Russell, E. W. (1974). The effect of acute lateralized brain damage on Halstead's biological intelligence factors. Journal of General Psychology, 90, 101-107.

Russell, E. W. (1979). Three patterns of brain damage on the WAIS. Journal of Clinical Psychology, 35, 611-620.

Russell, E. W. (1982). Factor analysis of the Revised Wechsler Memory Scale tests in the neuropsychological battery. Perceptual and Motor Skills, 54, 971-974.

Russell, E. W., Neuringer, C., & Goldstein, G. (1970). Assessment of brain damage: A neuropsychological key approach. New York: Wiley-Interscience.

Schreiber, D. J., Goldman, H., Kleinman, K. M., Goldfader, P. R., & Snow, M. Y. (1976). The relationship between independent neuropsychological and neurological detection and localization of cerebral impairment. Journal of Nervous and Mental Disease, 162, 360-365.

Shaw, D. J. (1966). The reliability and validity of the Halstead Category Test. Journal of Clinical Psychology, 22, 176-179.

Shore, C., Shore, H., & Pihl, R. O. (1971). Correlations between performance on the Category Test and the Wechsler Adult Intelligence Scale. Perceptual and Motor Skills, 32, 70.

Shure, G. H., & Halstead, W. C. (1958). Cerebral localization of intellectual processes. Psychological Monographs, 72 (12, Whole No. 465).

Simmel, M. L., & Counts, S. (1957). Some stable response determinants of perception, thinking, and learning: A study based on the analysis of a single test. Genetic Psychology Monographs, 56, 3-157.

Sohlberg, M. M., & Mateer, C. A. (1989). Introduction to cognitive rehabilitation. New York: Guilford Press.

SPSS, Inc. (1990). SPSS for the Macintosh: Operations Guide. Chicago: SPSS, Inc.

Strang, M. A. (1983). An examination of the Wisconsin Card Sorting Test and a comparison with the Halstead Category Test. (Doctoral dissertation, Wayne State University, 1983). Dissertation Abstracts International, 44, 929B.

Tarter, R. E., & Parsons, O. A. (1971). Conceptual shifting in chronic alcoholics. Journal of Abnormal Psychology, 77, 71-75.

Teuber, H. L., Battersby, W. S., & Bender, M. B. (1951). Performance of complex visual tasks after cerebral lesions. Journal of Nervous and Mental Disease, 114, 413-429.

Trabasso, T., & Bower, G. (1966). Presolution reversal and dimension shifts in concept identification. Journal of Experimental Psychology, 67, 398-399.

Vega, A., Jr., & Parsons, O. A. (1967). Cross-validation of the Halstead-Reitan Tests for brain damage. Journal of Consulting Psychology, 31, 619-625.

Walsh, K. W. (1987). Neuropsychology: A clinical approach (2nd ed.). New York: Churchill Livingstone.

Wechsler, D. (1944). Measurement of adult intelligence. Baltimore: Williams and Williams.

Weigl, E. (1941). On the psychology of the so-called process of abstraction. Journal of Abnormal and Social Psychology, 36, 3-33.

Weinberger, D. R., Berman, K. F., & Zec, R. F. (1986). Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia. Archives of General Psychiatry, 43, 114-124.

Wetherick, N. E., & Dominowski, R. L. (1976). How representative are concept attainment experiments? British Journal of Psychology, 67, 231-242.

Wetzel, L. C. (1983). Development of a short Booklet Category Test: Correlational and validity data. (Doctoral dissertation, Chicago Medical School, 1982). Dissertation Abstracts International, 43, 3381B.

Wiens, A. N., & Matarazzo, J. D. (1977). WAIS and MMPI correlates of the Halstead-Reitan Neuropsychology Battery in normal male subjects. Journal of Nervous and Mental Disease, 164, 112-121.

Winkleman, W. M. (1982). Neuropsychological indicators of right hemisphere cerebral damage in a medical rehabilitation population. (Doctoral dissertation, Southern Illinois University, 1981). Dissertation Abstract International, 43, 1634B.

Zillmer, E. A., Fowler, P. C., Newman, A. C., & Archer, R. P. (1986). Relationship between the WAIS and neuropsychological measures for neuropsychiatric inpatients. Journal of Clinical and Experimental Neuropsychology, 8, 34.

Zimmerman, S. F., Whitmyre, J. W., & Fields, F. R. (1970). Factor analytic structure of the Wechsler Adult Intelligence Scale in patients with diffuse and lateralized cerebral dysfunction. Journal of Clinical Psychology, 26, 462-465.

APPENDICES

APPENDIX A. Instructions For Modified Wisconsin Card Sorting Test

INSTRUCTIONS FOR MODIFIED WISCONSIN CARD SORTING TEST

Modification to standard instructions are in bold print. Allow patients to respond verbally to each response prior to providing feedback. Mark verbalized responses to the right of the "/" on the answer sheet, and actual responses to the left of the "/". If patients do not spontaneously provide verbal responses, say "because . . . " or "and you placed it there because . . . ". If you do not understand the reasoning, ask the patient to repeat, but do not inquire further. If the patient verbalizes a response which is inconsistent with card placement, repeat the response, and if the patient spontaneously self-corrects, record the correction; otherwise, record the original response.

THIS TEST IS A LITTLE UNUSUAL, BECAUSE I AM NOT ALLOWED TO TELL YOU VERY MUCH ABOUT HOW TO DO IT. YOU WILL BE ASKED TO MATCH EACH OF THE CARDS IN THESE DECKS TO ONE OF THE FOUR KEY CARDS, **AND I WILL TELL YOU WHETHER YOU ARE RIGHT OR WRONG.** YOU MUST ALWAYS TAKE THE TOP CARD FROM THE DECK, AND PLACE IT BELOW THE KEY CARD YOU THINK IT MATCHES, **AND ALSO SAY THE REASON WHY YOU PLACED THE CARD WHERE YOU PLACED IT--THE RULE YOU USED WHEN YOU PLACED IT IN THAT PILE.** I CAN'T TELL YOU HOW TO MATCH THE CARDS, BUT I WILL TELL YOU EACH TIME WHETHER YOU **PLACED IT IN THE RIGHT OR WRONG PILE, BUT REMEMBER WHEN I SAY RIGHT OR WRONG THAT APPLIES TO WHERE YOU PLACED THE CARD, NOT TO THE REASON YOU HAVE GIVEN.** IF YOU ARE WRONG, LEAVE THE CARD WHERE YOU'VE PLACED IT, AND TRY TO GET THE NEXT CARD CORRECT. USE THIS DECK FIRST, AND THEN CONTINUE WITH THE SECOND DECK. THERE IS NO TIME LIMIT ON THIS TEST.

APPENDIX B. Instructions For Modified Booklet Category Test

INSTRUCTIONS FOR MODIFIED BOOKLET
CATEGORY TEST

Modifications to standard instructions are in bold print. Allow patients to respond verbally to each response prior to providing feedback. Mark verbalized responses verbatim in the space to the right of each response on the answer sheet. (Do not inquire Subtests I and II). If patients do not spontaneously provide verbal responses, say "because . . . ". If you do not understand the reasoning, ask the patient to repeat, but do not inquire further. Inquire all responses, unless the patient has correctly identified the correct principle for that subtest three consecutive times. In that case, it is not necessary to continue inquiring, unless the patient makes a mistake, or the item is marked with a "?". If the patient spontaneously continues to verbalize reasoning, after three consecutive correctly verbalized responses, those responses should be recorded. Also all items marked with a "?" should be recorded, even when the patient has correctly identified the principle three consecutive times.

IN THIS BOOKLET YOU ARE GOING TO SEE DIFFERENT FIGURES AND DESIGNS. SOMETHING ABOUT THE PATTERN ON A PAGE WILL REMIND YOU OF A NUMBER BETWEEN ONE AND FOUR. ON THE STRIP IN FRONT OF YOU (pointing), YOU WILL SEE THE NUMBERS ONE, TWO, THREE, AND FOUR. FIRST LOOK AT THE PAGE AND DECIDE WHICH NUMBER THE PICTURE SUGGESTS, THEN POINT TO THAT NUMBER ON THE STRIP. FOR EXAMPLE, WHAT NUMBER DOES THIS REMIND YOU OF? (turn to the first page.) If the subject says, "one," ask him which number he should point to. After he has pointed to the number 1, say: CORRECT. THAT IS HOW I WILL RESPOND EVERY TIME YOU HAVE THE RIGHT ANSWER. I WILL RESPOND WITH "INCORRECT" WHEN YOU HAVE THE WRONG ANSWER. IN THIS WAY, YOU WILL KNOW EACH TIME WHETHER YOU HAVE THE RIGHT OR WRONG ANSWER. HOWEVER, FOR EACH PICTURE ON THE PAGE YOU ONLY GET ONE CHOICE. IF YOU MAKE A MISTAKE, WE JUST GO RIGHT ON TO THE NEXT PICTURE.

I. (Proceed with Subtest I) NOW, WHICH NUMBER WOULD YOU CHOOSE FOR THIS PICTURE?

II. After subtest I, say: THAT WAS THE END OF THE FIRST SUBTEST. THIS TEST IS DIVIDED INTO SEVEN SUBTESTS. IN EACH SUBTEST, THERE IS ONE IDEA OR PRINCIPLE THAT RUNS THROUGHOUT THE SUBTEST. ONCE YOU HAVE FIGURED OUT THE IDEA OR PRINCIPLE IN THE SUBTESTS, BY USING THIS IDEA YOU WILL GET THE RIGHT ANSWER EACH TIME. NOW WE ARE GOING TO BEGIN THE SECOND SUBTEST AND THE IDEA IN IT MAY BE THE SAME AS THE LAST ONE OR IT MAY BE DIFFERENT. WE WANT YOU TO FIGURE IT OUT. (Proceed with Subtest II.)

When you reach the first page with circles, say: YOU WILL NOTICE THAT WE FIRST SAW SQUARES, THEN LINES, AND NOW CIRCLES. EVEN THOUGH THE PATTERNS CHANGE, YOU SHOULD CONTINUE TO USE THE SAME IDEA TO GET THE RIGHT ANSWER.

III. After Subtest II, say: THAT WAS THE END OF THE SECOND SUBTEST AND AS YOU PROBABLY NOTICED, YOU DON'T NECESSARILY HAVE TO SEE A NUMBER TO HAVE A NUMBER SUGGESTED TO YOU. YOU SAW SQUARES, CIRCLES, AND OTHER FIGURES. ALSO, YOU PROBABLY NOTICED IN EACH OF THESE SUBTEST, THERE WAS ONLY ONE IDEA OR PRINCIPLE WHICH RAN THROUGHOUT. ONCE YOU FIGURED OUT THE IDEA, YOU CONTINUED TO APPLY IT TO GET THE RIGHT ANSWER. NOW WE ARE GOING TO START THE THIRD SUBTEST AND THE IDEA IN IT MAY BE THE SAME AS THE LAST ONE OR IT MAY BE DIFFERENT. I WANT TO SEE IF YOU CAN FIGURE OUT WHAT THE IDEA IS AND THEN USE IT TO GET THE RIGHT ANSWER, **AND I'D ALSO LIKE YOU TO SAY THE REASON YOU CHOSE THE ANSWER YOU CHOSE--WHY YOU PICKED THAT NUMBER, BUT AGAIN WHEN I SAY "RIGHT" OR "WRONG," THAT APPLIES TO THE NUMBER YOU HAVE CHOSEN, NOT TO THE REASON YOU HAVE GIVEN.**

REMEMBER, THE IDEA REMAINS THE SAME THROUGHOUT THE SUBTEST. I WILL TELL YOU WHEN WE COMPLETE ONE SUBTEST AND ARE READY TO BEGIN A NEW ONE.

IV. After Subtest III-V, say: THAT WAS THE END OF THAT SUBTEST. NOW WE ARE GOING TO BEGIN THE NEXT ONE. THE IDEA IN IT MAY BE THE SAME AS THE LAST ONE OR IT MAY BE DIFFERENT. WE WANT YOU TO FIGURE IT OUT.

VI. In Subtest VI after page #6 (first slide without numbers) say: THIS IS STILL THE SAME GROUP, BUT NOW THE NUMBERS ARE MISSING. THE PRINCIPLE IS STILL THE SAME.

VII. After Subtest VI, say: IN THE LAST SUBTEST THERE IS NO ONE IDEA OR PRINCIPLE THAT RUNS THROUGHOUT THE GROUP BECAUSE IT IS MADE UP OF ITEMS YOU HAVE ALREADY SEEN IN PRECEDING SUBTESTS. TRY TO REMEMBER WHAT THE RIGHT ANSWER WAS THE LAST TIME YOU SAW THE PATTERN AND GIVE THAT SAME ANSWER AGAIN.

APPENDIX C. Modified Wisconsin Card Sorting Test Form

WISCONSIN CARD SORTING TEST FORM

ID No ____________________ Age ________ Sex ________ Date ________

Education ____________________ Occupation ____________________

Comments ____

____

Trials C, F, N, C, F, N

__	1.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	33.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	2.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	34.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	3.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	35.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	4.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	36.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	5.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	37.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	6.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	38.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	7.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	39.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	8.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	40.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	9.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	41.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	10.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	42.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	11.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	43.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	12.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	44.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	13.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	45.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	14.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	46.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	15.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	47.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	16.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	48.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	17.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	49.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	18.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	50.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	19.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	51.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	20.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	52.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	21.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	53.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	22.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	54.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	23.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	55.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	24.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	56.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	25.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	57.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	26.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	58.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	27.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	59.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	28.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	60.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	29.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	61.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	30.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	62.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	31.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	63.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)
__	32.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)	__	64.	C F N O / C F N Seq Sum Pos Prev Des Ran Err Not (_)

Current Category _____ Number in Current Run _____

Trials C, F, N, C, F, N

__ 65. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 97. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 66. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 98. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 67. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 99. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 68. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 100. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 69. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 101. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 70. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 102. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 71. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 103. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 72. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 104. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 73. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 105. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 74. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 106. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 75. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 107. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 76. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 108. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 77. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 109. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 78. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 110. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 79. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 111. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 80. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 112. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 81. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 113. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 82. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 114. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 83. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 115. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 84. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 116. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 85. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 117. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 86. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 118. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 87. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 119. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 88. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 120. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 89. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 121. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 90. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 122. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 91. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 123. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 92. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 124. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 93. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 125. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 94. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 126. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 95. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 127. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)
__ 96. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)	__ 128. CFNO/CFN Seq Sum Pos Prev Des Ran Err Not (_)

Errors ________ Perseverative Responses _________ Categories ______ Learning to Learn ___________

Nonperseverative Errors ________ Perseverative Errors__________ % Perseverative Errors ____________

Trials to Complete 1st Category ______ Failure to Maintain Set ______ % Conceptual Level Response ____

If Right, Shift %_______ If Right, Stay %_________ If Wrong, Shift %_________ If Wrong, Stay %________

APPENDIX D. Modified Scoring And Recording Form For The Booklet Category Test

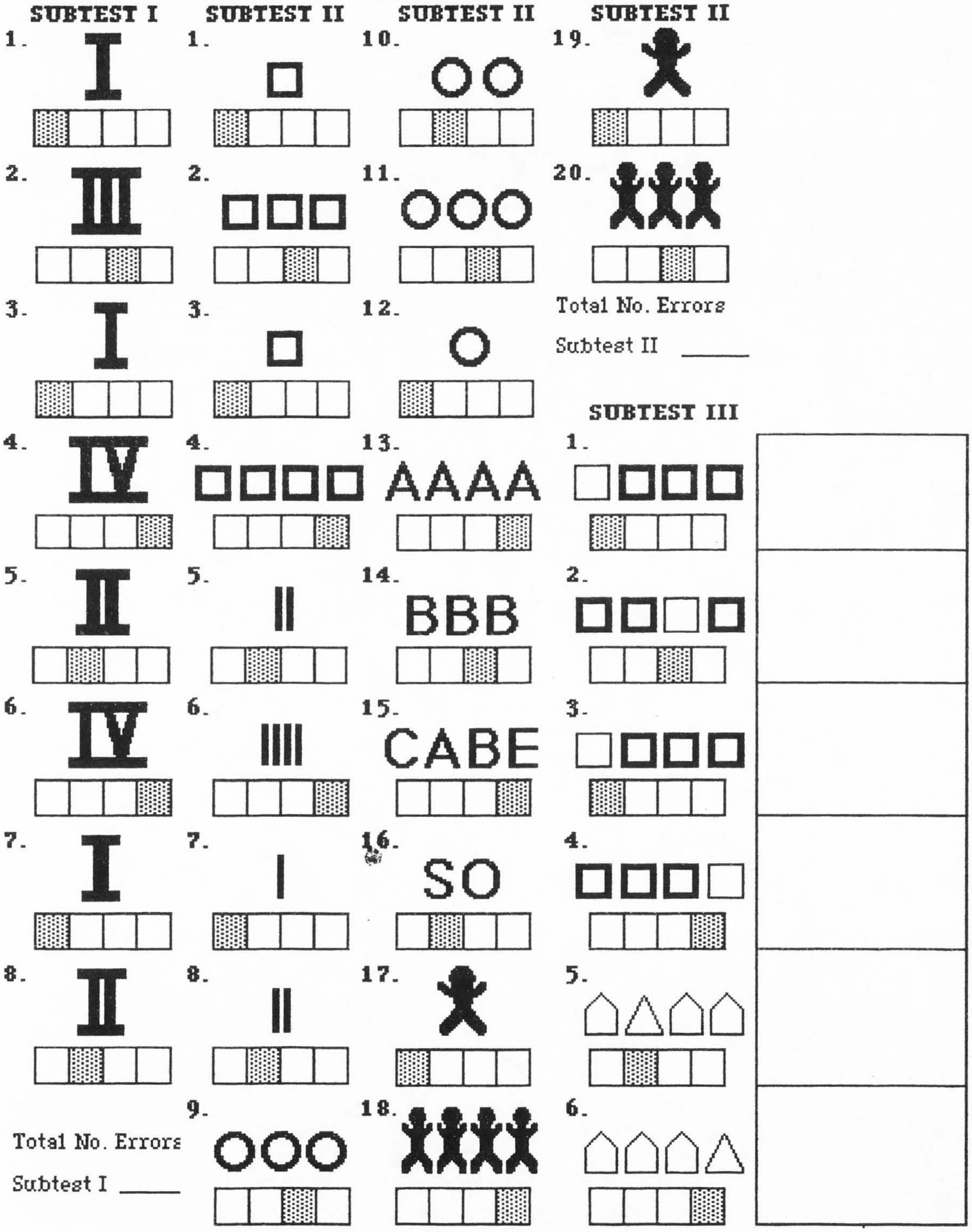
SCORING AND RECORDING FORM FOR THE BOOKLET CATEGORY TEST

Nick A. DeFilippis, Ph.D. and Elizabeth McCampbell, Ph.D.

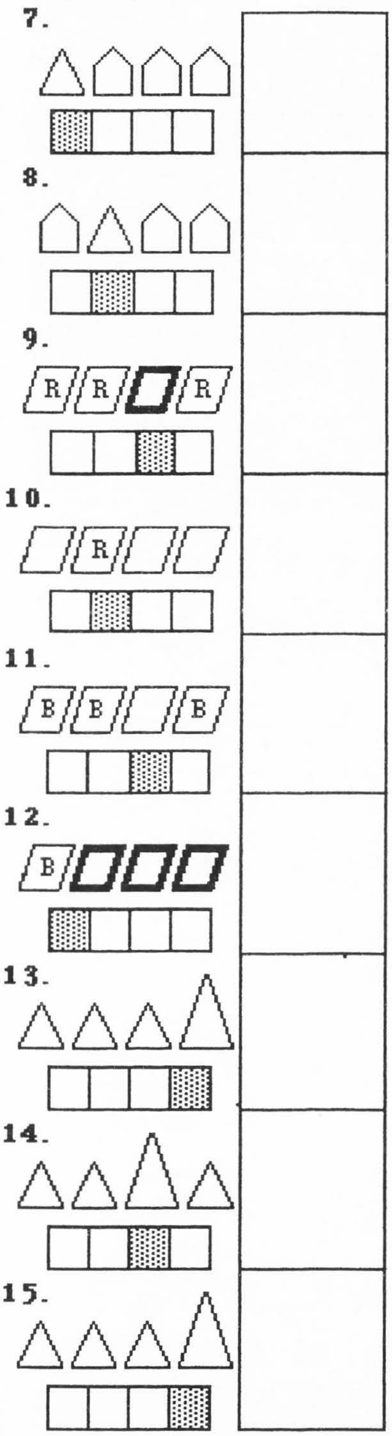
ID Number________________Date________________ Age________________Sex__________

Education______________Occupation________________Premorbid Intellectual Level_______

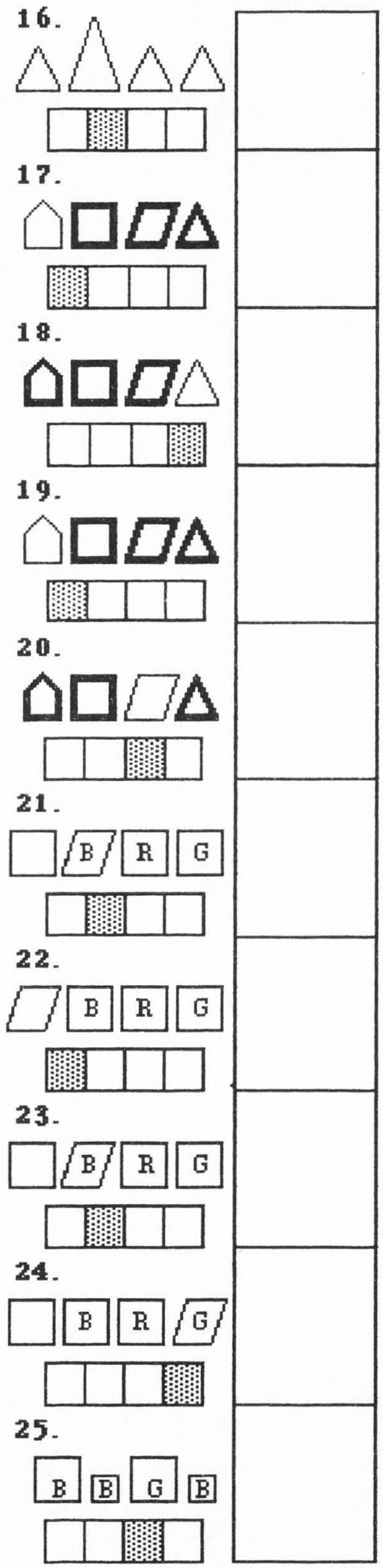
Total Number of Errors Subtests I-VII____



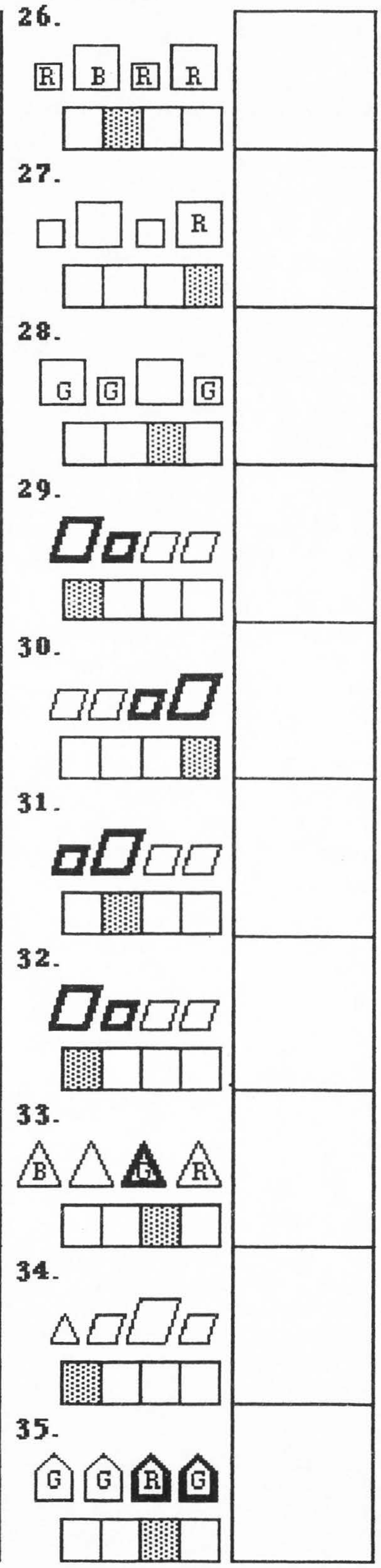
SUBTEST III



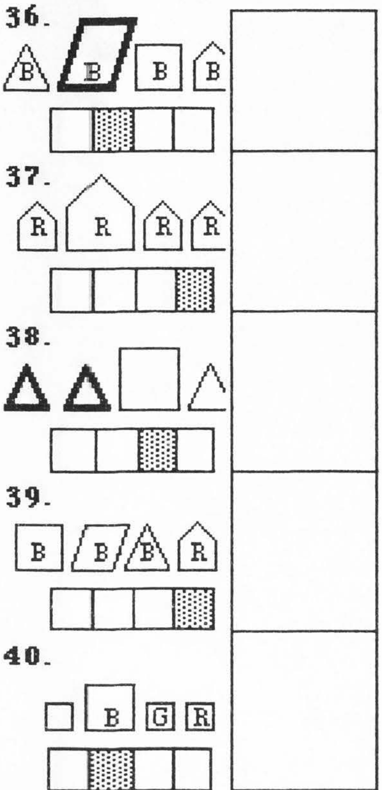
SUBTEST III



SUBTEST III



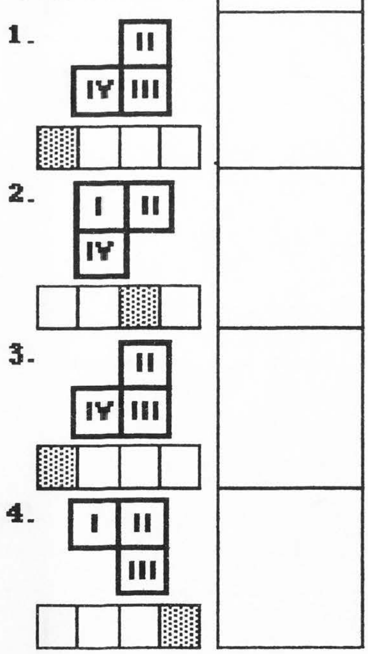
SUBTEST III



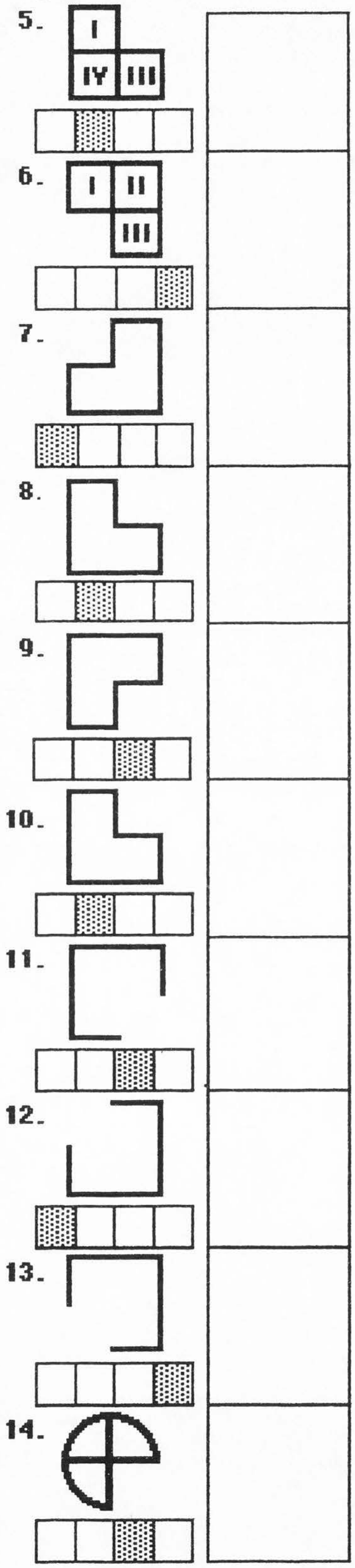
Total No. Errors

Subtest III ______

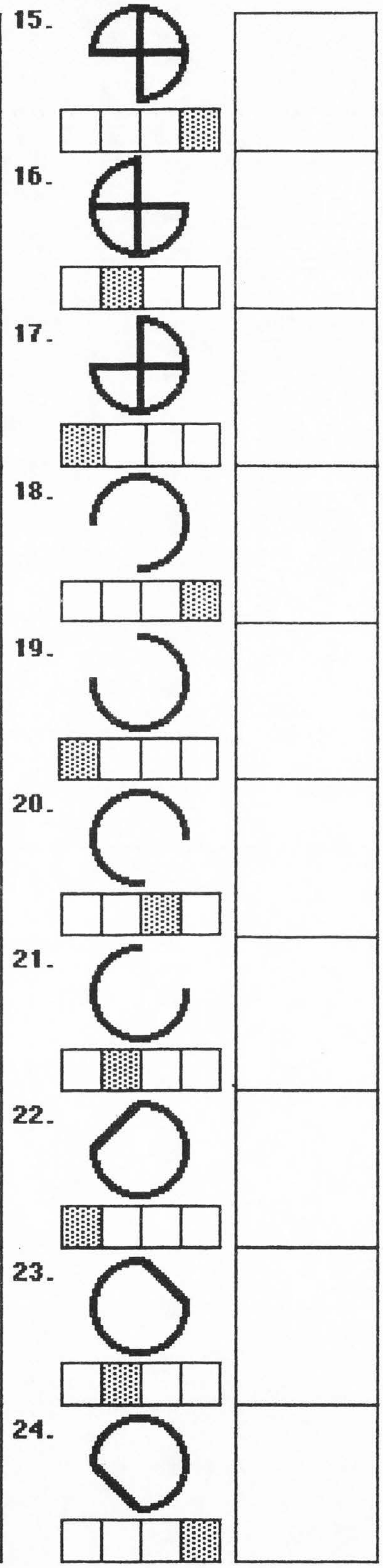
SUBTEST IV

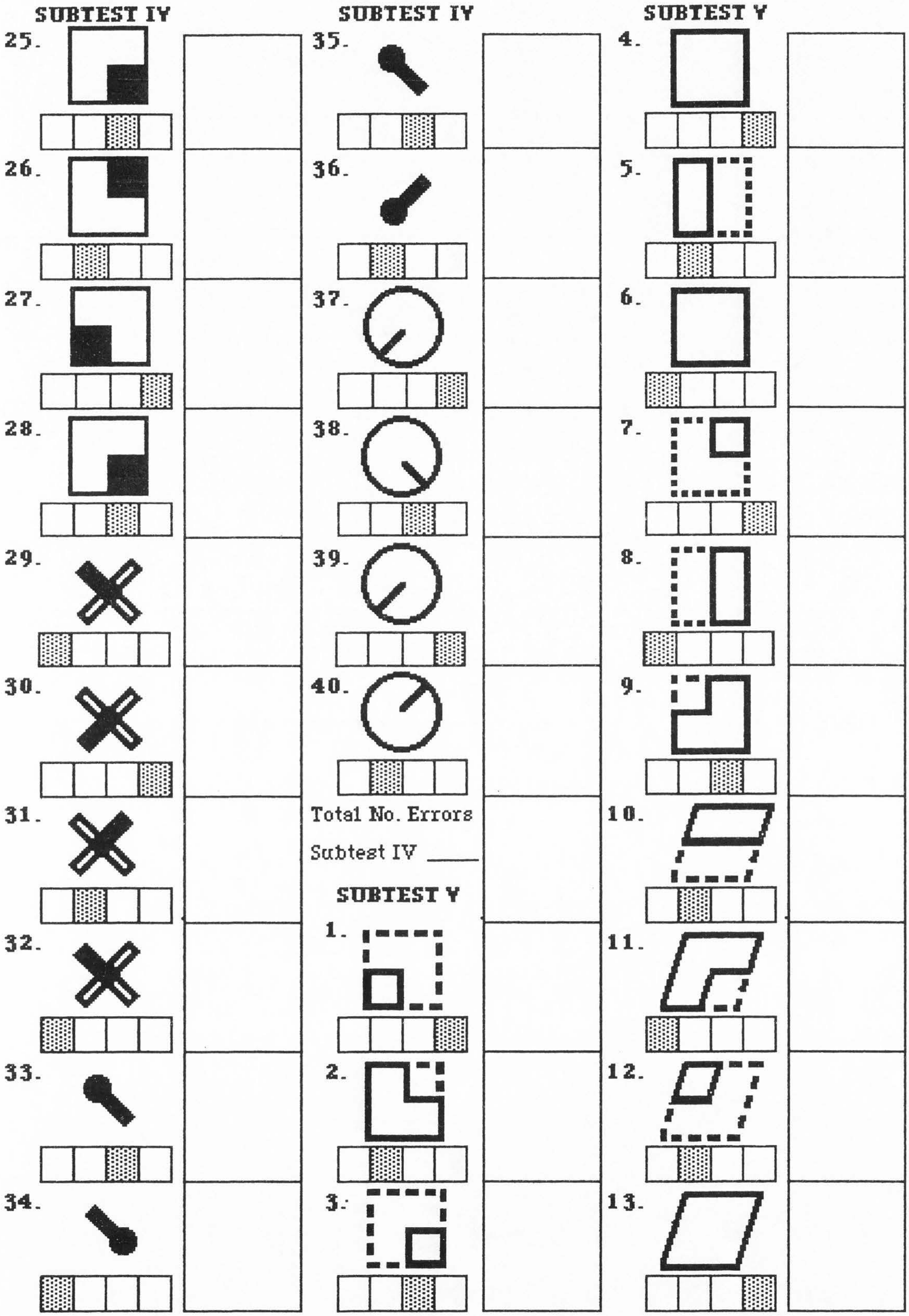


SUBTEST IV

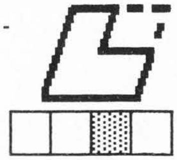


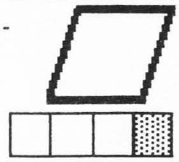
SUBTEST IV

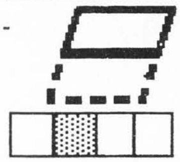


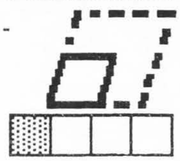

SUBTEST IV
25.
26.
27.
28.
29.
30.
31.
32.
33.
34.
SUBTEST IV
35.
36.
37.
38.
39.
40.
Total No. Errors
Subtest IV _____
SUBTEST V
1.
2.
3.
SUBTEST V
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.

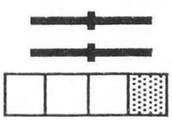
SUBTEST V

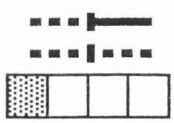
14.


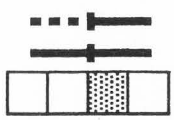
15.


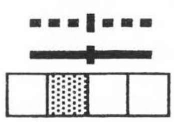
16.


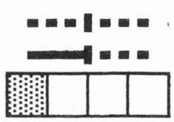
17.


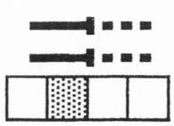
18.


19.


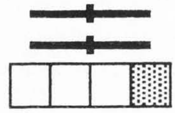
20.


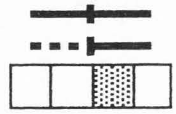
21.


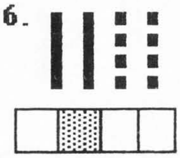
22.


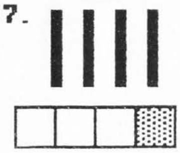
23.


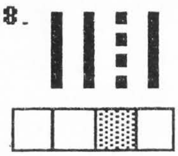
SUBTEST V

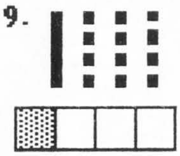
24.


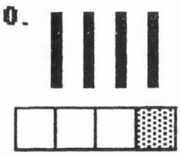
25.


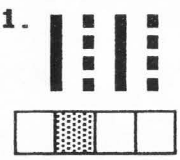
26.


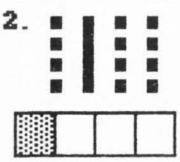
27.


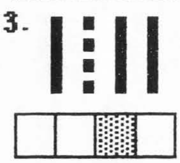
28.


29.


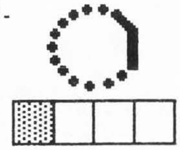
30.


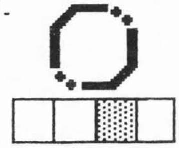
31.


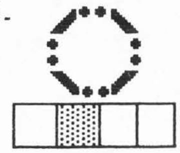
32.


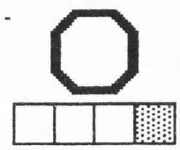
33.


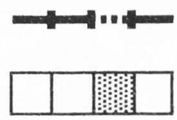
SUBTEST V

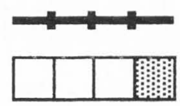
34.


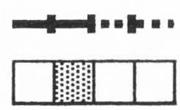
35.


36.


37.


38.


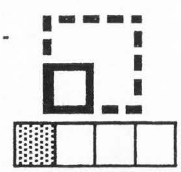
39.


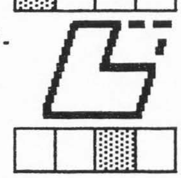
40.


Total No. Errors

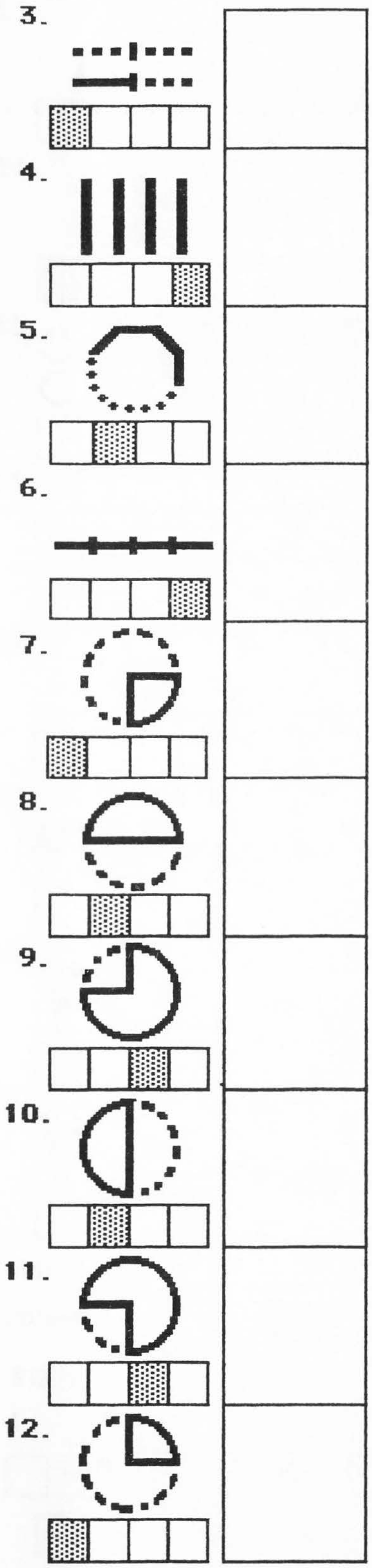
Subtest V ______

SUBTEST VI

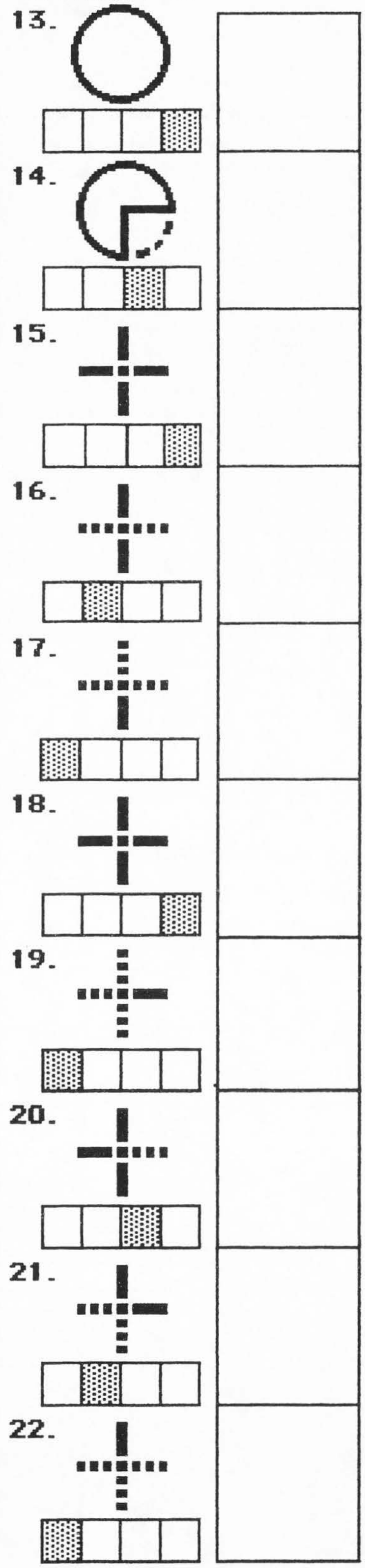
1.


2.


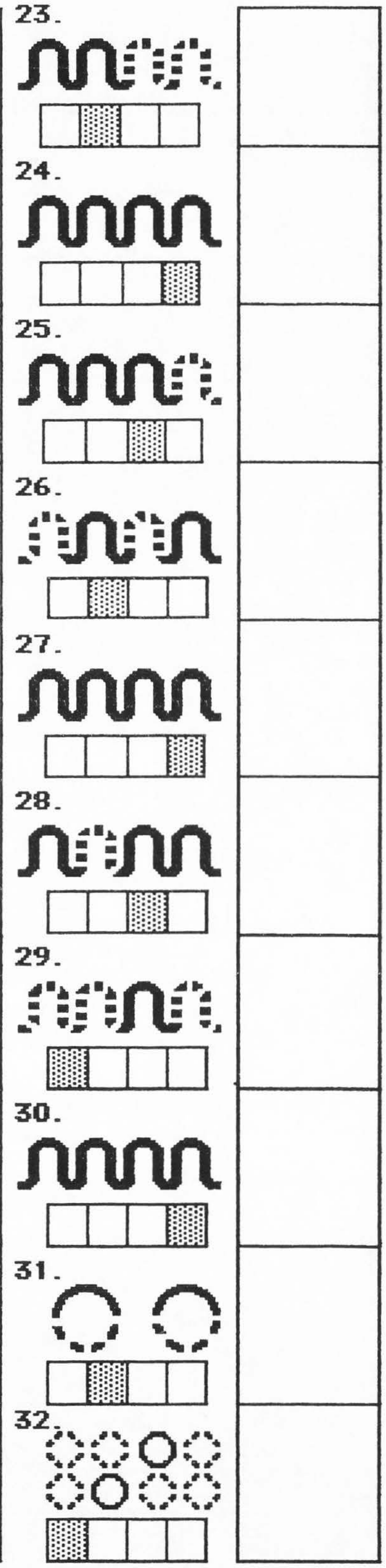
SUBTEST VI

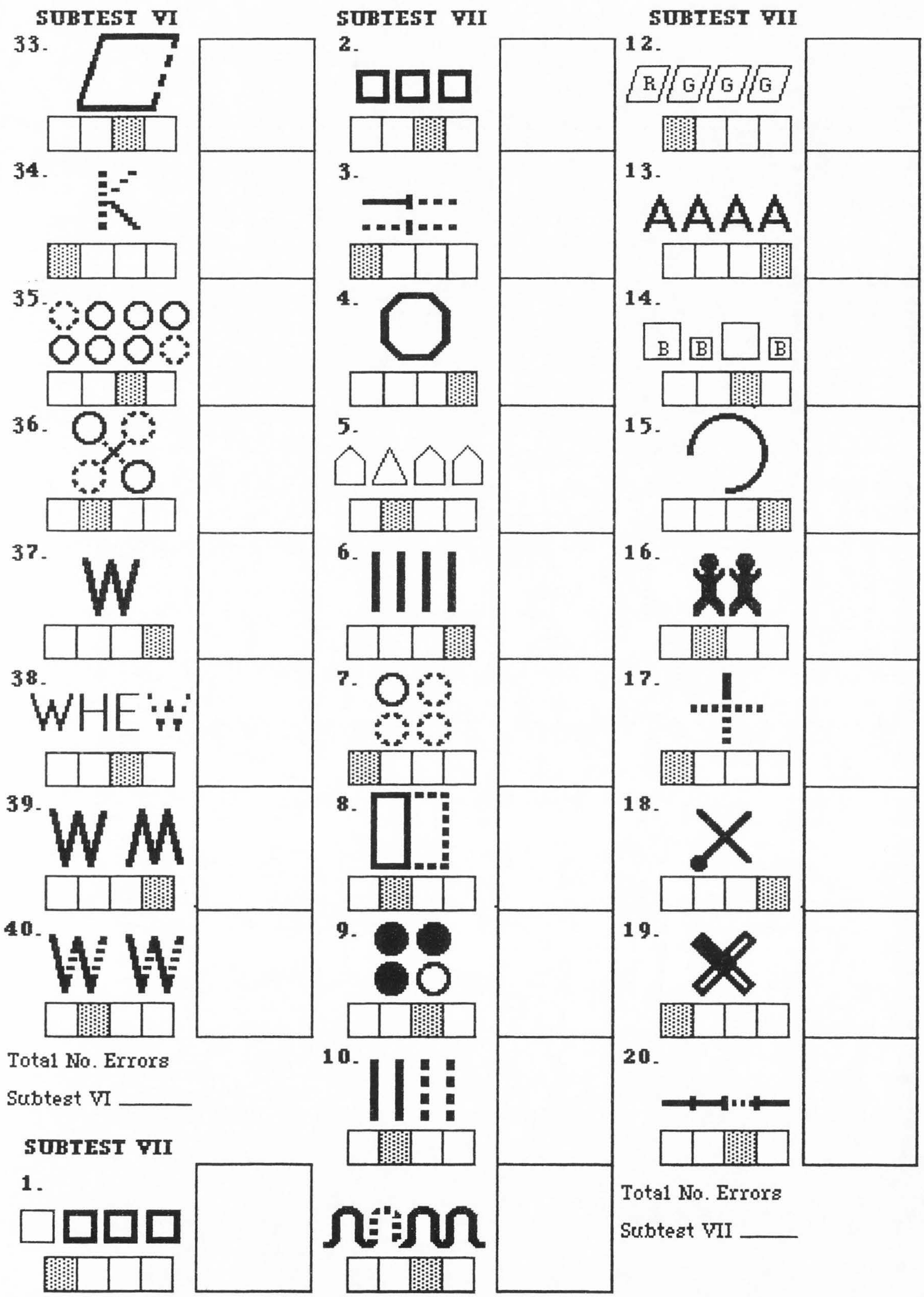


SUBTEST VI



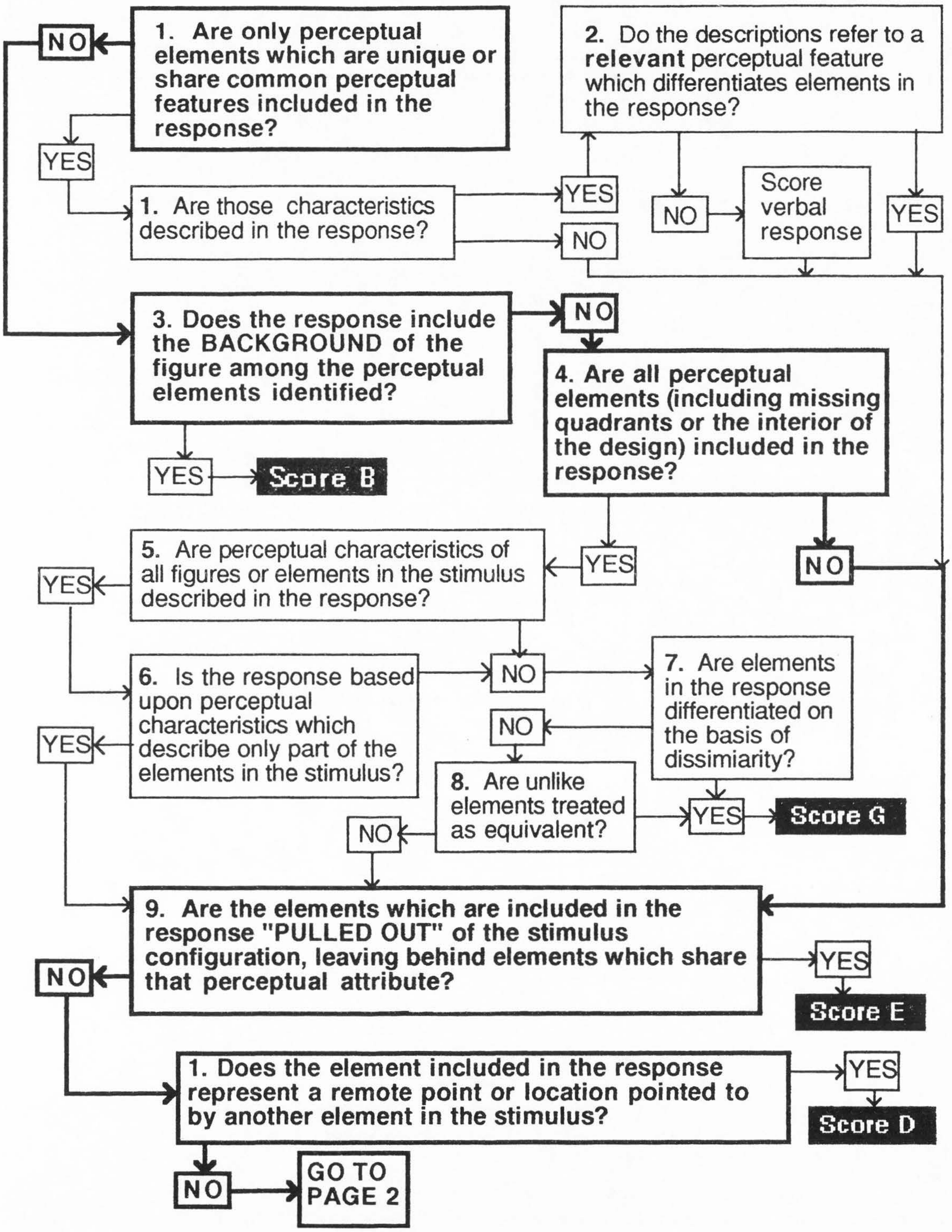
SUBTEST VI

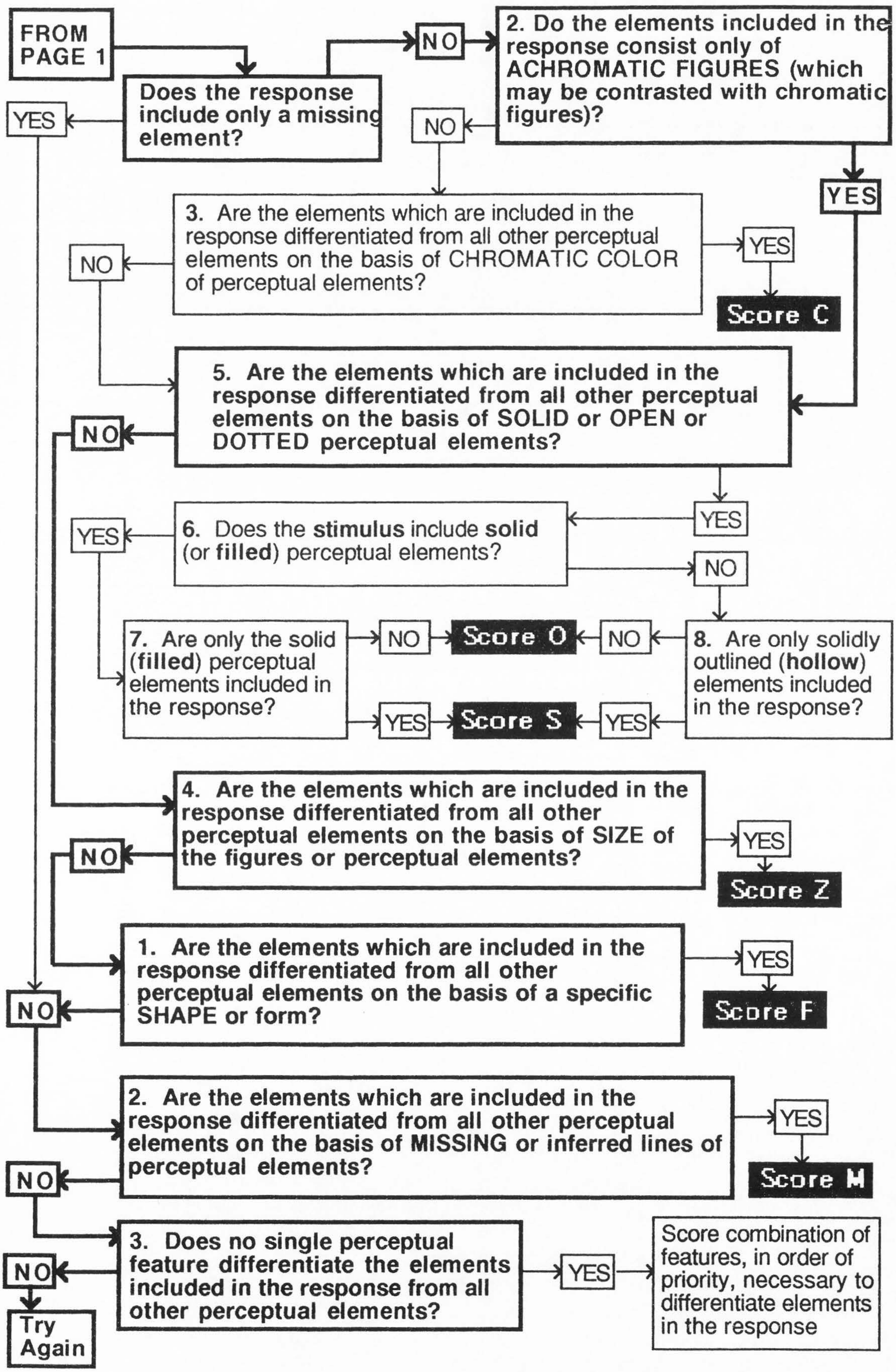



SUBTEST VI
33.
34.
35.
36.
37.
W
38.
WHEW
39.
W M
40.
W W
Total No. Errors
Subtest VI ______
SUBTEST VII
1.
SUBTEST VII
2.
3.
4.
5.
6.
7.
8.
9.
10.
SUBTEST VII
12.
R G G G
13.
AAAA
14.
B B B
15.
16.
17.
18.
19.
20.
Total No. Errors
Subtest VII ______

APPENDIX E. Decision Tree For Coding Configural Attributes On BCT

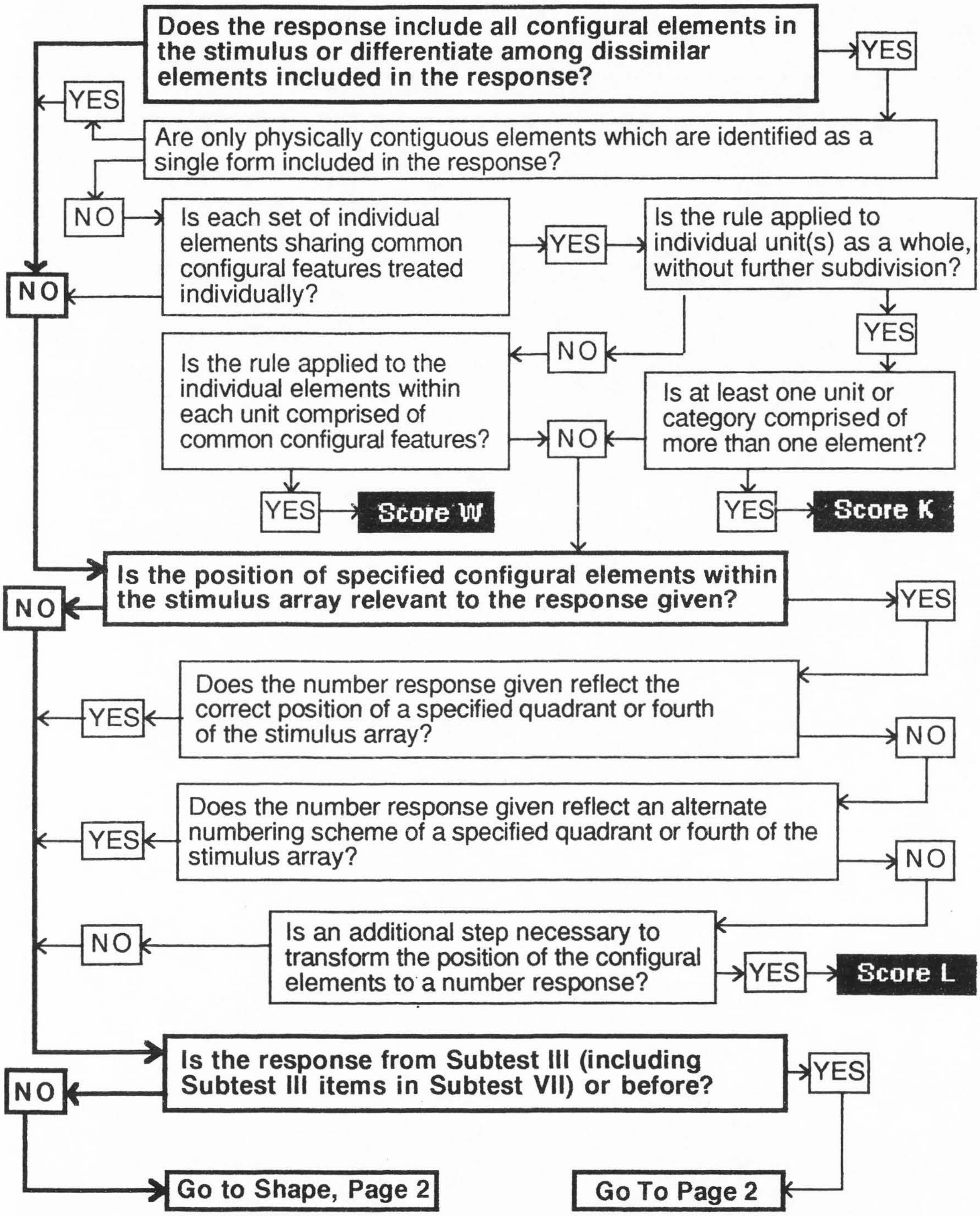
DECISION TREE FOR CODING CONFIGURAL ATTRIBUTES ON BCT

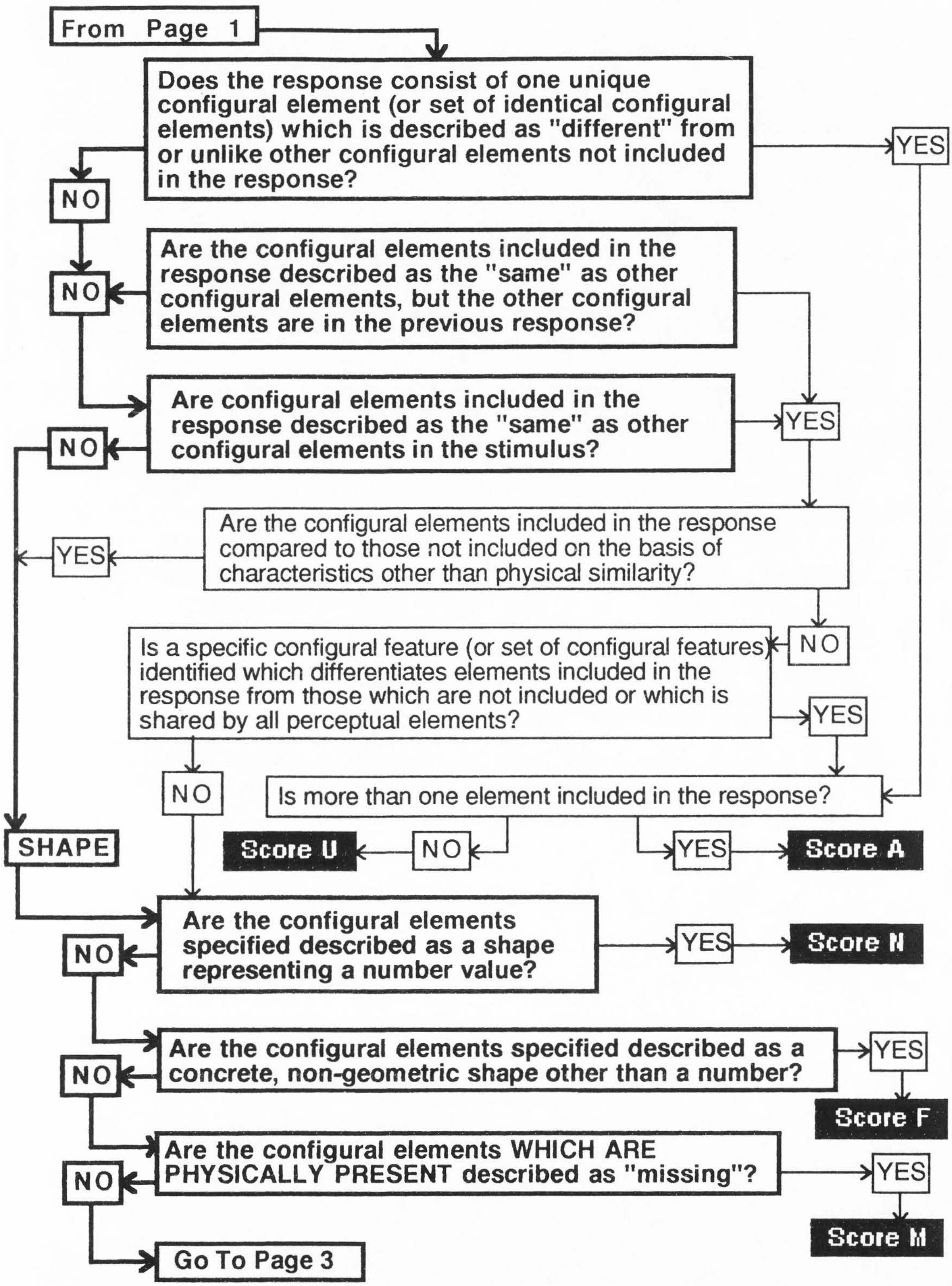


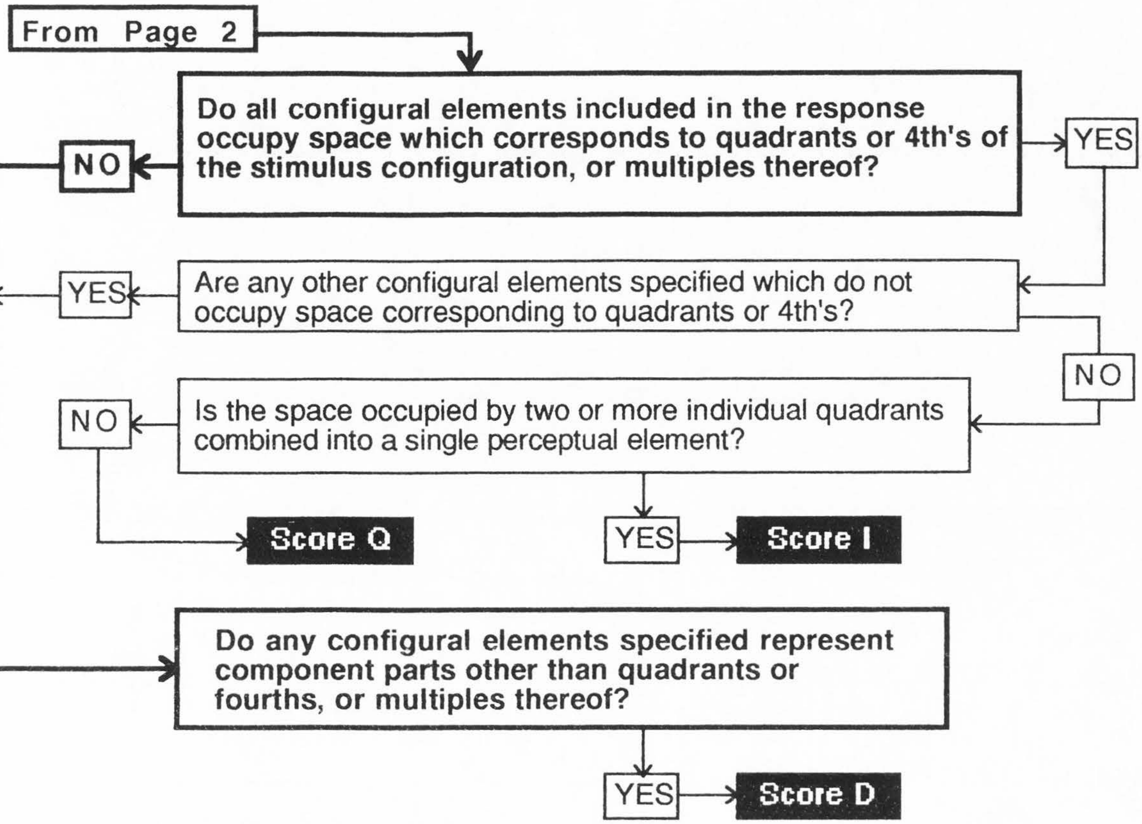

FROM PAGE 1
Does the response include only a missing element?
YES
NO
2. Do the elements included in the response consist only of ACHROMATIC FIGURES (which may be contrasted with chromatic figures)?
NO
YES
3. Are the elements which are included in the response differentiated from all other perceptual elements on the basis of CHROMATIC COLOR of perceptual elements?
NO
YES
Score C
5. Are the elements which are included in the response differentiated from all other perceptual elements on the basis of SOLID or OPEN or DOTTED perceptual elements?
NO
YES
6. Does the stimulus include solid (or filled) perceptual elements?
YES
NO
7. Are only the solid (filled) perceptual elements included in the response?
NO
Score O
NO
8. Are only solidly outlined (hollow) elements included in the response?
YES
Score S
YES
4. Are the elements which are included in the response differentiated from all other perceptual elements on the basis of SIZE of the figures or perceptual elements?
NO
YES
Score Z
1. Are the elements which are included in the response differentiated from all other perceptual elements on the basis of a specific SHAPE or form?
NO
YES
Score F
2. Are the elements which are included in the response differentiated from all other perceptual elements on the basis of MISSING or inferred lines of perceptual elements?
NO
YES
Score M
3. Does no single perceptual feature differentiate the elements included in the response from all other perceptual elements?
NO
Try Again
YES
Score combination of features, in order of priority, necessary to differentiate elements in the response

APPENDIX F. Decision Tree For Coding Abstract Attributes On BCT

DECISION TREE FOR CODING ABSTRACT ATTRIBUTES ON BCT

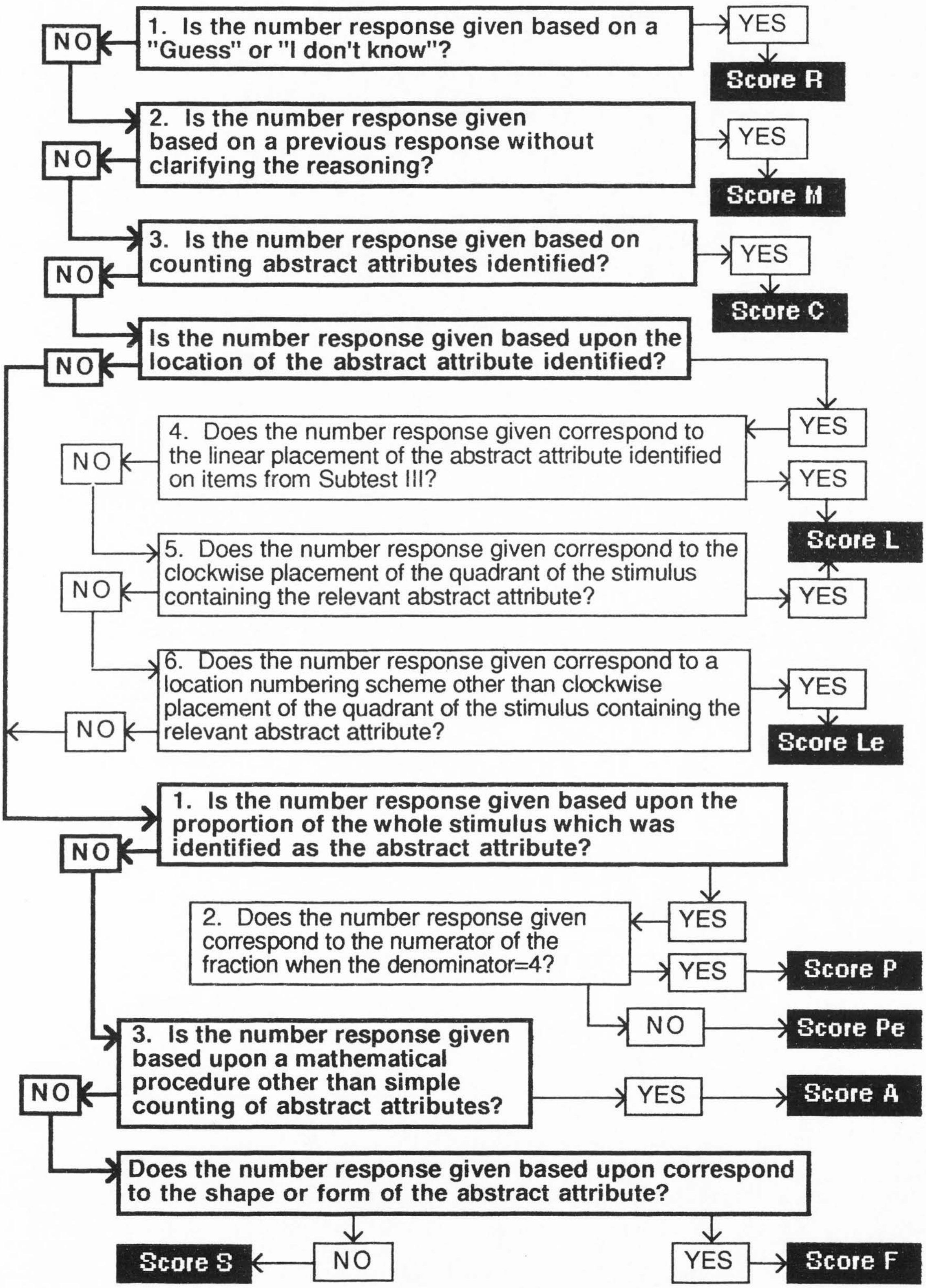



From Page 1
Does the response consist of one unique configural element (or set of identical configural elements) which is described as "different" from or unlike other configural elements not included in the response?
YES
NO
Are the configural elements included in the response described as the "same" as other configural elements, but the other configural elements are in the previous response?
NO
Are configural elements included in the response described as the "same" as other configural elements in the stimulus?
YES
NO
Are the configural elements included in the response compared to those not included on the basis of characteristics other than physical similarity?
YES
NO
Is a specific configural feature (or set of configural features) identified which differentiates elements included in the response from those which are not included or which is shared by all perceptual elements?
YES
NO
Is more than one element included in the response?
SHAPE
Score U
NO
YES
Score A
Are the configural elements specified described as a shape representing a number value?
NO
YES
Score N
Are the configural elements specified described as a concrete, non-geometric shape other than a number?
YES
NO
Score F
Are the configural elements WHICH ARE PHYSICALLY PRESENT described as "missing"?
NO
YES
Score M
Go To Page 3


From Page 2
Do all configural elements included in the response occupy space which corresponds to quadrants or 4th's of the stimulus configuration, or multiples thereof?
YES
NO
Are any other configural elements specified which do not occupy space corresponding to quadrants or 4th's?
YES
NO
Is the space occupied by two or more individual quadrants combined into a single perceptual element?
NO
YES
Score Q
Score I
Do any configural elements specified represent component parts other than quadrants or fourths, or multiples thereof?
YES
Score D

APPENDIX G. Decision Tree For Coding Abstract Attributes On BCT

DECISION TREE FOR CODING SCORING RULES ON BCT



APPENDIX H. Definitions For CT Ratio And Summary Variables

DEFINITIONS FOR CT RATIO AND SUMMARY VARIABLES

VARIABLE	DEFINITION
1. MSIS: Maintain Set Within Sets	Percent of responses in which all 3 determinants did not change following feedback of correct, when test stimuli remained consistent.
2. MSBS: Maintain Set Between Sets	Percent of responses in which all 3 determinants did not change following correct response, when test stimuli did change.
3. LR: Learning Rate	Mean item number across subtests which represents the third consecutive response utilizing all 3 correct determinants.
4. SLR: Maintain Set following Learning	Percent of responses following Learning Rate item in which all 3 determinants were correct.
5. RKC: Keep Concrete Score If Correct	Percent of responses in which the Concrete Score remained unchanged following a correct response.
6. RKA: Keep Abstract Score If Correct	Percent of responses in which the Abstract Score remained unchanged following a correct response.
7. RKR: Keep Rule If Correct	Percent of responses in which the Rule remained unchanged following a correct response.
8. RPCK: Percent Scores Kept When Correct	Percent of total determinant scores which remained unchanged following a correct response.
9. WSC: Shift Concrete Score If Wrong	Percent of responses in which the Concrete score shifted following an incorrect response.
10. WSA: Shift Abstract Score If Wrong	Percent of responses in which the Abstract score shifted following an incorrect response.
11. WSR: Shift Rule If Wrong	Percent of responses in which the Rule shifted following an incorrect response.
12. WS1: Shift 1 Score If Wrong	Percent of responses in which one determinant score changed following an incorrect response.
13. WS2: Shift 2 Scores If Wrong	Percent of responses in which two determinant scores changed following an incorrect response.
14. WS3: Shift 3 Scores If Wrong	Percent of responses in which three determinant scores changed following an incorrect response.
15. WPCS: Percent Scores Shifted When Wrong	Total percent of determinant scores changed following an incorrect response.
16. NC3: Percent Near Correct Subtest III	Percent of responses in Subtest III in which Location comprised the Abstract score rather than the Rule.
17. NC4: Percent Near Correct Subtest IV	Percent of responses in Subtest IV in which quadrant location was not based on clockwise rotation.
18. PRAC4: Percent Subtest IV Practice Correct	Percent of first six responses in Subtest IV (where quadrants are labeled) which are correct.
19. PSV5-6: Percent Subtest V-VI Perseverations	Percent of responses in Subtests V & VI in which the Rule (L or Le) is perseverated from Subtest IV.

APPENDIX I. Definitions for WCST Ratio and Summary Variables

DEFINITIONS FOR WCST RATIO AND SUMMARY VARIABLES

VARIABLE	DEFINITION
1. CAT: Categories Completed	Number of sets of 10 consecutive correct responses.
2. CLR: Conceptual Level Responses	Percent of responses which are correct occurring in runs of 3-10 cards.
3. FTMS: Failure to Maintain Set	Number of sets of 5-9 consecutive correct responses, followed by an incorrect response.
4. LTL: Learning to Learn	Mean of the summed error rate for each category.
5. MLOP: Missed Learning Oppor.	Sum of correct unambiguous responses and incorrect ambiguous responses (which provide sufficient information to deduce principle)
6. NPSVE: Percent Nonperseverative Errors	Percent of responses which are incorrect and do not meet criteria for perseverations.
7. PSVE: Percent Perseverative Errors	Percent of responses which are incorrect and are also perseverative.
8. PSVR: Percent Perseverative Responses	Percent of responses which meet the following criteria: a) would have been correct in the previous category; b) match the first incorrect unambiguous response prior to completion of the first category; c) is an ambiguous responses that is both preceded and followed by unambiguous perseverative responses; or d) is the 2nd or subsequent of three consecutive incorrect unambiguous responses which establish a new "perseverate-to" principle for that category, although ambiguous perseverative responses may be interspersed among those three consecutive perseverative responses.
9. TE: Percent Total Errors	Percent of responses which are incorrect.
10. TC: Percent Correct	Percent of correct responses which are not included in criteria runs.
11. MAX C: Percent Color Resp.	Percent of responses utilizing the color determinant.
12. MAX F: Percent Form Resp.	Percent of response utilizing the form determinant.
13. MAX N: Percent Number	Percent of responses utilizing the number determinant.
14. MAX SD: S.D. of Determinant Responses	Standard deviation of Color, Form, and Number responses (measure of difficulty shifting).
15. SIN C: Unambiguous Color	Percent of responses using unambiguous color determinant.
16. SIN F: Unambiguous Form	Percent of responses using unambiguous form determinant.

VARIABLE	DEFINITION
17. SIN N: Unambiguous No.	Percent of response using unambiguous number determinant.
18. SIN SD: S.D. Single Determinant Resp.	Standard deviation of unambiguous color, form, and number responses.
19. R SH: Right-Shift Responses	Number of responses in which the determinant that was correct in the previous response shifted.
20. WRG ST: Wrong-Stay Responses	Number of responses which repeated the determinants incorrect in the previous response.

CURRICULUM VITAE

Phillip R. Wolfe

518 Clough Avenue
Superior, Wisconsin 54880
(715) 392-4823

Division of Rehabilitation
Polinsky Medical Rehabilitation Center
530 E. Second Street
Duluth, Minnesota 55805
(218) 725-5332

EDUCATION

Ph.D. Program Professional-Scientific Psychology (Clinical, Counseling, School)	Degree Anticipated 12/92 Utah State University Logan, Utah (APA Accredited)
M.A., Clinical Psychology 6/79	George Mason University Fairfax, Virginia
B.A., Major: Social Science 12/72 Minor: English	University of Northern Colorado Greeley, Colorado

CLINICAL EXPERIENCE

September, 1989 - January, 1992
Cognitive Rehabilitation Specialist, Polinsky Medical Rehabilitation Center, Division of Rehabilitation

- Perform psychological and neuropsychological testing and assessment with children and adults
- Provide cognitive rehabilitation treatment for brain-injured patients
- Provide adjustment to disability counseling for patients who have suffered strokes, head injuries, spinal cord injuries, or other neurological impairments

September, 1989 - January, 1992
Cognitive Rehabilitation Specialist, Iowa Methodist Medical Center, Department of Clinical Psychology (710 bed general medical center, children's hospital, and rehabilitation center)

- Perform psychological and neuropsychological testing and assessment with children and adults
- Provide cognitive rehabilitation treatment for brain-injured patients
- Provide adjustment to disability counseling for patients who have suffered strokes, head injuries, spinal cord injuries, or other neurological impairments

- Conduct group and individual psychotherapy for outpatients and inpatients experiencing emotional or mental distress (particular interest in gender issues, substance abuse, and sexual and physical violence)
- Member of Head Injury Team

September, 1988 - August, 1989
Psychology Intern, Knoxville VA Medical Center, Knoxville, Iowa (APA accredited predoctoral internship site)

• Behavioral Health rotation: Group and individual treatment and patient and staff education seminars for habit control, pain control, stress management, assertiveness training, grief and loss issues, depression

• Substance Abuse rotation: Group and individual therapy for substance abusers in long-term treatment program, based on 12-step model. Initiated and co-led gender issues group.

• Neuropsychology rotation: Neuropsychological evaluations for patients with dementia, post-acute head injury and stroke, intracranial tumours, and other neurological disorders

October, 1984 - August, 1988
Graduate student, Utah State University, completed following practicum placements

• Intermountain Sexual Abuse Treatment Center: Performed psychological evaluations and conducted group and individual therapy for sexual offenders referred by Division of Parole and Probation

• Bear River Community Mental Health Center: Conducted individual psychotherapy and performed psychological evaluations and assessments for clients with chronic mental illness or who were experiencing emotional or mental distress

• USU Psychology Community Clinic: Performed psychological evaluations and conducted individual psychotherapy with students and members of the community seeking treatment for a variety of problems ranging from depression to academic and social problems.

• Adult Probation and Parole: Conducted individual psychotherapy with probationers convicted of various crimes

• Clinical Services, Developmental Center for Handicapped Persons: Performed cognitive, academic, and behavioral assessment of children referred for a variety of problems, including suspected Attention Deficit Disorders, learning disorders, enuresis, encopresis, autism, gender identify disorders

• Public School District, Blackfoot, Idaho: Performed psychoeducational evaluations for children referred for learning and behavioral difficulties

July, 1973 - July, 1984
Washington D. C. Department of Corrections, Employed full-time in the following positions:

- Counseling Psychologist, Diagnostic Center: Performed pre-sentence and post-sentence psychological evaluations for youthful offenders and adults
- Counseling Psychologist, Maximum Security Facility: Conducted group and individual therapy and performed psychological evaluations for offenders convicted of capital crimes or presenting management problems
- Correctional Treatment Specialist, Diagnostic Center: Performed social evaluations for youthful offenders prior to sentencing
- Parole Officer, Community Services: Supervised youthful offenders released to parole supervision
- Correctional Treatment Specialist, Youth Center II: Provided case management services for youthful offenders sentenced under the Federal Youth Corrections Act

RESEARCH EXPERIENCE

- Dissertation research in progress, A Cognitive Process Approach to Interpreting Performance on the Halstead Category Test and the Wisconsin Card Sorting Test

TEACHING EXPERIENCE

- Teaching Assistantship: Taught courses (nine sections) in Human Development, General Psychology, and Psychopathology both on-campus and at off-campus extensions